THESE DE DOCTORAT DE L'UNIVERSITE PARIS 6

Spécialité : **Informatique**

Option : **Intelligence Artificielle**

Présentée pour obtenir le titre de

DOCTEUR DE L'UNIVERSITE PARIS 6

Par

Catherine FARON

Sujet de la thèse :

Connaissances taxinomiques : Représentation de taxinomies comportant des exceptions et construction d'hypermédias à base de connaissances taxinomiques

devant le jury composé de :

M. Michel CHEIN	Rapporteur
M. Pascal COUPEY	Examinateur
Mme Rose DIENG	Examinateur
Mme Anne DOUCET	Président
M. Jean-Gabriel GANASCIA	Directeur
M. Amedeo NAPOLI	Rapporteur
M. François RECHENMANN	Examinateur

Remerciements

Je tiens tout d'abord à exprimer toute ma gratitude à mon directeur de thèse, Jean-Gabriel Ganascia, professeur à l'Université Paris 6. Ses conseils judicieux, son exigence, ses critiques pertinentes ainsi que sa présence confiante et bienveillante m'ont permis de mener à bien mon travail de thèse. Mon engagement dans la recherche et l'enseignement doit beaucoup à son enthousiasme contagieux en cours de DEA puis pendant nos réunions d'équipe.

Je remercie Michel Chein, professeur à l'Université Montpellier II, d'avoir accepté de rapporter ma thèse. Je lui suis très reconnaissante pour ses nombreux commentaires éclairants sur mon travail. Ils m'ont été des plus précieux.

Je remercie Amedeo Napoli, chargé de Recherches au CNRS d'avoir accepté de rapporter ma thèse. Je le remercie pour l'intérêt qu'il a bien voulu porter à mon travail, sa lecture attentive de ma thèse, ses conseils avisés et bienveillants.

Je remercie Pascal Coupey, maître de conférence à l'Université Paris 13, Rose Dieng, directeur de recherche à l'INRIA, Anne Doucet, professeur à l'Université Paris 6, et François Rechenmann, directeur de recherche à l'INRIA qui ont bien voulu faire partie de mon jury de thèse. Je remercie Anne Doucet d'avoir présidé ce jury. Une mention particulière à Pascal Coupey pour les discussions enrichissantes que nous avons pu avoir, ses encouragements amicaux et sa lecture attentive de ma thèse.

Je remercie tous les membres de la "famille" ACASA : Irina Akoulchina, Isabelle Bournaud, Christophe Carry, Quyen Kieu, Geber Rahmalo, Irina Tchoumatchenko, Jean-Daniel Zucker et les autres. Chacun, à sa manière, a apporté sa pierre à mon travail. Je remercie tout particulièrement Isabelle Bournaud, la première lectrice de mes premières versions de ce manuscript. Son soutien tout au long de cette difficile période de rédaction m'a été d'une aide précieuse. Je remercie également Quyen Kieu pour son soutien moral durant ces quatre années de thèse. Je remercie Irina Akoulchina pour son entêtement lors des expérimentations de SATELIT en orchidologie. Je leur exprime ici à toutes trois toute ma gratitude et mon amitié.

Je remercie Henri Soldano, maître de conférence à l'Université Paris 13 pour nos discussions amicales, son humour et ses conseils.

Je remercie Marc Champesme, maître de conférence à l'Université Paris 13 pour son aide ponctuelle.

Je remercie Nicole Bidoit, Kazem Lellahi et Monique Pariente aux côtés de qui j’ai agréablement rempli mes fonctions d’enseignante à l’Université Paris 13 en même temps que je rédigeais ma thèse.

Je remercie l’ensemble des membres du LIFO de l’Université d’Orléans pour l’accueil chaleureux qu’ils m’ont fait au sein de leur laboratoire. Je remercie spécialement Siva Anantharaman et Gilles Richard, soucieux de concilier ma charge d’ATER et de bonnes conditions de préparation de ma soutenance de thèse.

Je remercie Pierre Rebufat et François Rebufat pour nos discussions mathémagiques nocturnes enthousiasmées qui ont contribué à mon engagement dans la recherche. Je remercie Anne-Marie Rebufat de les avoir supportées.

Je remercie François Rebufat, mon interlocuteur quotidien, mon compagnon. Je remercie Tartine pour le réconfort poilu et ronronnant qu'elle m'a apporté pendant mes heures de travail.

Catherine

TABLE DES MATIÈRES

Introduction

Motivations

Depuis plus d'un siècle, l'homme est la cause de la disparition de nombreuses espèces animales et végétales. Cet appauvrissement de la diversité biologique ou *biodiversité* est la conséquence d'une utilisation anarchique des ressources de la biosphère. Celles-ci ont été et sont encore utilisées sans retenue dans la fabrication des produits alimentaires, pharmaceutiques, industriels. La prise de conscience des limites des ressources de la biosphère et de l'utilité de la biodiversité est cependant croissante et tend à se mondialiser. La protection de la biodiversité n'est pas une préoccupation nouvelle ; dès 1909 en effet, un premier congrès international s'est tenu à La Haye sur la sauvegarde des paysages. Ce sujet est devenu aujourd'hui source d'une véritable inquiétude. A preuve le Sommet de la Terre (CNUED, Rio de Janeiro, 1992) au cours duquel a été édifiée une convention mondiale (PNUE) visant à réglementer l'appropriation par l'homme des ressources naturelles. Il s'agit de protéger, conserver, voire enrichir la biodiversité [Auroi, 92][Chauvet, 93]. En cette fin de XX° siècle, l'inventaire des espèces vivantes et l'analyse de la diversité biologique redeviennent ainsi un objectif important des biologistes, et la systématique une science contemporaine.

La systématique est la “science des classifications des formes vivantes” [Le Robert, 94]. Elle a pour programme “l'étude et la description de la diversité des êtres vivants, la recherche de la nature et des causes de leurs ressemblances, la mise en évidence des relations de parenté entre eux et l'élaboration de classifications traduisant ces relations de parenté” [Matile, 87]. La systématique est donc une discipline essentielle à l’étude et la protection de la biodiversité. Ainsi, à l'issue du Sommet de la Terre, un projet international a été lancé, baptisé “Systématique Agenda 2000”, ayant pour but l'inventaire de la biosphère. Il s’agit d’inventorier les descriptions de toutes les espèces vivantes terrestres, puis d'analyser et synthétiser ces informations et les organiser sous une forme accessible. Une telle “bibliothèque du vivant” rendrait abordable l'énorme masse d'informations actuellement dispersées dans les collections d'histoire naturelle. De la sorte, l'ensemble des informations utiles à l'analyse de la biodiversité, à la gestion des collections, aux programmes de protection de la nature, de gestion et d'utilisation rationnelle des ressources naturelles seraient aisément disponibles.

Étant donné le gigantisme de la masse d'informations à traiter, l’usage d’outils informatiques devient nécessaire pour mémoriser ces informations sous une forme accessible et faciliter ainsi la tâche d’analyse des systématiciens. C’est dans cette perspective générale que s’inscrivent nos travaux de recherche.

Objectifs

Notre projet est d'élaborer un “catalogue informatique”, sur le modèle des catalogues “papiers” des systématiciens. Un tel outil est destiné à optimiser l'exploitation des informations botaniques ou zoologiques. Ces informations à mémoriser et à organiser sont dispersées sur différents médias : textes, images, croquis, classifications, etc. Elles ont trait à la fois à la description du domaine, de sa terminologie et de ses spécimens, et à la description de l’organisation du domaine proposée par les systématiciens. Les spécimens sont en effet regroupés au sein de classes, encore appelées en systématique *taxons*, qui sont hiérarchiquement organisés en *taxinomie*.

Nous faisons appel aux techniques hypermédias pour concevoir un système qui intègre différents types d'information sur différents médias et offre un accès “non linéaire” à cette information. Pour construire, maintenir et exploiter ce système hypermédia, nous faisons appel à des techniques d'intelligence artificielle et plus précisément celles de représentation et d'acquisition de connaissances. L’organisation et la maintenance du réseau hypermédia sont basées sur l'indexation des documents multimédias par les éléments d'une base de connaissances. La mise en place de primitives d'accès à l'information mémorisée et l’exploitation de cette information font appel à la manipulation formelle des connaissances de la base.

Connaissances Taxinomiques

Nous nous intéressons à la représentation de connaissances pour des domaines tels que la botanique ou la zoologie, domaines qui ont en commun une organisation taxinomique. Considérons en guise d'illustration la famille des Orchidées [Lecoufle, 81], exemple que nous reprendrons dans la suite de cette thèse. Les orchidées sont des plantes *phanérogames monocotylédones* ; elles forment une famille dans l’ordre des *Orchidales*, la famille des *Orchidaceae*. Celle-ci se divise en sous-familles, tribus et sous-tribus ; les sous-tribus se divisent encore en genres qui se divisent eux-mêmes en espèces. Ces niveaux d’organisation en orchidologie sont communs à l’ensemble des domaines des sciences naturelles. En orchidologie, et plus généralement en botanique ou en zoologie, les êtres vivants ayant des traits descriptifs communs sont en effet regroupés et ces groupes sont organisés hiérarchiquement. De telles classifications des êtres vivants sont appelées *taxonomies* ou *taxinomies*. Un *taxon* désigne tout groupe naturel, à quelque niveau de la taxinomie que ce soit. Ainsi, une “famille” est un taxon, de même qu'une “espèce”. La motivation d'une telle organisation des domaines des sciences naturelles est la mise en évidence des traits descriptifs que partagent certains spécimens regroupés sous un même taxon. Ces traits descriptifs constituent ensuite des *clefs d'identification*, qui permettent de reconnaître l'espèce à laquelle appartient un spécimen donné du domaine.

A chaque taxon est donc associé un identificateur unique auquel est associée une description que partagent les spécimens que regroupe le taxon, c’est-à-dire ses instances. Ces descriptions sont structurées, selon une “logique de composition” des parties en sous-parties. Par exemple, la partie aérienne d'une orchidée est formée d'une tige garnie de feuilles et terminée par une inflorescence accompagnée de bractées ; la fleur est formée d'un périanthe et d'un labelle. Les caractéristiques de ces parties et sous-parties participent également aux descriptions des taxons : les feuilles des orchidées de l'espèce *Orchis pallens* sont oblongues et d'un vert uniforme, la forme de l'inflorescence est celle d'un épi ovale serré, la couleur des fleurs est généralement jaune, jaune verdâtre pâle pour le périanthe et jaune vif immaculé pour le labelle [Lecoufle, 81]. Pour représenter les descriptions associées aux taxons, nous avons choisi le formalisme des graphes conceptuels [Chein et Mugnier, 92][Sowa, 84]. Ce formalisme permet à la fois de rendre compte de l'organisation taxinomique du domaine et de la structuration des descriptions associées aux taxons.

Étant donnée la structure taxinomique d'un domaine, nous nous intéressons à l'élaboration d'un langage de description par l'explicitation des traits descriptifs discriminant entre les taxons. La mise en évidence de ces traits discriminants est en effet une manière de rendre compte de la structure d’une taxinomie et d'expliciter ses clefs d'identification. Notre tâche de construction des descriptions des taxons est subordonnée à cet objectif de discrimination. Les graphes conceptuels que nous construisons pour représenter ces descriptions correspondent donc à des conditions nécessaires et suffisantes. Ils constituent les *définitions* des identificateurs des taxons auxquels ils sont associés, c'est-à-dire les représentations intentionnelles de ces taxons [Leclère, 95].

Taxinomies et exceptions

Dans les sciences naturelles, la diversité du vivant rend particulièrement difficile la tâche d'explicitation que nous nous sommes fixée. Dans ces domaines, des spécimens existent en effet qui constituent des exceptions à certains traits descriptifs associés aux taxons dont ils ont été déclarés instances. Or les conditions d'appartenance à un taxon qui constituent la description associée à ce taxon ne couvrent pas ces instances exceptionnelles et les clefs explicitées à partir de telles descriptions ne permettent d'identifier que les instances “typiques” des taxons.

Pour pallier ce problème, une solution consisterait à mettre en disjonction dans les descriptions associées aux taxons les traits descriptifs relatifs à leurs instances exceptionnelles. Cependant, cela ne ferait pas apparaître le caractère exceptionnel de ces traits descriptifs et les descriptions des taxons ainsi modifiées ne suffiraient plus à discriminer certains taxons frères. Elles ne permettraient donc plus de rendre compte de la structure taxinomique du domaine.

Or notre objectif est d’expliciter des descriptions qui précisément permettent de rendre compte de l’organisation taxinomique des domaines modélisés et d’identifier toutes les instances des taxons. Les descriptions explicitées doivent donc être à même de fonder la discrimination des taxons et leur organisation doit coïncider avec celle de la taxinomie à représenter. Le problème qui se pose est de rendre *explicitement* compte des exceptions que comporte une taxinomie, et pour ce faire de mentionner explicitement le caractère exceptionnel de certains traits descriptifs des taxons. Nous proposons pour cela une extension du formalisme des graphes conceptuels qui permet de représenter dans un graphe des exceptions et des défauts. Notre démarche est celle proposée dans [Coupey et Fouqueré, 97] dans le cadre des logiques de descriptions.

Base de Connaissances et Hypermédia

Des connaissances, représentées (par exemple) à l’aide de graphes conceptuels, constituent un modèle de domaine et des facilités d’accès à l’information formalisée. Cependant, elles ne suffisent pas à rendre compte de toute l’information disponible sur un domaine. Cela est particulièrement vrai des domaines des sciences naturelles où un croquis ou une description textuelle ne peuvent être remplacés par une description formelle. Le second aspect de notre travail concerne l'intégration de connaissances à des supports d'information non formalisée et plus précisément à des réseaux d’information hypermédias.

Du point de vue d’une application hypermédia, l’intégration de connaissances formalisées qui représentent un modèle du domaine assure une structuration du réseau hypermédia sémantiquement riche et pallie ainsi les difficultés d’accès à l’information que peuvent rencontrer les utilisateurs. Le risque de désorientation est ainsi évité aux utilisateurs de l’application hypermédia qui accèdent à l’information mémorisée en naviguant sur le réseau hypermédia. De plus, des procédures de recherche d’information “intelligentes” peuvent être intégrées à l'application hypermédia qui reposent sur la manipulation formelle de la base de connaissances intégrée au réseau hypermédia.

Nous nous intéressons plus particulièrement à la conception d’applications hypermédias à base de connaissances du domaine. Nous proposons une méthode de conception fondée sur la simultanéité de l'organisation d’un réseau d'information hypermédia et de l’acquisition d’une base de connaissances taxinomique. Les documents multimédias sont indexés par les éléments de la base de connaissances au fur et à mesure de l’acquisition de ces derniers. La méthode d'acquisition mise en oeuvre constitue de fait les fondements d'une méthode d'organisation du réseau hypermédia semi-automatique. L'organisation initiale donnée ainsi à l'information sous la forme d’un réseau hypermédia est calquée sur celle de la base de connaissances. Ce réseau hypermédia initial peut ensuite être raffiné manuellement, en tirant cependant partie de la structure de la base de connaissances auxquelles ses éléments sont indexés.

Le système SATELIT

Une validation de notre travail est le système de conception d’applications hypermédias à base de connaissances du domaine que nous avons implanté. Il est dédié aux domaines des sciences naturelles et destiné à des utilisateurs non nécessairement experts des domaines modélisés. Nous l’avons baptisé SATELIT, mot valise pour “Système d’Acquisition de Taxinomies et d’ELIcitation de Traits discriminants”. SATELIT met en oeuvre la méthode de construction d’applications hypermédias à base de connaissances que nous proposons.

Plan du mémoire

Le chapitre suivant est consacré à la représentation de connaissances taxinomiques. Nous proposons une caractérisation des connaissances taxinomiques et spécifions les fonctionnalités attendues d’une telle base de connaissances. Partant, nous discutons du choix des graphes conceptuels comme formalisme de représentation. Nous présentons alors ce formalisme en nous attachant à montrer son adéquation à la représentation d'une base de connaissances taxinomiques.

Un problème majeur de représentation rencontré dans les domaines tels que les sciences naturelles est l'occurrence d'exceptions dans les descriptions des taxons à expliciter. Ce sujet a motivé de nombreux travaux en représentation des connaissances. Dans le chapitre 2, nous présentons les solutions qui ont été proposées dans le cadre de différents formalismes de représentation de connaissances et proposons une extension du modèle des graphes conceptuels pour la gestion des exceptions, qui s'inspire d'une démarche issue de travaux en logiques de descriptions.

Les raisonnements que nous mettons en oeuvre sur des connaissances taxinomiques représentées dans le modèle des graphes conceptuels sont décrits dans le chapitre 3. Ils consistent d'abord en l'acquisition de connaissances taxinomiques. La structure taxinomique du domaine étant connue, il s'agit d'expliciter des définitions des taxons dont les relations d'ordre entre elles rendent bien compte de cette taxinomie. Pour cela, nous adoptons une démarche descendante d'explicitation de la définition d'un taxon à partir de celle de son père dans la taxinomie. Sur la base de connaissances ainsi construite, les raisonnements effectués consistent ensuite en l'identification d'objet, c'est-à-dire la reconnaissance du taxon auquel appartient un objet donné du domaine, par comparaison de leurs descriptions respectives.

Dans le chapitre 4, nous nous intéressons à la collaboration entre connaissances taxinomiques et hypermédia. Nous nous concentrons plus particulièrement sur la conception d'hypermédias à base de connaissances du domaine. Nous proposons une méthode de conception d'hypermédia à base de connaissances taxinomiques en mettant en évidence son originalité par rapport aux approches existantes. Elle consiste d'une part en une organisation du réseau hypermédia fondée sur la structure d'une base de connaissances taxinomiques représentées à l'aide de graphes conceptuels. D'autre part, la méthode de construction du réseau est fondée sur la méthode d'acquisition des connaissances présentée dans le chapitre précédent, les éléments du réseau hypermédia étant indexés à ceux de la base de connaissances au fur et à mesure de leur acquisition.

Le chapitre 5 est dédié à la mise en oeuvre du système SATELIT qui implante la méthode de conception d'hypermédia à base de connaissances décrite dans le chapitre 4. Nous présentons l'architecture du système et décrivons un exemple d'utilisation pour la construction d'une application hypermédia à base de connaissances en orchidologie.

Chapitre 1

Représentation de connaissances taxinomiques

Introduction

Dans ce chapitre, nous commençons par préciser ce que nous entendons par “connaissances taxinomiques”. Pour ce faire, nous dégageons leurs particularités en distinguant celles liées à la spécificité des domaines modélisés, d’une part, et celles qui résultent de l’exploitation qui est faite des bases de connaissances taxinomiques, d’autre part. Sur cette base, nous produisons dans la seconde partie les qualités attendues d’un formalisme pour représenter des connaissances taxinomiques. Nous mettons alors en évidence l’adéquation du formalisme des graphes conceptuels à cette tâche particulière de représentation. Dans la troisième partie de ce chapitre, nous décrivons le formalisme des graphes conceptuels simples [Mugnier et Chein, 96]. Enfin, nous montrons sur un exemple en orchidologie l’utilisation que nous faisons de ce formalisme pour représenter des connaissances taxinomiques.

1.1. Connaissances taxinomiques

1.1.1. Caractéristiques descriptives des connaissances taxinomiques

Nous nous efforçons de dégager dans cette partie la spécificité des connaissances explicitées en sciences naturelles. C’est en effet en botanique et en zoologie qu’a vu le jour cette manière particulière d'organiser les connaissances d'un domaine dans des structures d'ordre appelées classifications (§ 1.1.1.1.). La terminologie de l'activité classificatoire en est issue, et notamment la qualification de "taxinomiques" des connaissances d'un domaine organisées sur le modèle des classifications botaniques ou zoologiques (§ 1.1.1.2.).

Les organisations taxinomiques sont caractérisées par leur complexité (au sens ordinaire du terme). Plus précisément, à une complexité en terme de *taille* des classifications s'ajoutent d'une part une complexité liée à la *multiplicité des critères* classificatoires (§ 1.1.1.3.) et d'autre part celle des *descriptions structurées* des classes (§ 1.1.1.4). C'est l'ensemble de ces propriétés que nous prendrons en considération pour choisir un formalisme adapté à la représentation de connaissances taxinomiques.

1.1.1.1. Organisation du vivant : unité et diversité

Une caractéristique essentielle du vivant est sa diversité. Comme le souligne Mayr, "Il n'y a pas deux individus identiques au sein d'une population se reproduisant sexuellement, pas deux populations au sein d'une espèce, pas deux espèces au sein d'un taxon d'ordre supérieur, pas deux associations d'espèces au sein d'un écosystème, et ainsi de suite à l'infini. Où que l'on regarde, on trouve la singularité et la singularité signifie la diversité" [Mayr, 89].

Cependant, malgré leur diversité de taille, de forme ou de mode de vie, beaucoup d'êtres vivants sont construits sur un même modèle de base, appelé *plan d'organisation*. Pour cette raison, les êtres vivants sont classés dans des embranchements, ceux d'un même embranchement partageant certaines caractéristiques générales.

Les embranchements ne sont pas l'unique niveau d'organisation du monde vivant. Traditionnellement en botanique ou en zoologie, ils sont divisés en classes, elles-mêmes divisées en ordres, à leur tour divisés en familles, divisées en genres, divisés en espèces. Ces différents niveaux d'organisation sont présentés sur la figure 1.1. Ils sont issus de la classification des plantes proposée par Linné et admis depuis le XVIII° siècle. Des niveaux intermédiaires sont souvent intercalés (sous-embranchements, sous-familles, etc.) ou ajoutés aux feuilles d'une classification (groupes, sous-groupes) pour raffiner l'organisation hiérarchique de celle-ci.

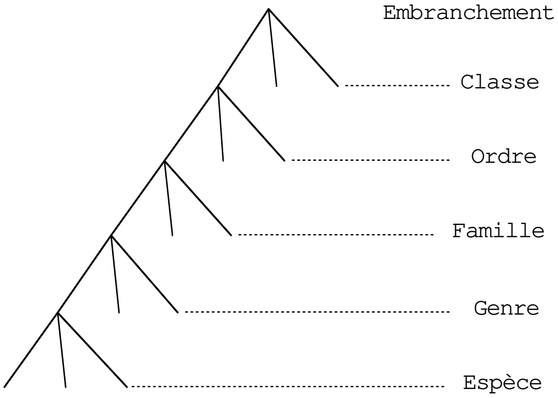


figure 1.1. Les rangs taxinomiques proposés par Linné

Le nombre de niveaux (la profondeur des organisations hiérarchiques construites) et le nombre de regroupements pour chaque niveau (la largeur) témoignent du degré élevé de conceptualisation de la classification du vivant. On compte en effet jusqu'à 21 rangs ; on connaît plus de 100000 genres, plus d'un million d'espèces.

1.1.1.2. Terminologie

La *systématique* est la science dont l'objet d'étude est la diversité du vivant ; les systématiciens cherchent en effet à mettre en évidence et à expliquer les régularités qui existent dans la diversité du vivant. Pour ce faire, ils rassemblent, condensent et organisent tout le savoir concernant les organismes vivants [Mayr, 89]. La tâche d'organisation apparaît donc comme centrale dans l'activité des systématiciens, comme en témoigne la définition de la systématique proposée dans [Le Robert, 94] : elle désigne la "science des classifications des formes vivantes et des espèces chimiques".

En systématique, les classifications d'êtres vivants sont appelées des *taxinomies*. Une taxinomie est un ensemble de classes ordonnées par une relation d'inclusion stricte, c'est-à-dire dont la structure est celle d'une hiérarchie stricte ou plus formellement celle d'un arbre. Un *taxon* est un regroupement d'êtres vivants de niveau d'organisation, ou *rang taxinomique*, quelconque [Mayr, 89]. Étymologiquement, la racine grecque *taxis* signifie "arrangement, ordre" et exprime l'idée de "placer où il convient, selon une organisation" [Le Robert, 94]. Pour poursuivre en ces termes, une fonction essentielle des taxinomies est de "placer où il convient" un être vivant, c'est-à-dire de *l'identifier* en déterminant le taxon auquel il appartient. En effet, à chaque rang taxinomique, l'ensemble des taxons de ce rang constitue une partition plus ou moins fine de l'ensemble des spécimens instances du taxon racine de la taxinomie.

Nous qualifions de *taxinomiques* l'ensemble des connaissances relatives à une taxinomie, à sa description, à la description de ses taxons, à la description des plans d'organisation.

L'usage de cette terminologie propre aux sciences naturelles s'est étendu à d'autres domaines dont l'organisation des connaissances est similaire. De fait, [Le Robert, 94] propose une définition plus générale du terme “systématique” comme “ensemble de vues et de méthodes relevant d'un système de pensée”. Des connaissances taxinomiques sont en effet explicitées non seulement en botanique et en zoologie mais aussi en minéralogie, en archéologie, en linguistique (taxinomies de caractères chinois [Zucker, 96][Bournaud, 96]). En ingénierie des connaissances, des taxinomies sont encore élaborées pour modéliser des domaines aussi variés que l'imputation comptable [Bournaud, 96], l'analyse de risque des transports [Darricau et al, 97]... Le lecteur intéressé par un exposé des enjeux de la classification pour ces nouveaux domaines d'application pourra se reporter au premier chapitre de la thèse de Bournaud [Bournaud, 96].

1.1.1.3. Complexité des taxinomies

Une taxinomie organise une masse de connaissances considérable. Nous avons précédemment souligné l'importance de la taille des taxinomies en sciences naturelles, du nombre de taxons et de niveaux taxinomiques (cf. § 1.1.1.1.). La complexité du vivant ne peut cependant être toute contenue dans une taxinomie, aussi fine soit-elle. Des espèces sont encore à découvrir, tandis que d'autres disparaissent ou évoluent. Une taxinomie est donc vouée à la *réorganisation*, chaque fois qu'un événement vient remettre en cause l'organisation en vigueur du domaine.

De plus, une structure taxinomique qui organise un ensemble d'organismes vivants n'est pas unique. Au contraire, des taxinomies différentes peuvent coexister, concurrentes ou complémentaires. Une première explication est l'usage de traits descriptifs différents pour fonder ces taxinomies. Ces différences sont le résultat de l'existence de différents *points de vue* possibles en biologie. En particulier, les descriptions construites dans une perspective fonctionnaliste de compréhension du mode d'action des éléments structuraux d'un organisme et de leurs interactions sont évidemment différentes de celles construites dans une perspective évolutionniste de compréhension des causes des caractéristiques d'un organisme et de ses adaptations. Nous renvoyons le lecteur intéressé à [Mayr, 89] pour une présentation de la biologie fonctionnaliste et de la biologie évolutionniste. Des différences de *niveau de description* sont également à l'origine de différences de points de vue. Plus précisément, les taxinomies construites à partir de descriptions qui ont trait à la morphologie des êtres vivants diffèrent de celles qui traitent d’anatomie, de cytologie, de biochimie, ou encore de génome [Conruyt, 94]. La concordance ou au contraire la divergence des points de vue qui président à la construction d’une taxinomie peuvent être mises en relief par la comparaison, la recherche de relations entre les taxinomies qui en résultent. Enfin, même sur un ensemble identique de connaissances descriptives, des structures taxinomiques concurrentes peuvent être élaborées, selon l’ordre dans lequel on considère les traits descriptifs discriminants. Par exemple, “chez les animaux, on arrive à des classifications totalement différentes selon que l’on retient comme premiers caractères différentiels la présence ou l’absence de sang, la présence ou l’absence de poils, la présence de deux ou quatre pieds” [Mayr, 89].

La multiplicité des critères classificatoires pour construire une taxinomie s’ajoute à la taille des taxinomies construites. Ces caractéristiques définissent ensemble la complexité des *structures* taxinomiques. Nous les distinguons en cela des caractéristiques des *descriptions* des taxons, liées à la complexité des organismes vivants décrits.

1.1.1.4. Complexité des organismes vivants

Pour rendre compte de la structure des taxinomies qu’ils construisent, les systématiciens mettent en évidence à chaque niveau d’organisation d’une taxinomie les critères classificatoires qui ont présidé à sa construction. Il s'agit des traits descriptifs qui permettent de discriminer entre taxons frères. Une description associée à un taxon est commune à l'ensemble des spécimens qu’il regroupe. La construction d’une taxinomie repose donc sur une étude descriptive soignée et détaillée des organismes vivants à classer.

Eu égard à la *complexité* des organismes vivants, décrire ces derniers peut sembler difficile. Mais cette complexité ne répond pas au hasard. Au contraire, les êtres vivants sont organisés : ils sont dotés de *structures* qui les caractérisent. Plus précisément, un organisme est descriptible par la structure hiérarchique de ses parties et sous-parties et les caractéristiques de celles-ci, selon une logique dite *de composition* [Conruyt, 94]. Remarquons en effet que les éléments d’un organisme sont dépourvus de sens hors de l’organisme : “les ailes, les jambes, les têtes, les reins ne peuvent vivre par eux-mêmes, mais seulement en tant que parties de l’ensemble” [Mayr, 89]. La logique de composition consiste donc en un mécanisme de décomposition en parties et sous-parties, itéré autant de fois que nécessaire pour atteindre la finesse de description souhaitée. Elle définit une structure arborescente de parties et sous-parties selon laquelle progresse la description, les caractéristiques d'une partie étant décrites avant celles de ses sous-parties.

D'autre part, la difficulté à décrire des organismes vivants réside également dans la définition et l'usage d'un vocabulaire spécial. Particulièrement pour rendre compte de la structure d'une taxinomie, les descriptions associées à ses taxons doivent en effet être très précises. Elles font de fait appel à une terminologie riche et nuancée. Ainsi la complexité des organismes vivants se traduit non seulement par des descriptions structurées mais aussi par un vocabulaire du domaine spécialisé.

1.1.1.5. Conclusion

Les taxinomies sont propres à des domaines pour lesquels la masse des connaissances et leur diversité nécessitent de les organiser pour les appréhender en mettant à jour une unité. La complexité des connaissances taxinomiques demeure cependant importante. Nous avons en effet mis en évidence dans cette partie la grande taille des taxinomies et la multiplicité des critères de classification d'une part et d'autre part des descriptions de taxons structurées et exprimées dans une riche terminologie.

La représentation formelle de connaissances taxinomiques est un moyen de mieux les appréhender en les manipulant plus facilement. Un formalisme adapté à la représentation de connaissances taxinomiques doit permettre de faire face à leur complexité. Plus précisément, il doit permettre (1) de représenter des descriptions riches et nuancées d'objets structurés, (2) de représenter un grand nombre de telles descriptions, (3) de comparer, discriminer ces descriptions et de représenter leur organisation taxinomique, (4) de modifier facilement une description ou une partie de la structure taxinomique, (5) de comparer différentes taxinomies.

Ces contraintes dans le choix d'un formalisme de représentation de connaissances taxinomiques que nous avons mises en évidence sont celles imposées par les caractéristiques descriptives des connaissances taxinomiques. La partie suivante est dédiée aux contraintes imposées par l'utilisation qui est faite de ces connaissances.

1.1.2. Caractéristiques fonctionnelles d'une base de connaissances taxinomiques

Nous nous attachons ici à caractériser une base de connaissances taxinomiques eu égard à ses différentes fonctions. En premier lieu, une base de connaissances taxinomiques a un rôle “statique” de *mémorisation* des informations du domaine. Cela est commun à l'ensemble des bases de connaissances descriptives mais nous montrons qu'un domaine pour lequel les connaissances sont organisées en taxinomie impose en outre une démarche particulière dans l'élaboration d'une base de connaissances (§ 1.1.2.1.). L'appréhension du domaine modélisé procède alors de la manipulation formelle d'une base de connaissances. Nous mettons en évidence les procédures d'accès aux connaissances particulières à l'organisation taxinomique du domaine (§ 1.1.2.2.). Enfin, la structure taxinomique d'une base de connaissances est destinée à supporter un accès hypermédia aux informations textuelles ou graphiques indexées par les éléments de cette base (§ 1.1.2.3.).

1.1.2.1. Un modèle du domaine

La motivation première dans l'élaboration d'une base de connaissances taxinomiques est le modèle du domaine qu'elle représente. Des connaissances descriptives sont explicitées, qui représentent les concepts du domaine et les relations entre concepts.

Il faut distinguer deux activités connexes dans l'élaboration d'une base de connaissances taxinomique. La première consiste à *construire* une taxinomie, c'est-à-dire à organiser les connaissances relatives à un domaine et en extraire une structure hiérarchique. C'est ce qu'on appelle le *regroupement conceptuel*[1] [Bournaud et Ganascia, 96] [Furtado Vasco et al, 96]. La seconde consiste non pas à élaborer une organisation des connaissances mais à expliciter cette organisation. Plus précisément, il s'agit dans cette approche d'acquérir et de représenter des connaissances dont la structure taxinomique a été antérieurement construite. La tâche d'explicitation s'adresse alors non pas à la structure du domaine mais à la *terminologie* nécessaire à la définition des concepts du domaine. C'est à ce second point de vue dans l'élaboration de bases de connaissances que nous nous intéressons.

Étant *donnée* la structure taxinomique d'un domaine, nous nous appliquons à définir les concepts du domaine, essentiellement les taxons. Pour ce faire, nous explicitons les primitives de langage nécessaires à leurs définitions. La méthode d'acquisition de connaissances descriptives et terminologiques à laquelle nous faisons appel est décrite dans le chapitre 3 de cette thèse.

---

[1] En anglais : *conceptual clustering*.

1.1.2.2. Un accès au domaine

Une structure taxinomique permet de faciliter l'accès aux connaissances acquises. Une taxinomie peut en effet être envisagée comme un système *d'index* pour stocker et retrouver aisément les connaissances sur un domaine. Dans cette perspective, les systématiciens utilisent une taxinomie comme un *système de détermination* des taxons dont sont instances les spécimens examinés. Cette activité est encore appelée *identification* d'un spécimen, par détermination ou reconnaissance du taxon le plus spécifique dont le spécimen est instance. Ce mode particulier d'accès à l'information emmagasinée est une fonctionnalité essentielle d'une base de connaissances taxinomiques qui modélise le domaine.

Toute base de connaissances est dotée de mécanismes d'exploitation qui permettent d'accéder aux connaissances mémorisées, d'exprimer des requêtes, et d'appréhender ainsi le domaine modélisé. Ces procédures sont fondées sur la *manipulation formelle* des connaissances représentées. Dans le cas particulier d'une base de connaissances taxinomiques, le formalisme de représentation de connaissances utilisé doit permettre des opérations formelles qui modélisent l'activité d'identification, spécifique aux taxinomies.

Pour identifier un spécimen, les systématiciens procèdent par comparaison de la description du spécimen avec celles des taxons et plus précisément avec les parties discriminantes de ces descriptions. Ils s'appuient en effet sur les *clefs* d'identification que sont les critères classificatoires sur lesquels est fondée la taxinomie. Sur ce modèle, des opérations formelles d'identification sur une base de connaissances taxinomiques consistent en la manipulation des représentations des clefs de la taxinomie. Aussi pouvons-nous préciser la tâche d'explicitation à laquelle nous nous intéressons. Étant donnée la structure taxinomique d'un domaine, nous nous appliquons à expliciter les primitives de langage nécessaires à la définition des taxons et qui permettent de les discriminer. En d'autres termes, les descriptions des taxons doivent être élaborées *en fonction d'un but*, qui est l'identification. Il s'agit d'expliciter les clefs de la taxinomie en faisant apparaître dans la description des taxons de même rang les caractéristiques qui les distinguent. Ces clefs taxinomiques doivent être manipulables, représentées dans la base de connaissances. Elles sont les “briques” des procédures d'accès au domaine.

1.1.2.3. Le support d'un accès hypermédia au domaine

Des connaissances formalisées, si elles constituent un moyen appréciable de modélisation d'un domaine et de manipulation de ses concepts, ne conviennent cependant pas à l'appréhension de tous les aspects d'un domaine. En particulier en sciences naturelles, elles permettent de mettre en place des modes d'accès au domaine sous la forme de procédures d'identification mais elles ne remplacent pas des descriptions textuelles et graphiques des objets. C'est pourquoi nous proposons de faire collaborer base de connaissances et hypermédia au sein d'un même système. Cet aspect de notre travail est développé dans le chapitre 4 de cette thèse. Nous nous limitons ici à mettre en évidence les caractéristiques requises par le formalisme de représentation d'une base de connaissances associée à un hypermédia.

Nous avons précédemment envisagé une taxinomie comme un système d'index selon lequel organiser des connaissances. Ces connaissances taxinomiques peuvent à leur tour être utilisées pour organiser un ensemble de documents multimédias. Plus précisément, notre intention est d'indexer différents médias d'information aux éléments d'une telle base de connaissances : à la taxinomie, aux taxons, à leurs descriptions, aux concepts qui participent à ces descriptions et plus généralement à l'ensemble des concepts du domaine modélisé. De la sorte, un accès hypermédia au domaine sera basé sur l'organisation taxinomique de ce dernier. Cela nécessite en premier lieu que les connaissances représentées soient *descriptives* afin d'associer la représentation textuelle ou graphique d'un concept à sa représentation formelle. De manière corollaire, cette entreprise requiert des connaissances *déclaratives*, modulaires et *structurées* qui assurent une indexation habile et précise des différents médias aux éléments de la base de connaissances.

1.1.2.4. Conclusion

De l'examen des fonctionnalités d'une base de connaissances taxinomiques, nous retenons certaines exigences imposées à leur formalisme de représentation que nous résumons ici. (1) Des connaissances taxinomiques constituent la partie terminologique de la base de connaissances, relative aux concepts du domaine. (2) Elles doivent être représentées de façon déclarative. (3) De manière générale, les capacités de manipulation formelle des connaissances doivent être aussi étendues que possible, pour offrir un accès aux connaissances du domaine aussi large que possible. (4) La manipulation des connaissances doit permettre l'explicitation et la représentation des clefs taxinomiques, destinées à l'identification des objets du domaine.

1.2. Choix d'un formalisme de représentation

Dans cette partie, nous nous intéressons au choix d'un formalisme pour la représentation de connaissances taxinomiques. Nous traduisons les caractéristiques d'une base de connaissances taxinomiques que nous avons dégagées dans la partie précédente en termes de contraintes qu'elles nous imposent dans le choix d'un formalisme de représentation (§ 1.2.1.). Nous discutons alors de l'adéquation du modèle des graphes conceptuels à la représentation de connaissances taxinomiques (§ 1.2.2.).

1.2.1. Qualités requises

1.2.1.1. Intelligibilité

L'intelligibilité d'une base de connaissances est un critère important d'évaluation du formalisme de représentation de connaissances utilisé. Cette qualité est ici essentielle pour que la base de connaissances taxinomiques formalisées offre un modèle du domaine accessible. Cette base est en effet consultée comme mémoire de la taxinomie et des descriptions des taxons, comme zone de travail lors de l'explicitation de la terminologie du domaine ou de l'indexation d'un réseau hypermédia, puis de mises à jour.

Les caractéristiques d'un formalisme de représentation de connaissances qui participent à son intelligibilité consistent en des connaissances déclaratives, c'est-à-dire distinctes des mécanismes d'exploitation, en une séparation de la partie terminologique de la base de connaissances de celle asertionnelle relative aux objets et plus généralement une modularité des éléments de la base de connaissances formalisées.

1.2.1.2. Pouvoir de représentation

Le pouvoir de représentation du formalisme vient en réponse à la complexité des connaissances à représenter. Plus précisément, le formalisme choisi doit d'une part permettre de représenter des bases de connaissances de grande taille. Nous avons en effet souligné la taille importante des taxinomies et donc le nombre élevé de descriptions de taxons à représenter, mais aussi les tailles importantes de chacune des descriptions et de la terminologie employée, riche et nuancée. D'autre part, la qualité *compliquée* de la terminologie du domaine et des descriptions structurées de taxons doit également être prise en compte dans le choix d'un formalisme de représentation approprié.

1.2.1.3. Efficacité de raisonnement

L'efficacité de raisonnement du formalisme est déterminante pour l’exploitation des connaissances représentées. L’exploitation de bases de connaissances taxinomiques fait appel en effet à des manipulations formelles en grand nombre. Qui plus est, ces manipulations des connaissances sont particulièrement complexes. Il s’agit essentiellement de comparer les descriptions (nombreuses et complexes) des taxons, pour soit rendre compte de la structure taxinomique du domaine représenté, soit comparer différentes structures possibles, soit expliciter des clefs d’identification, soit encore identifier un objet et plus généralement répondre aux requêtes des utilisateurs de la base de connaissances.

1.2.2. Adéquation du formalisme des graphes conceptuels

Sur la base des qualités requises à un formalisme de représentation de connaissances taxinomiques que nous venons de dégager, nous montrons dans cette partie l’adéquation du formalisme des graphes conceptuels à cette tâche de représentation. Plus précisément, nous discutons l’intelligibilité des graphes conceptuels, leur puissance de représentation et de raisonnement et soulignons l’intérêt de l’existence d’une communauté scientifique active autour de ce formalisme de représentation.

1.2.2.1. Intelligibilité des graphes conceptuels

Les graphes conceptuels sont des graphes bipartis qui expriment des relations entre les concepts du domaine représenté (cf. § 1.3.1.). Ils sont étiquetés par des items lexicaux. Ce formalisme offre ainsi un mode de *représentation graphique* qui permet une appréhension *visuelle* des connaissances formalisées. La lecture, l’interprétation et la manipulation des éléments du formalisme sont de ce fait aisées et c’est à la fois la représentation et le raisonnement sur cette représentation qui sont intelligibles pour l'homme.

L’intelligibilité de la représentation des connaissances est essentielle à la réalisation de la base de connaissances, sa mise à jour, sa consultation et son exploitation. Ces activités sont en effet accomplies par des utilisateurs non nécessairement informaticiens.

L’intelligibilité des raisonnements effectués sur les connaissances formalisées permet de suivre pas à pas et visuellement ces raisonnements. Ils peuvent donc être contrôlés, vérifiés, voire expliqués à un non expert.

1.2.2.2. Puissance de représentation des graphes conceptuels

Les graphes conceptuels sont un modèle *déclaratif* de représentation de connaissances. Il est muni d'une *sémantique logique* consistante et complète. A tout graphe conceptuel G, on peut associer une formule de la logique du premier ordre, notée $\Phi(G)$ [Sowa, 84]. Le lecteur intéressé trouvera dans [Chein et Mugnier, 92] les preuves de la complétude et de la consistance de Φ. $\Phi(G)$ est une formule conjonctive, positive, ayant pour termes des variables ou des constantes, et fermée existentiellement [Mugnier et Chein, 96].

1.2.2.3. Puissance de raisonnement des graphes conceptuels

Les raisonnements sur les graphes conceptuels sont non pas des déductions logiques sur des représentations graphiques de formules logiques mais des manipulations directes de graphes [Mugnier et Chein, 96]. En d'autres termes, les calculs d'inférences s'appuient sur des *algorithmes* de graphes. Le formalisme des Graphes Conceptuels profite donc des algorithmes classiques en *Théorie des Graphes* et plus généralement des travaux en algorithmique combinatoire.

1.2.2.4. Communauté scientifique autour des Graphes Conceptuels

Outre les motivations propres à la spécificité de notre tâche de représentation, l'activité scientifique “au service” du formalisme des graphes conceptuels est à prendre en considération dans le choix de celui-ci. Nous avons précédemment évoqué les résultats algorithmiques du domaine de la Théorie des Graphes dont il bénéficie. Les utilisateurs du formalisme profitent également des logiciels existants, tels que les éditeurs de graphes. La demande en ingénierie des connaissances et en linguistique informatique sont le moteur de nombreuses recherches sur le formalisme. Ces recherches s'adressent à une communauté scientifique active, que réunit annuellement la conférence ICCS, International Conference on Conceptual Structures. La recherche française tient une place importante dans la communauté internationale comme en témoignent le volume 10 de la Revue d'Intelligence Artificielle (1996) ainsi que les thèmes de recherche de l'équipe du LIRMM à l'université Montpellier II, du projet ACACIA à l'INRIA Sophia-Antipolis. Parmi les travaux en ingénierie des connaissances qui s'appuient sur les graphes conceptuels comme formalisme de représentation, citons plus particulièrement [Aïmeur et Ganascia, 93] et [Bournaud et Ganascia, 95] qui ont traité, avant nous, dans l'équipe ACASA du LIP6, de ce que nous appelons *connaissances taxinomiques*. Dans leurs thèses, Bournaud [Bournaud, 96] et Aïmeur [Aïmeur, 94] s'intéressent respectivement à l'organisation et à l'explicitation de connaissances.

1.3. Formalisme des graphes conceptuels

Cette partie est dédiée à la présentation du formalisme des graphes conceptuels. Nous montrerons dans la partie suivante (§ 1.4.) son adéquation à la représentation de connaissances taxinomiques sur l'exemple d'une base de connaissances exprimée dans ce formalisme pour le domaine de l'Orchidologie.

Le formalisme des graphes conceptuels a été proposé par Sowa en 1984 [Sowa, 84]. Depuis, les fondements logiques du formalisme et la définition formelle des notions fondamentales ont fait l'objet de nombreux autres travaux et des extensions ont été proposées au formalisme initial. La version du formalisme des graphes conceptuels que nous exposons dans cette partie est fondée sur les travaux de Chein et Mugnier. Il s'agit du *modèle des Graphes Simples* dont le lecteur intéressé trouvera un exposé plus détaillé dans [Chein et Mugnier, 92] et [Mugnier et Chein, 96].

Nous commençons par présenter la notion de *graphe conceptuel simple* (§ 1.3.1.) et en donnons la définition formelle. Élément essentiel du formalisme, un graphe conceptuel simple permet de représenter des connaissances de diverses natures : entités, attributs, états, événements. La notion de *support* fait l'objet de la partie suivante (§ 1.3.2). Un support renferme le vocabulaire du modèle avec lequel représenter des connaissances, c'est-à-dire construire des graphes conceptuels. Dans le modèle des graphes simples, les éléments du support sont atomiques. Pour pouvoir en définir certains, nous adoptons une extension du modèle des graphes simples, proposé dans [Chein et Leclère, 94][Leclère, 95], qui consiste précisément en un mécanisme de définition des éléments du support (§ 1.3.3.). Enfin, nous détaillons les opérations définies dans le modèle qui permettent de raisonner sur les connaissances représentées (§ 1.3.4.).

1.3.1. Graphe conceptuel

Intuitivement, un graphe se dessine comme un schéma constitué d'un ensemble de points reliés par des flèches. Les points sont les sommets du graphe et les flèches les arcs du graphe. Plus formellement, la définition d'un graphe est la suivante [Berge, 70] :

Définition 1.1. [Graphe] *: Un graphe* $G=(X, U)$ *est l'ensemble constitué par*

- un ensemble de noeuds $X=\{x_1, \dots x_n\}$*,*

- un ensemble d'arcs $U=\{u_1, \dots u_n\}$ *d'éléments du produit cartésien* X¥X*.*

Un graphe conceptuel est un graphe fini et biparti, qui possède des noeuds *concepts* et des noeuds *relations conceptuelles* - les relations entre ces concepts. Un graphe conceptuel est étiqueté par des items lexicaux, ce qui présente l'avantage immédiat d'intelligibilité des connaissances représentées. En effet, un graphe peut être appréhendé visuellement grâce à sa présentation graphique conventionnelle. Plus précisément, il est représenté graphiquement par ses items lexicaux reliés par des flèches qui figurent les arcs du graphe ; ces items sont encadrés ou encerclés selon qu'ils sont les étiquettes de concepts ou des relations conceptuelles.

Le graphe de la figure suivante est un exemple de graphe conceptuel. Intuitivement, il représente des informations sur la chatte Tartine que cherche Annie dans la maison alors qu'elle est dans le jardin derrière la maison, à la recherche d'une souris. Les noeuds [`chat:Tartine`], [`humain:Annie`], [`jardin:`] [`maison:`] et [`souris:`] sont des concepts tandis que (`cherche`), (`guette`), (`dans`) et (`devant`) sont des relations entre ces concepts.

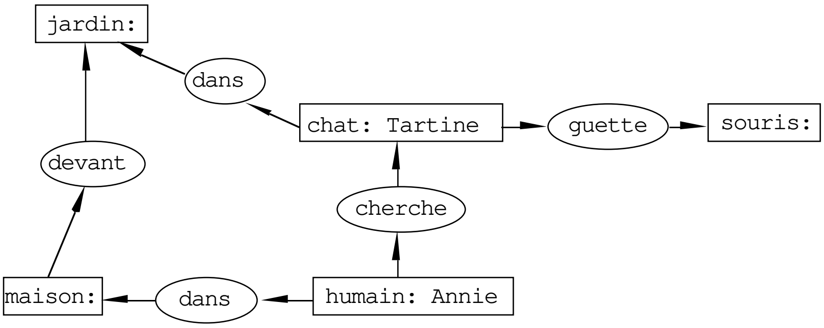


figure 1.2. Un exemple de graphe conceptuel

Nous adoptons comme définition formelle d'un graphe conceptuel celle établie dans [Mugnier et Chein, 96].

Définition 1.2. [Graphe conceptuel] : *Un graphe conceptuel* `G` *est un quadruplet* `(C, R, U, étiq)` *constitué par :*

- *l'ensemble* `C` *non vide des noeuds concepts de* `G`*,*
- *l'ensemble* `R` *des noeuds relations conceptuelles de* `G`*,*
- *l'ensemble* `U` *des arêtes de* `G`*,*
- *la fonction* `etiq` *qui associe à chaque noeud de* `G` *un item lexical pour étiquette.*

Chaque noeud d'un graphe conceptuel `G` est étiqueté. L'étiquette `étiq(r)` d'un noeud relation `r` est un *type* qui précise la nature du lien entre les noeuds concepts connectés par ce noeud relation. Par exemple, les noeuds concepts [`chat:Tartine`] et [`jardin:`] du graphe de la figure 1.2. sont connectés par le noeud relation conceptuelle (dans) d'étiquette est-dans. T_r est l'ensemble des types de relations disponibles pour construire un graphe conceptuel G.

L'étiquette étiq(c) d'un noeud concept c d'un graphe conceptuel G est un couple composé d'un *type* et d'un *référent* qui précisent respectivement la nature et l'identité de l'entité que représente ce noeud concept. Par exemple, le noeud concept [chat:Tartine] a pour type chat et pour référent Tartine. Tc est l'ensembles des types de concepts disponibles pour construire un graphe conceptuel G. I est l'ensemble des *marqueurs individuels*, utilisés comme référents des noeuds concepts. Dans notre exemple, Tartine et Annie sont des marqueurs individuels. Le symbole * représente le *référent générique* et est utilisé lorsque l'identité exacte d'un noeud concept est inconnue. Par convention, lorsque le référent d'un concept est générique, il peut ne pas être dessiné. C'est par exemple le cas du concept [souris:] sur la figure 1.2.

Pour chaque noeud relation r de G, l'ensemble des arêtes qui lui sont adjacentes est totalement ordonné de 1 à *degré*(r), où *degré*(r) désigne le nombre total d'arêtes adjacentes à r. Ainsi, un noeud concept c est le $i^{ème}$ voisin d'un noeud relation r de G s'il existe une arête rc étiquetée par *i*. Dans le cas particulier de relations binaires, étiqueter les arêtes du graphe n'est pas nécessaire. Pour distinguer les deux voisins d'un noeud relation, il suffit en effet d'utiliser des flèches pour figurer les arcs. C'est par exemple le cas des arcs du graphe de la figure 1.2.

Un graphe conceptuel n'a de sens que par rapport à un domaine ; il est en effet construit sur un *support* qui définit le vocabulaire du domaine. T_C, T_r et $I \cup \{*\}$ en sont les éléments essentiels.

1.3.2. Support et canonicité

La représentation d'un domaine passe par la représentation de sa terminologie, avant même celle de ses objets. Ces derniers seront en effet décrits à l'aide des éléments de la terminologie du domaine. Dans le modèle des graphes simples, la terminologie d'un domaine est représentée par un *support* et les graphes conceptuels qui représentent les objets de ce domaine sont construits à l'aide des éléments de ce support.

1.3.2.1. Support

Définition 1.3. [Support] : *Un support S est un quintuplet* $(T_c, T_r, I, \tau, \sigma)$ *où :*

- T_c *est l'ensemble partiellement ordonné des types de concepts, possédant un plus grand et un plus petit élément (type universel noté* $\top$ *et type absurde noté* $\bot$*) ;*
- T_r *est l'ensemble des types de relations, partitionné en ensembles de types de relations de même arité admettant chacun un plus grand et un plus petit élément ;*
- σ *associe à tout type de relation le type maximal que peuvent prendre chacun de ses arguments ;*
- I *est l'ensemble des marqueurs individuels ; ils sont deux à deux incomparables et tous plus petits que le référent générique noté *,*
- $\tau : I \rightarrow T_c\backslash\{\bot\}$ *associe à tout marqueur individuel un type de concept dit conforme.*

Par exemple, le support de la figure 1.3. ci-dessous représente une terminologie qui permet de construire le graphe conceptuel de la figure 1.2.

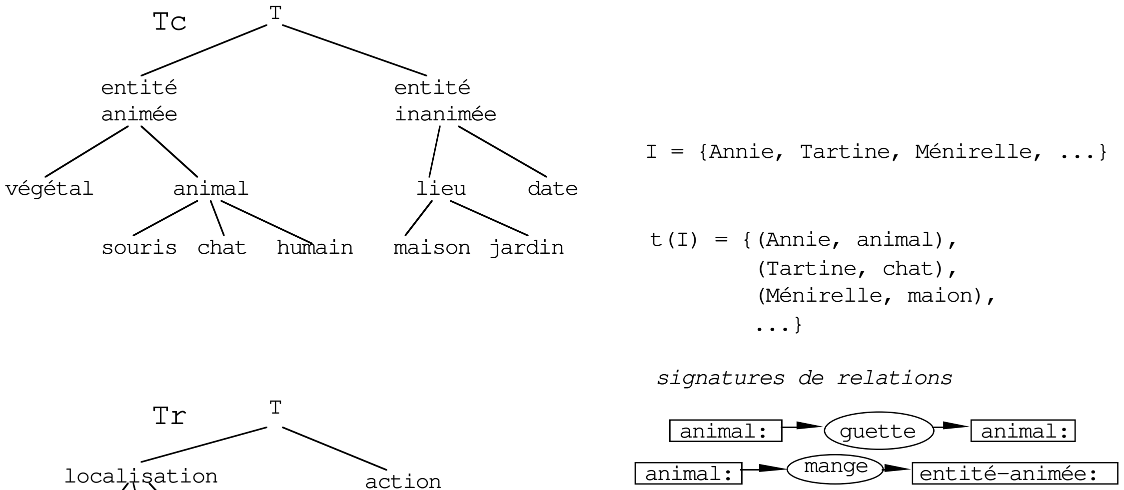


figure 1.3. Représentation d'une terminologie par un support

T_c et T_r sont tous deux ordonnés par des relations qui peuvent être interprétées comme des liens *sorte de*. Par exemple, `souris` est une *sorte d'*`animal`, `est-dans` une *sorte de* `localisation`.

Dans le modèle des graphes simples, T_c et T_r sont des ensembles de types de concepts et de relation *atomiques*, c'est-à-dire non définis [Mugnier et Chein, 96]. Le modèle initial proposé par Sowa comporte un mécanisme de *définition de types de concepts* [Sowa, 84]. A partir de celui-ci, Leclère propose une extension du modèle des graphes simples pour permettre de définir des types de concepts ou de relations [Leclère, 95]. La partie 1.3.3. est consacrée à ce mécanisme de définition de types structurés à partir des types existants.

Les types de relations sont munis de signatures à l'aide de l'application σ. Une signature impose des restrictions quant aux noeuds concepts voisins d'un noeud relation conceptuelle dans un graphe conceptuel. Plus précisément, elle impose les types dont les types de ces concepts doivent être des sous-types, eu égard à la relation d'ordre définie sur T_C. On représente la signature d'un type de relation par un graphe *étoile* [Chein et Mugnier, 92]. Il possède un unique noeud relation conceptuelle, de type celui dont on veut représenter la signature ; ses noeuds concepts ont précisément pour types ceux les plus généraux qu'autorise la signature (cf. figure 1.3.). Dans notre exemple, la signature du type de relation `mange` impose d'une part que le mangeur soit un animal et d'autre part que cet animal mange un animal ou un végétal ; elle empêche ainsi par exemple qu'un jardin ne mange une maison. En d'autres termes, les signatures permettent d'imposer des restrictions syntaxiques et sémantiques dans l'usage d'un support, c'est-à-dire dans la construction de graphes conceptuels. L'ensemble des signatures des types de relations d'un support constitue la *base* de ce support [Mugnier et Chein, 93].

Les marqueurs individuels sont typés à l'aide de l'application τ : à tout individu `m` est associé le plus petit type de concept `t` tel que `m` *est un* `t`. Un marqueur individuel `m` est dit conforme à un type de concept `t` si et seulement si $\tau(m) \leq t$. Dans notre exemple, τ(`Tartine`) est égal au type `chat` (cf. figure 1.3.) ; le référent `Tartine` est donc conforme au type `animal` et non au type `maison`. En effet, `chat` est un sous-type de `animal` et non de `maison`. La notion de conformité permet donc d'introduire des restrictions sémantiques dans l'usage du support et plus précisément de contrôler la relation d'instanciation entre marqueurs individuels et types de concepts.

1.3.2.2. Base canonique

La notion de *canonicité* est introduite dans [Sowa, 84] pour assurer de la correction sémantique des graphes conceptuels construits. Une base *canonique* est un ensemble noyau de graphes conceptuels déclarés canoniques par le concepteur de la base et utilisés pour construire des graphes conceptuels dits canoniques, c'est-à-dire doués de sens, à l'aide des opérations élémentaires de spécialisation (cf. § 1.3.4.). Plus formellement :

Définition 1.4. [Base canonique] : *Une* base canonique B *associée à un support* S *est un ensemble fini de graphes conceptuels simples bien formés relativement à* S*, c'est-à-dire construits à partir des éléments de* S *et en respectant les signatures des types de relations.*

Une base canonique permet d'imposer des contraintes minimales pour la construction d'une base de connaissances dans le modèle des graphes conceptuels. Ces contraintes sous forme de graphes canoniques peuvent être "syntaxiques" ; c'est le cas des graphes conceptuels qui représentent les signatures des types de relation. La plus petite base canonique d'un support est de fait sa base de graphes étoiles [Mugnier et Chein, 93]. Les contraintes que représentent une base canonique peuvent également être liées au domaine, c'est-à-dire "sémantiques" ; c'est par exemple le cas d'un graphe représentant la description générique d'une orchidée utilisé comme générateur des graphes décrivant les différentes familles de la taxinomie des orchidées (cf. § 1.4.).

1.3.3. Définition de types de concepts

Le mécanisme de définition de types de concepts dans le formalisme des graphes conceptuels est basé sur la notion d'*abstraction* [Sowa, 84].

Définition 1.5. [Abstraction] : *Une* abstraction *n-aire* $\lambda x_1, \ldots x_n G$ *consiste en un graphe* G*, appelé* corps*, et une liste de n noeuds concepts génériques de* G*, appelés* paramètres formels*. Les n variables* $x_1, \ldots x_n$ *distinguent les paramètres formels des autres noeuds concepts de* G*. Ils figurent dans le champ référent des noeuds.*

Définition 1.6. [Définition de type, Genre et Différence] : *Un type de concept* t *est défini par une abstraction unaire* $\lambda x G$*. On note :* `type t(x) is G`

G *est appelé la* différence *de* t*. La variable* x *distingue le noeud concept (unique) paramètre formel du graphe* G*. Ce noeud est appelé la* tête *de* t *et le type de ce noeud le* genre *de* t*.*

Dans le modèle des graphes simples, étant donné un support S, Leclère propose, partant de [Sowa, 84], de définir un type de concept t à l'aide d'un type de concept t' de S, déclaré être le *genre* de t, et d'un graphe conceptuel G construit à partir des éléments de S et dont un noeud concept a pour étiquette (t',*). Ce graphe exprime en quoi le type t nouvellement défini diffère de son genre t', c'est la *différence* de t [Leclère, 95].

Considérons par exemple le graphe G de la figure 1.4. Nous pouvons définir à l'aide de ce graphe un nouveau type de concept `chat_dom` qui représente un chat domestique.

En considérant que `G` exprime la différence entre les types `chat_dom` et `chat`, nous marquons son sommet [`chat:`] par la variable `x`. `chat` est ainsi le genre de `chat_dom`. Nous définissons alors `chat_dom` par l'abstraction unaire `type chat_dom(x) is G`.

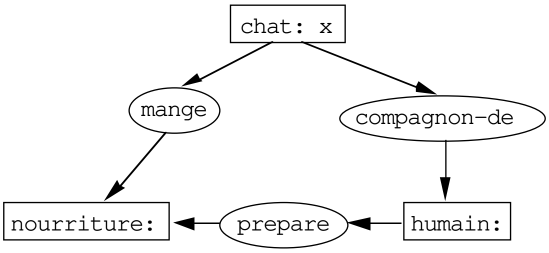


figure 1.4. Définition d'un type de concept qui représente un chat domestique

La relation d'ordre sur l'ensemble des types de concepts n'est plus alors simplement posée par le constructeur du support. Entre deux types dont l'un au moins est défini, elle est calculable. Il s'agit plus précisément d'une relation de spécialisation entre types, fondée sur la relation de spécialisation entre graphes. Elle est décrite dans la partie suivante.

Nous ne nous intéresserons pas ici aux possibilités de définition des types de relations car nous ne les utilisons pas dans notre travail. Le lecteur intéressé trouvera l'exposé d'un tel mécanisme dans [Leclère, 95].

1.3.4. Relation de spécialisation/généralisation

La notion fondamentale dans le modèle des graphes conceptuels est la relation de généralisation/spécialisation, notée ≤, définie sur l'ensemble des graphes construits sur un même support. L'interprétation logique de la relation de généralisation est l'implication. Elle permet en effet de déterminer si les connaissances représentées par un graphe sont la conséquence logique de celles représentées par un autre. La relation de généralisation est donc la base de tout raisonnement avec des graphes conceptuels.

La relation ≤ est comprise indifféremment en termes de spécialisation ou de généralisation. Plus formellement,

Propriété 1.1. : *Étant donnés deux graphes conceptuels G_1 et G_2 définis sur un même support, G_1 est une spécialisation de G_2, noté $G_1 \leq G_2$, si et seulement si G_2 est une généralisation de G_1, également noté $G_1 \leq G_2$.*

Notons que la relation de généralisation est aussi appelée *subsomption* et que dans certains domaines tels qu'en apprentissage symbolique la notation est inversée. On note en effet $a \leq b$ le fait que `a` soit une généralisation de `b` (`a` subsume `b`).

La relation de spécialisation peut être définie en termes de *projection*. Une projection est un morphisme de graphes, qui conserve les arêtes et autorise la restriction des étiquettes des sommets. De façon équivalente, la relation de spécialisation peut également être définie en termes d'une séquence (finie) d'opérations élémentaires de spécialisation. L'équivalence entre ces deux définitions est montrée dans [Mugnier et Chein, 96]. Nous adoptons ici une définition en termes d'opérations élémentaires de spécialisation et nous renvoyons le lecteur intéressé par une présentation de la projection de graphes à [Mugnier et Chein, 96]. La définition "constructive" de la relation de spécialisation en termes d'opérations est en effet la mieux appropriée à notre travail. Nous nous intéressons en effet à la *construction* de graphes par spécialisation d'autres graphes (cf. § 1.4.2.). La définition de la relation de spécialisation en termes de projection s'adresse davantage à la vérification ou la recherche de relations de spécialisation entre graphes.

1.3.4.1. Opérations élémentaires de spécialisation

Nous adoptons ici les opérations de spécialisation définies dans le modèle des graphes conceptuels simples [Mugnier et Chein, 96]. Il s'agit des opérations unaires de simplification, restriction de relation, restriction de concept et joint interne et des opérations binaires de somme disjointe et joint externe.

Simplification :

Suppression d'un noeud relation redondant d'un graphe `G`, c'est-à-dire un noeud ayant les mêmes voisins qu'un autre noeud de même type et dans le même ordre.

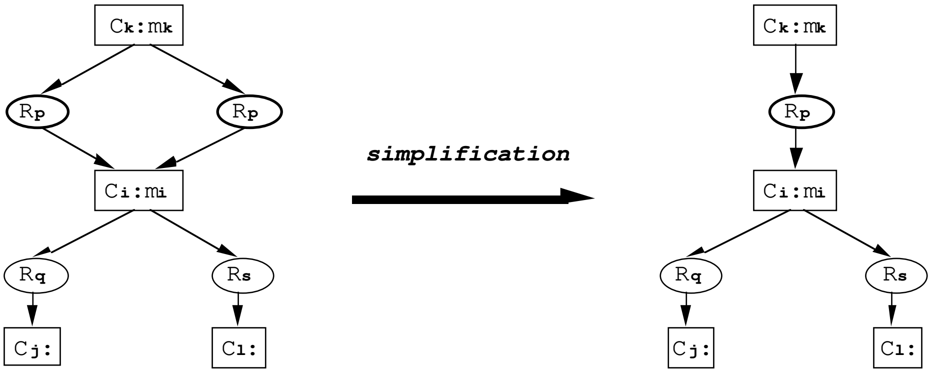


figure 1.7. Simplification

Restriction de relation : Diminution de l'étiquette d'un noeud relation `r` d'un graphe `G`, c'est-à-dire remplacement du type de `r` par un sous-type dans T_r.

Restriction de concept : Diminution de l'étiquette d'un noeud concept `c` d'un graphe `G`, c'est-à-dire remplacement du type de `c` par un sous-type dans T_C et/ou, si le marqueur de `c` est générique, remplacement de celui-ci par un marqueur individuel.

Joint interne : Fusion de deux noeuds concepts d'un graphe `G` ayant même étiquette.

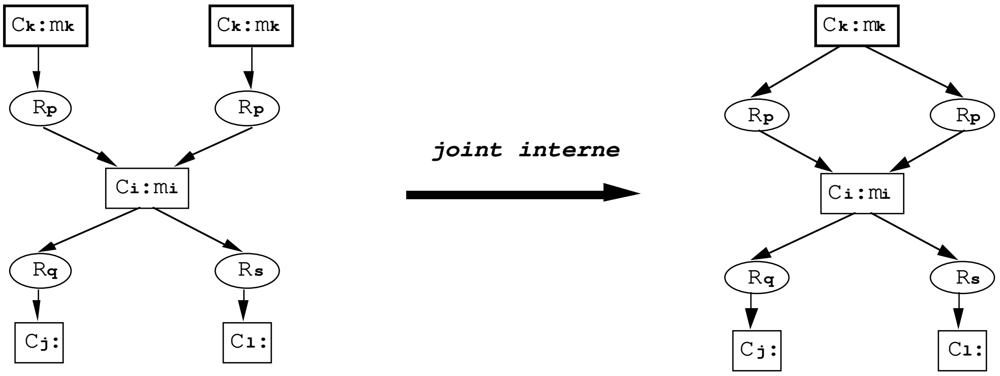


figure 1.6. Joint interne

Somme : somme disjointe de deux graphes G_1 et G_2, le résultat étant donc un graphe non connexe.

Joint externe : Fusion entre un noeud concept c_1 d'un graphe G_1 et un noeud concept c_2 d'un graphe G_2 ayant même étiquette.

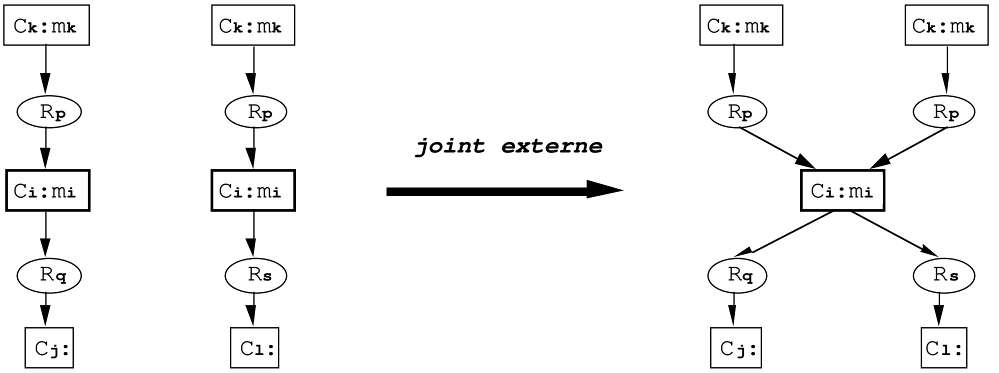


figure 1.5. Joint externe

Dans le cas général, cette opération n'est pas élémentaire mais la composée d'une somme et d'un joint interne. Cependant, dans le cas particulier de graphes connexes qui pourra être le nôtre (cf. § 1.4.2.), l'opération de somme n'est pas considérée puisque son résultat est un graphe non connexe et l'opération de joint externe devient élémentaire.

Des *opérations de généralisation* sont définies dans le modèle des graphes simples comme les opérations inverses de celles de spécialisation : duplication, augmentation, éclatement, décomposition. Nous ne les présentons pas ici car nous ne les utilisons pas dans notre travail. Nous renvoyons le lecteur intéressé à [Mugnier et Chein, 96].

1.3.4.2. Relation de spécialisation entre graphes conceptuels

A l'aide des opérations de spécialisation de graphes que nous venons de présenter, nous pouvons maintenant définir la relation de généralisation entre graphes construits sur un même support.

Définition 1.7. [Relation de généralisation/spécialisation entre graphes] : *Soient deux graphes conceptuels G_1 et G_2. G_1 est une* généralisation *de G_2 et G_2 est une* spécialisation *de G_1, noté $G_2 \leq G_1$, si et seulement si G_2 peut être obtenu à partir de G_1 par une séquence finie d'opérations élémentaires de spécialisation.*

Une définition plus formelle, proposée dans [Mugnier et Chein, 96], est fondée sur la notion de *dérivation* dont la définition est la suivante :

Définition 1.8. [Dérivation] : *Une* dérivation *d'un graphe conceptuel simple G est un graphe fini, orienté, sans circuit, dont les sommets sont étiquetés par des graphes conceptuels simples et tel que :*

- il a un unique noeud qui ne soit l'extrémité finale d'aucun arc ; ce noeud est étiqueté par G ;

- tout autre sommet soit est une source, soit a un (ou deux) prédécesseurs à partir duquel (desquels) il peut être obtenu par une opération élémentaire de spécialisation unaire (ou binaire, c'est-à-dire l'opération de joint externe).

Définition 1.9. [Relation de généralisation] : *Soient deux graphes conceptuels simples G_1 et G_2. $G_2 \leq G_1$ s'il existe une dérivation de G_2 contenant un sommet étiqueté par G_1.*

Nous pouvons maintenant introduire la notion de *canonicité* d'un graphe telle qu'elle est définie dans [Chein et Mugnier, 93].

Définition 1.10. [Graphe canonique] : *Un graphe G est* canonique *relativement à une base canonique B s'il est* dérivable *à partir des graphes de B. L'ensemble des graphes canoniques relativement à une base B est la* fermeture *de B par les opérations élémentaires de spécialisation.*

1.3.4.3. Relation de spécialisation entre types de concept

Entre deux types de concepts atomiques, la relation d'ordre est établie par le constructeur du support ; entre deux types dont l'un au moins est défini, elle est calculable. Plus précisément, la relation d'ordre dont est muni l'ensemble des types de concepts d'un support est une relation de spécialisation induite par celle définie sur les graphes conceptuels [Leclère, 95].

Définition 1.11. [Relation de Généralisation entre types de concept] *: Soit* `t` *un type de concept défini par* `type t(x) is G` *et soit* t' *un type de concept appartenant au support utilisé pour la construction de* `G`*.*

- Si t' *est atomique, alors* $t \leq t'$ *ssi genre(*`t`*)* $\leq t'$ *;*

- Si t' *est défini par* `type t'(x) is G'`*, alors* $t \leq t'$ *ssi il existe une dérivation de* `G` *contenant* `G'` *telle que la tête de* `G'` *devienne celle de* `G`*.*

Cette définition de la relation de généralisation entre types dont l'un au moins est défini permet d'introduire les types de concepts définis dans l'ensemble T_C du support, initialement composé de types atomiques. Un protocole d'insertion automatique de types est défini dans [Chein et Leclère, 94] :

Protocole Insertion (t, S) :

(1) Calcul de l'ensemble `PGS` des Plus Grands Spécialisés de `t` dans `S`, c'est-à-dire recherche des types de concepts t' de `S` vérifiant $t' \leq t$ et $\neg \exists p$, $p \leq t$ et $t' \leq p$.

(2) Calcul de l'ensemble `PPG` des Plus Petits Généralisés de `t` dans `S`, c'est-à-dire recherche des types de concepts t' de S vérifiant $t \leq t'$ et $\neg \exists p$, $t \leq p$ et $p \leq t'$.

(3) Ajout des relations $s \leq t$, pour tout $s \in$ `PGS` et des relations $t \leq g$, pour tout $g \in$ `PPG`.

(4) Vérification des propriétés de `S`.

Pour comparer deux types définis, on considère leurs descriptions *atomiques*, c'est-à-dire des graphes `G` et `G'` dont les noeuds ont tous des types atomiques. Si tel n'est pas le cas, on se ramène à une description atomique par une succession d'opérations d'expansion qui consistent à remplacer un concept dont le type est défini par la définition de celui-ci [Leclère, 95]. Plus formellement,

Définition 1.12. [Expansion] : *Soit* `G` *un graphe conceptuel dont le type* `t` *d'un noeud concept* `c` *est défini par* `type t(x) is D(x)`. *L'expansion de* `c` *dans* `G` *consiste à :*

- dans `G`*, remplacer* `t` *par son genre* `t'`*;*

- dans `D`*, remplacer le référent de la tête* `c'` *par celui de* `c` *dans* `G`*;*

- faire un joint entre `G` *et* `D` *sur les sommets* `c` *et* `c'`*.*

L'opération inverse est celle de contraction. Elle consiste à supprimer dans un graphe un sous-graphe correspondant à la définition d'un type et à le remplacer par un sommet concept étiqueté par ce type [Leclère, 95].

Définition 1.13. [Expansion] : *Soit* `G` *un graphe possédant un sous-graphe* `G'` *isomorphe à la définition* `D(x)` *d'un type de concept* `t` *et tel que* `G'` *est connecté au reste de* `G` *uniquement par un sommet concept correspondant à la tête* `c` *de* `D(x)`. *La contraction de* `G'` *dans* `G` *consiste à :*

- remplacer le type de `c` *par* `t,`

- supprimer le sous-graphe `G' \ [c]` *dans* `G`*.*

Les éléments du formalisme des graphes conceptuels nécessaires à la compréhension de notre travail étant présentés, nous développons dans la partie suivante un exemple de formalisation d'une base de connaissances taxinomiques à l'aide de graphes conceptuels. Cet exemple constitue une illustration de l'utilisation du formalisme des graphes conceptuels que nous venons de présenter et témoigne de l'adéquation de ce formalisme à la représentation de connaissances taxinomiques. En outre, nous précisons sur cet exemple quelles restrictions simplificatrices peuvent être imposées dans l'usage du modèle des graphes conceptuels pour cette tâche particulière de représentation de connaissances taxinomiques.

1.4. Exemple d’une base de connaissances taxinomiques représentée dans le formalisme des graphes conceptuels

Dans cette partie, nous présentons un extrait d’une base de connaissances formalisées dans le modèle des graphes conceptuels destinée à représenter la taxinomie des orchidées décrite par l'orchidéiste Lecoufle [Lecoufle, 81]. Dans un premier temps, nous décrivons les éléments de la base de connaissances en montrant l’adéquation des graphes conceptuels aux caractéristiques d'une taxinomie énoncées au § 1.1. de ce chapitre. Nous expliquons ensuite la construction d'une telle base de connaissances.

1.4.1. Description des éléments de la base de connaissances

Les orchidées sont des plantes *phanérogames monocotylédones* ; elles forment une famille dans l’ordre des *Orchidales*, la famille des *Orchidaceae*. Dans la classification proposée dans [Lecoufle, 81], la famille des Orchidaceae se divise en sous-familles, tribus et sous-tribus ; les sous-tribus se divisent encore en genres qui se divisent eux-mêmes en espèces. Sur cet exemple de la taxinomie des orchidées, nous montrons dans cette partie comment représenter dans le modèle des graphes conceptuels d'une part chaque taxon, quel que soit son niveau taxinomique, et d'autre part la structure de l'ensemble des taxons, c'est-à-dire la taxinomie. Nous donnons également une représentation du langage de description des taxons et des clefs d'identification sur lesquelles se fondent la détermination des espèces qui est une fonctionnalité essentielle pour une base de connaissances taxinomiques, pour les systématiciens.

1.4.1.1. Représentation d'un taxon

Rappelons qu'en sciences naturelles un organisme est décrit selon une logique de composition qui s'appuie sur la structure de l'organisme à décrire (cf. § 1.1.1.4.). Plus précisément, un organisme est décrit par la structure hiérarchique de ses parties et sous-parties et les caractéristiques de celles-ci. Par exemple, dans la description d'une orchidée, on distingue sa partie aérienne de sa partie souterraine. Une caractéristique de la partie aérienne est sa hauteur ; elle est composée d'une tige, de feuilles, de bractées et d'une inflorescence. La fleur est elle-même composée d'un labelle, d'un périanthe et d'un éperon, etc. La taille, la forme, la couleur sont les caractéristiques usuelles de ces différentes parties.

A partir de ces considérations anatomiques, nous pouvons préciser quels graphes conceptuels sont adéquats à la représentation des descriptions associées aux taxons. Tout d'abord, eu égard à cette logique de composition adoptée dans les descriptions de taxons, nous pouvons ne considérer que des arbres. Celui de la figure 1.8. ci-dessous en est un exemple ; il représente la description associée à la tribu des *Vandeae*. Les types de concepts et de relations utilisés dans les arbres de description des taxons constituent le support représentant le langage de description des orchidées. Celui-ci sera détaillé dans la partie 1.4.1.3. Les relations utilisées sont les relations *binaires* de types `composé_de`, `caractérisé_par` ou encore inférieurs à `caractérisé_par` tels que `de_texture`, `de_forme`, etc. Les relations de types `composé_de` sont établies entre deux concepts dont le type de l'un est inférieur au type `composé` et l'autre au type `composant`. Les relations de types inférieurs ou égaux à `caractérisé_par` sont établies entre deux concepts dont le type de l'un est inférieur à `composé` et celui de l'autre à `caractéristique`.

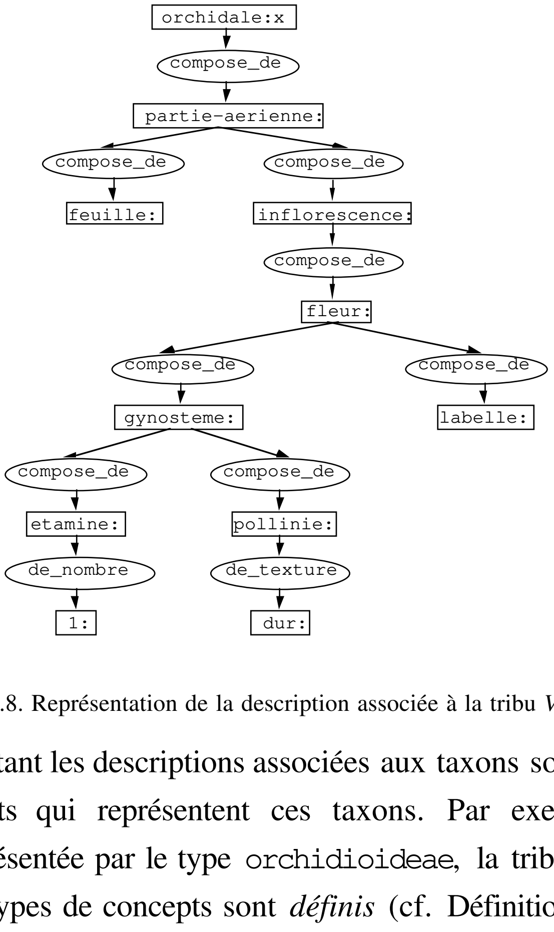


figure 1.8. Représentation de la description associée à la tribu *Vandeae*

Les arbres représentant les descriptions associées aux taxons sont utilisés pour définir des types de concepts qui représentent ces taxons. Par exemple, la sous-famille *Orchidioideae* est représentée par le type `orchidioideae`, la tribu *Vandeae* par le type `vandeae`, etc. et ces types de concepts sont *définis* (cf. Définition 1.5.). Par exemple, l'arbre `G` de la figure 1.8. ci-dessus permet de définir le type de concept `vandeae` qui représente la tribu *Vandeae* : `type vandeae(x) is G`.

L'arbre de définition d'un type de concept représentant un taxon peut être vu comme le résultat d'un *joint* entre des graphes de la forme $[C_i] \rightarrow (R_i) \rightarrow [C'_i]$ représentant chacun un trait descriptif. Nous appellerons de tels graphes des *arcs conceptuels* [Bournaud, 96]. La figure 1.9. ci-dessous présente l'ensemble des arcs conceptuels qui représentent les traits descriptifs de la tribu *Vandeae* ; le résultat d'une séquence de joints entre ces arcs serait le graphe de définition du type `vandeae` de la figure 1.8. Un trait descriptif désigne soit une caractéristique d'une partie de l'objet décrit, soit une relation de composition entre deux parties. Un trait dit “de caractérisation” est représenté, dans le modèle des graphes conceptuels, par un noeud relation dont le type est inférieur au type `caracterisé_par` et deux noeuds concepts dont le type de l'un est inférieur au type `composant` et celui de l'autre inférieur au type `caractéristique`. Un trait dit “de composition” est représenté par un noeud relation de type `composé_de` et deux noeuds concepts dont le type de l'un est inférieur au type `composé` et celui de l'autre inférieur au type `composant`.

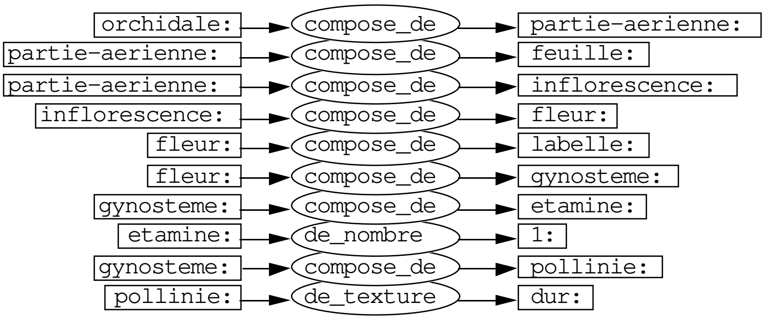


figure 1.9. Représentation des traits descriptifs de la tribu des *Vandeae*

Les relations qui interviennent dans les arbres conceptuels décrivant les taxons étant toutes binaires, la décomposition d'un tel arbre en un ensemble d'arcs conceptuels représentant chacun un trait descriptif est directe [Mineau, 90]. De plus, cette décomposition est sans perte d'information. Ce résultat a été établi dans le cas où les étiquettes des noeuds de l'arbre sont toutes distinctes [Bournaud, 96]. Il est encore valide dans le cas qui est le notre d'arbres dont les noeuds concepts de type inférieur ou égal à `caractéristique` peuvent avoir la même étiquette. En effet, ces noeuds sont des feuilles de l'arbre ; les noeuds internes, de types inférieurs à `composant` sont tous distincts (il suffit pour cela de désigner par des noms différents les parties différentes des objets décrits). En ne considérant que les joints entre arcs conceptuels sur des noeuds concepts de types inférieurs à COMPOSANT, on a bien une unique décomposition possible d'un arbre en un ensemble d'arcs conceptuels et une unique “recomposition” possible de cet arbre par une séquence de joints entre les arcs conceptuels. Cet isomorphisme entre un graphe conceptuel décrivant un taxon et l'ensemble des arcs conceptuels qui composent ce graphe permet de mettre en évidence la notion de “trait descriptif”. Nous reviendrons dans le chapitre 3 de cette thèse sur l'intérêt que présente d'un point de vue algorithmique une représentation de graphes sous forme d'ensembles d'arcs.

1.4.1.2. Représentation d'une taxinomie

Chaque taxon d'une taxinomie donnée étant représenté par un type de concept, la taxinomie, c'est-à-dire la structure de l'ensemble des taxons, est représentée par la structure de l'ensemble des types de concepts qui représentent ces taxons. Dans le modèle des graphes conceptuels, les types de concepts sont ordonnés par une relation de généralisation (cf. Définition 1.11. et Propriété 1.2.). La relation de généralisation entre types de concepts représentant des taxons constitue donc le niveau d'abstraction qui permet d'appréhender une taxinomie dans les termes du modèle des graphes conceptuels.

Plus précisément, un taxon de niveau taxinomique inférieur à un second taxon est représenté par un type de concept qui est une spécialisation du type de concept représentant le second taxon. Par exemple, comme cela apparaît sur la figure 1.10. ci-dessous, le type de concept vanda qui représente le genre *Vanda* est une spécialisation du type sarcanthineae qui représente la sous-tribu *Sarcanthineae*. Par transitivité de la relation de spécialisation, c'est aussi une spécialisation du type vandeae qui représente la tribu des *Vandeae* ou encore du type orchidioideae qui représente la sous-famille des *Orchidioideae*. orchidaceae, le type de concept le plus général qui représente la famille des *Orchidaceae* est une spécialisation du type orchidale qui représente l'ordre des *Orchidales*. Nous introduisons ce type comme étant atomique. Nous nous intéressons en effet à la taxinomie des orchidées et à ce titre seules nous intéressent les descriptions des taxons qui regroupent des spécimens d'orchidées, c'est-à-dire la famille des *Orchidaceae* et les taxons inférieurs.

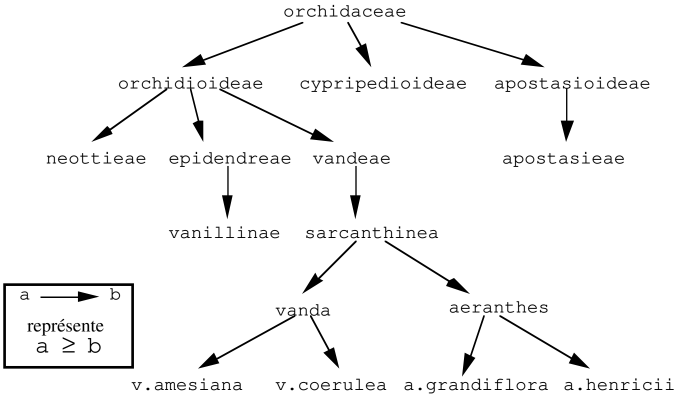


figure 1.10. Représentation d'un extrait de la taxinomie des orchidées

1.4.1.3. Représentation du langage de description

La terminologie nécessaire à la définition des concepts du domaine représenté - les taxons en orchidologie par exemple - apparaît dans le support sur lequel sont construits les graphes conceptuels qui définissent les types de concepts représentant les taxons (cf. définition 1.3.). Elle consiste en types de relations et types de concepts *atomiques*, qui constituent les *primitives de langage* nécessaires à la définition des types de concepts qui représentent les taxons. La figure 1.11. ci-dessous offre un extrait du support nécessaire à la représentation de la taxinomie des orchidées. $\mathtt{T}_{\mathtt{c}}$ et $\mathtt{T}_{\mathtt{r}}$ ont des structures d'*arbres*. Leurs types supérieurs ont été mis en évidence précédemment, lors de l'examen des graphes de définition des types de concepts représentant des taxons. Il s'agit des types de concepts `orchidale`, `composé`, `composant` et `caractéristique` et des types de relations `composé_de` et `caractérisé_par`. Les types de concepts définis représentant des taxons sont insérés au support sous le type atomique `orchidale`. Les signatures (cf. Définition 1.3.) des types de relations `composée` et `caractérisé_par` représentent des contraintes sur les noeuds concepts voisins de ces types ; nous les avons exprimées lors de l'examen des graphes de définition des types représentant des taxons. Elles imposent d'une part que les types des noeuds concepts voisins d'un noeud relation `composé_de` soient des sous-types l'un de `composé` et l'autre de `composant` et d'autre part que les types des noeuds concepts voisins d'un noeud relation `caractérisé_par` soient des sous-types l'un de `composé` et l'autre de `caractéristique`.

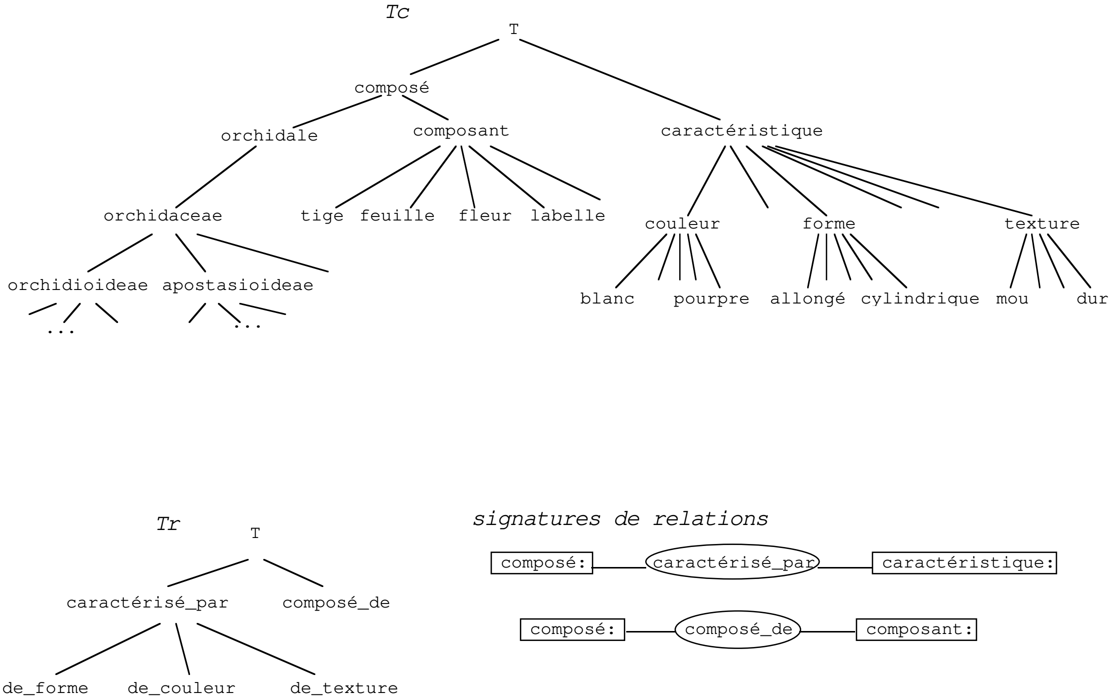


figure 1.11. Représentation du langage de description des orchidées

1.4.1.4. Représentation des clefs d'identification

Rappelons qu'une fonction essentielle d'une base de connaissances taxinomiques est de permettre la mise en place d'un système de détermination basé sur des clefs d'identification (cf. § 1.1.2.2.). Les clefs d'identification d'une taxinomie expriment les critères classificatoires sur lesquels celle-ci est fondée. Plus précisément, une clef est associée à chaque taxon qui n'est pas une feuille de la taxinomie, c'est-à-dire une espèce ; cette clef permet de distinguer les fils du taxon dans la taxinomie. En d'autres termes, une clef est un ensemble de traits descriptifs permettant de discriminer des taxons frères.

Une description de taxon consiste en une structure hiérarchique de parties et sous-parties et les caractéristiques de celles-ci. Un trait descriptif discriminant est donc soit une caractéristique qui diffère d'un taxon à l'autre, soit une sous-partie présente chez les uns et absente chez les autres. Par exemple, d'après le botaniste Hunt, dans la sous-famille des *Orchidioideae*, on peut distinguer les tribus des *Vandeae*, des *Neottieae* et des *Epidendreae* par la texture des pollinies de leurs fleurs respectivement dure, molle et demi-molle. Pour discriminer les sous-familles des *Orchidioideae*, des *Cypripedioideae* et des *Apostasioideae*, deux traits descriptifs sont nécessaires ; les *Orchidioideae* se distinguent des *Cypripedioideae* et des *Apostasioideae* par le nombre d'étamines de leurs fleurs ; elles n'en possèdent qu'une à la différence de celles-ci qui en possèdent deux. *Cypripedioideae* et *Apostasioideae* diffèrent quant à elles par la présence d'un staminode sur le gynostème de la fleur chez les premières et son absence chez les secondes [Lecoufle, 81].

Dans le modèle des graphes conceptuels, une clef est représentée par un graphe (non nécessairement connexe). Ses composantes connexes représentent chacune un trait descriptif discriminant, dont la conjonction constitue la clef à représenter. Le cas particulier d'une clef vide associée à un taxon possédant un unique fils peut être représenté par le graphe réduit à un noeud concept générique de type universel. Chaque taxon qui n'est pas une feuille de la taxinomie étant muni d'une clef permettant de discriminer ses fils, on peut construire un arbre isomorphe à la taxinomie privée de ses feuilles et dont les noeuds sont étiquetés par les graphes représentant les clefs d'identification. On appelle un tel arbre un *arbre de détermination*. La détermination d'une espèce peut alors être vue comme la recherche d'un chemin dont l'extrémité initiale est la racine de l'arbre et celle terminale la feuille qui représente l'espèce. La figure 1.12. ci-dessous présente l'arbre de détermination associé à l'extrait de la taxinomie des orchidées décrite sur la figure 1.10.

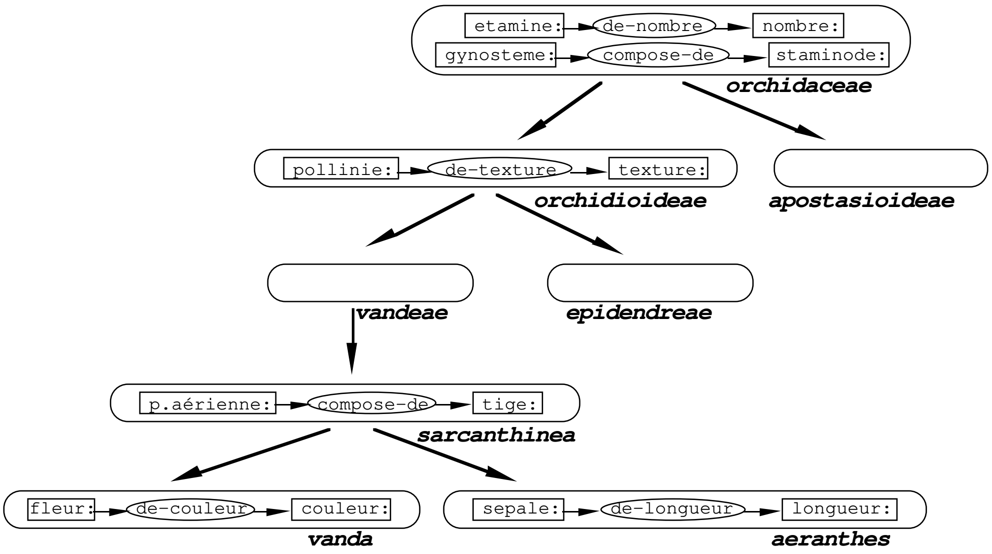


figure 1.12. Représentation de l'arbre de détermination associé à la taxinomie des orchidées

1.4.2. Construction de la base de connaissances

Dans cette partie, nous nous intéressons à la construction d'une base de connaissances taxinomiques en termes des éléments du modèle des graphes conceptuels. Nous ne présentons pas les aspects méthodologique et algorithmique de l'acquisition des connaissances ; ils font l'objet du chapitre 3 de cette thèse. Notre intention est ici de préciser la description des différents éléments d'une base de connaissances taxinomiques que nous avons fournie dans la partie précédente, en considérant maintenant la base décrite dans une perspective dynamique d'acquisition. Plus précisément, nous envisageons une base de connaissances comme le résultat d'un travail d'explicitation de connaissances. En effet, notre démarche ne consiste pas à définir des types de concepts pour représenter des taxons *à partir de* primitives de langage, et ce dans le but d'expliciter une structure taxinomique. Au contraire, la structure de la taxinomie étant connue, ce sont les primitives de langage que nous cherchons à expliciter, celles nécessaires à la définition de types de concepts représentant des taxons dont les relations de généralisation qu'ils entretiennent rendent compte de leur organisation taxinomique (cf. § 1.1.2.). En d'autres termes, notre objectif est d'expliciter un langage de description discriminant et ce en explicitant des traits discriminants entre taxons frères. L'élaboration des descriptions des taxons est subordonnée à cet objectif.

1.4.2.1. Base canonique

Dans une perspective dynamique de construction des graphes conceptuels qui définissent les types de concepts représentant des taxons, la correction de ces graphes peut être vue en termes de canonicité. Une base canonique permet en effet d'imposer des contraintes minimales dans la construction d'une base de connaissances (cf. Définition 1.9.). Étant donnée une base canonique, un graphe adéquat à la définition d'un type de concept représentant un taxon est un graphe canonique, c'est-à-dire dérivable à partir des graphes de cette base (cf. Définition 1.10.). C'est cette base canonique que nous précisons ici.

La structure de la taxinomie est connue et en particulier sa racine dans notre exemple, il s'agit de la famille des *Orchidaceae* dans l'ordre des *Orchidales*. De fait, nous avons établi que le type atomique `orchidale` est une généralisation commune à tous les types de concepts définis représentant des taxons d'orchidées (cf. § 1.4.1.2.). Donc, d'après la définition de la relation de généralisation entre un type de concept atomique et un type défini (Définition 1.11.), le genre de tout type de concept défini représentant un taxon doit être inférieur ou égal au type atomique `orchidale`. En particulier, le genre du type `orchidaceae` est `orchidale` ; c'est-à-dire que la tête du graphe de définition du type `orchidaceae` a pour type `orchidale`. La description la plus sommaire de la famille des *Orchidaceae* est en effet de dire que ce sont des *Orchidales*, ce qui peut être représenté dans le modèle des graphes conceptuels en définissant le type `orchidaceae` par le graphe `[orchidale:x]` réduit à un noeud concept unique : sa tête, de type `orchidale`. D'après la définition de la relation de généralisation entre deux types de concepts définis (Définition 1.11.), tout type défini représentant un descendant de la famille des *Orchidaceae* est inférieur au type défini `orchidaceae` et doit donc être défini par un graphe tel qu'il existe une dérivation de ce graphe contenant le graphe de définition du type `orchidaceae`. En d'autres termes, pour respecter la structure de la taxinomie des orchidées à représenter, les types des concepts représentant les taxons de cette taxinomie doivent tous être définis par des graphes qui soient dérivables à partir du graphe de définition du type `orchidaceae`. Pour observer cette contrainte tout au long de la construction de la base de connaissances, le graphe de définition de la racine de la taxinomie est *déclaré* canonique, c'est-à-dire appartenant à la base canonique. Ainsi tous les types définis seront effectivement dérivés de ce graphe.

Outre cette contrainte sémantique imposée dans la construction d'une base de connaissances taxinomiques, nous avons stipulé des exigences d'ordre syntaxiques quant aux graphes utilisés pour définir des types de concepts représentant des taxons : elles imposent que chaque graphe soit le résultat d'une séquence de joints entre arcs conceptuels qui représentent les traits descriptifs du taxon décrit (cf. § 1.4.1.). En termes de canonicité, ces règles syntaxiques de construction des graphes représentant des traits de composition ou de caractérisation sont formulées à l'aide de la base du support : les graphes signatures des relations `composé_de` et `caractérisé_par`. Une représentation correcte d'un trait descriptif est un graphe dérivable par une succession d'opérations de *restriction de concept et de relation* à partir du graphe canonique signature de la relation `composé_de` ou de la relation `caractérisé_par`.

Une représentation correcte d'un taxon est donc dérivable par une séquence de *joints* entre le graphe qui permet de définir le type représentant la racine de la taxinomie considérée et les graphes canoniques de la base du support qui assurent de la correction de la représentation des traits descriptifs. La figure 1.13. ci-dessous présente la base canonique à partir de laquelle sont définis dans notre exemple les types qui représentent les taxons de la taxinomie des orchidées.

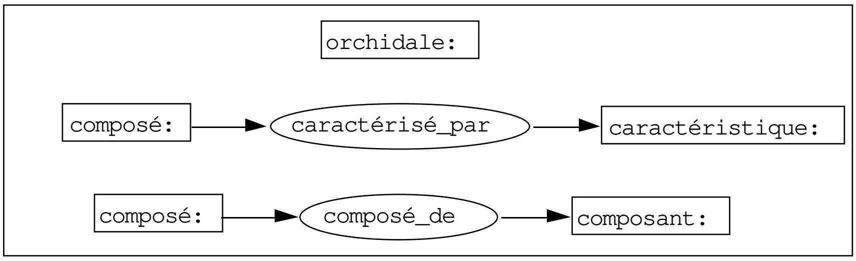


figure 1.13. Base canonique pour la taxinomie des orchidées

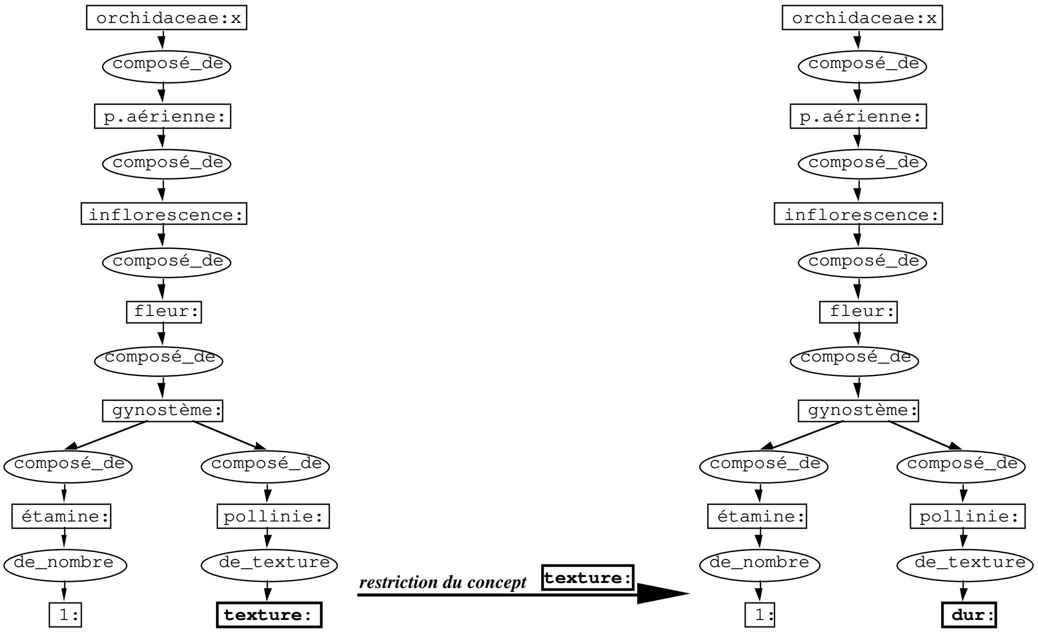
1.4.2.2. Définition de types de concepts représentant des taxons

La racine d'une taxinomie à acquérir est représentée par un type de concept dont le genre est un type représentant un taxon d'ordre supérieur ; dans notre exemple, la famille des *Orchidaceae* est représentée par le type `orchidaceae` dont le genre est le type (atomique) `orchidale` qui représente l'ordre des *Orchidales*. Les taxons de rangs inférieurs sont représentés par des types de concept dont les formes atomiques de leurs définitions ont également pour genre - atomique - le type `orchidale`. En d'autres termes, le graphe canonique [`orchidale:`] constitue la tête commune à l'ensemble des définitions des types représentant des taxons d'orchidées.

De plus, pour que les relations de généralisation entre les types de concepts représentant les taxons représentent la structure de la taxinomie, le graphe de définition de chaque type est construit à partir de celui du type représentant le taxon père, en spécialisant ce dernier par restriction d'un trait de caractérisation ou par joint avec un graphe représentant un nouveau trait de composition ou de caractérisation. De la sorte, dans les formes atomiques des définitions de types, la description d'un taxon “comprend” explicitement celles de ses ancêtres dans la taxinomie.

Un moyen de mieux faire apparaître les relations de filiation entre taxons serait de présenter des définitions de types sous forme contractée, et de représenter ainsi une sous-famille par un type dont le genre est le type `orchidaceae`, une tribu par un type dont le genre est le type qui représente la sous-famille de la tribu, etc. Pour cela, il faudrait pouvoir remplacer dans la forme atomique d'une définition de type le sous-graphe correspondant à la description du taxon père par le type (défini par ce graphe) qui représente ce taxon. Cependant, l'opération de contraction (Définition 1.13) telle qu'elle est définie par Leclère ne nous le permet pas puisque dans la description `G` d'un taxon le sous-graphe `G'` correspondant à la description du taxon père n'est pas connecté au reste du graphe uniquement par sa tête. En effet, des joints avec des graphes représentant de nouveaux traits de composition ou de caractérisation peuvent être opérés à n'importe quel noeud de `G'` de type inférieur à `composant`. De plus, certains concepts de `G'` de types inférieurs à `caractéristique` peuvent être restreints dans `G`. Nous considérerons donc toujours les formes atomiques des définitions de types.

Pour représenter la structure de la taxinomie par les relations de généralisation entre les types de concepts qui représentent ses taxons, il ne suffit pas de représenter les relations de “filiation” ; il faut encore représenter celles de “fraternité”. Deux types de concepts représentant deux taxons fils d'un même taxon doivent ne pas être comparables par la relation de généralisation. C'est la règle qui régit l'explicitation des descriptions de taxons. Considérons par exemple les types représentant les tribus des *Vandeae*, des *Neottieae* et des *Epidendreae*. Pour qu'ils ne soient pas comparables, leurs définitions doivent comporter la représentation de la texture des pollinies de leurs fleurs qui diffère d'une tribu à l'autre (cf. § 1.4.1.4.). Quant au type qui représente le père de ces tribus, la sous-famille des *Orchidioideae*, on peut en préciser la définition en représentant le fait que sa fleur est caractérisée par sa texture. Ces caractéristiques sont représentées par des graphes qui sont des spécialisations du graphe signature du type de relation `caractérisé_par`, par restriction de concepts et de relations. Comparons le graphe de définition du type `orchidioideae` avec les graphes de définition des types `vandeae`, `neottieae` et `epidendreae`. Comme le montre la figure 1.14. suivante, ces derniers sont des spécialisations du graphe de définition du type `orchidioideae` par *restriction* du type de concept `texture`. De plus, ils sont incomparables entre eux puisque les types atomiques `mou`, `dur` et `demi_mou` sont incomparables, comme cela apparaît sur l'arbre T_C du support (cf. figure 1.11.). La structure de la taxinomie est donc bien représentée.



graphe de définition
du type orchidioideae

graphe de définition
du type vandeae

figure 1.14. Définition du type vandeae par spécialisation du type orchidioideae

Dans le cas où le trait descriptif discriminant est un trait de composition, la démarche est similaire. Considérons par exemple les types représentant la sous-tribu des *Sarcanthinea* de la tribu des *Vandeae* et les genres *Vanda* et *Aeranthes* (cf. figure 1.10.). Ces taxons sont discriminés par la présence d'une tige chez les spécimens de l'un et son absence chez les spécimens de l'autre (cf. figure 1.12.). La présence d'une tige est représentée par un graphe qui est une spécialisation du graphe signature du type de relation composée. L'absence d'un composant est représentée par l'absence d'un tel sous-graphe dans la description du taxon. Les types vanda et aeranthes ainsi définis sont des spécialisations du type sarcanthinea. Plus précisément, comme le montre la figure 1.15. suivante, le graphe de définition du type vanda est le résultat d'un joint entre le graphe de définition du type sarcanthinea et celui représentant la présence d'une tige sur la partie aérienne de la plante.

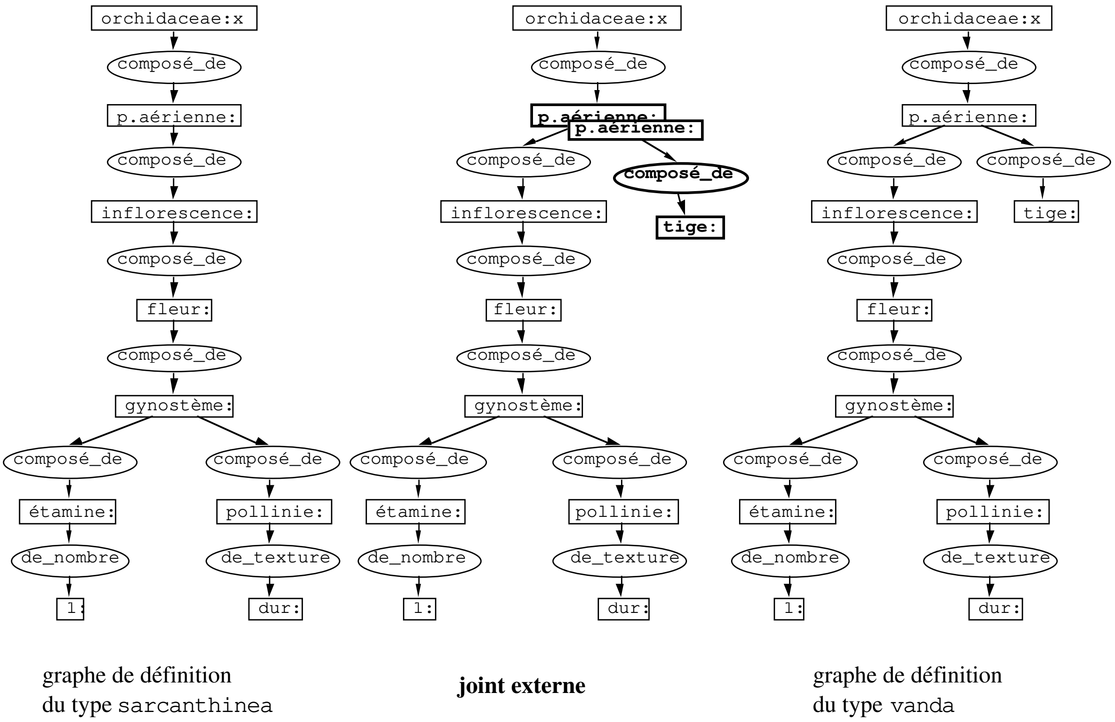


figure 1.15. Définition du type `vanda` par spécialisation du type `sarcanthinea`

Reste à vérifier que les graphes de définition des types représentant des taxons ainsi construits sont bien canoniques. Pour cela, il suffit de montrer qu'ils sont tous dérivables à partir des graphes de la base canonique que nous avons définie, par une séquence finie d'opérations de spécialisation. Chaque type représentant un taxon est défini par un graphe dérivé de celui de son père : le graphe de définition d'un type représentant un taxon est construit par une séquence d'opérations de *restriction* de types de concepts et de *joint* avec le graphe de définition du type représentant le père du taxon à décrire. Plus précisément, il s'agit de restriction de types de concepts représentant des caractéristiques et de *joint* avec des arcs conceptuels représentant des traits de composition ou de caractérisation. Ces arcs conceptuels sont eux-mêmes canoniques puisqu'ils sont des spécialisations des graphes signatures des types de relations `composé_de` et `caractérisé_par`. Le graphe de définition d'un type de concept représentant un taxon est donc canonique si celui du type représentant son père est lui-même canonique. On est ainsi ramené à montrer la canonicité du graphe de définition du type représentant la racine de la taxinomie, ce qui est immédiat puisqu'il appartient à la base canonique. Tous les taxons sont donc bien décrits par des graphes canoniques.

1.4.2.3. Explicitation d'un langage de description

Sans entrer dans le détail de la méthode d'acquisition que nous utilisons (cf. chapitre 3), soulignons ici que le graphe de définition d'un type de concept représentant un taxon n'est pas construit "une fois pour toutes". Au contraire, il est spécialisé chaque fois que cela est nécessaire, au moment de l'introduction d'un taxon frère ou fils, pour le discriminer de son frère ou discriminer ses fils. Le langage de description des taxons est explicité au fur et à mesure que de nouveaux taxons sont introduits dans la base de connaissances, que de nouveaux traits descriptifs sont introduits dans les descriptions des taxons. Plus précisément, de nouveaux types de concepts atomiques inférieurs à `composant` sont introduits dans le support au fur et à mesure qu'il est fait mention de nouvelles parties et sous-parties dans la description des taxons. De nouveaux types de relations atomiques inférieurs à `caractérise_par` et de nouveaux types de concepts atomiques inférieurs à `caractéristique` sont introduits au fur et à mesure que de nouveaux traits caractéristiques sont introduits par joint dans les définitions des types représentant les taxons et spécialisés par restriction de concept. Cette spécialisation des définitions des types de concepts représentant les taxons est guidée par la discrimination des taxons qu'ils représentent. De fait, le support ainsi construit représente un langage de description discriminant, juste suffisant pour l'élaboration de descriptions discriminantes des taxons introduits dans la base.

1.5. Conclusion

Dans ce chapitre, nous avons posé le cadre général de notre travail, la représentation de connaissances pour des domaines structurés tels que les sciences naturelles. Nous avons caractérisé les connaissances taxinomiques de ces domaines et mis en évidence l'adéquation du modèle des graphes conceptuels simples pour les représenter. Puis nous avons décrit dans le détail les particularités d'une base de connaissances taxinomiques telles que nous les représentons dans le modèle des graphes conceptuels.

Les connaissances du monde réel que nous voulons représenter sont souvent empruntes d'imprécision qui se traduisent par des exceptions à l'héritage sur les taxinomies établies. Le modèle des graphes conceptuels simples que nous venons de proposer pour représenter ces connaissances ne permet pas de rendre compte de ces imprécisions. Pour cela, nous nous intéressons dans le chapitre suivant à étendre ce formalisme.

Chapitre 2

Représentation des exceptions sur les taxinomies

Introduction

Dans les sciences naturelles, la diversité du vivant rend particulièrement difficile la tâche de description des taxons telle que nous l'avons définie dans le chapitre précédent, en termes de conditions nécessaires et suffisantes d'appartenance aux taxons. Dans ces domaines, des spécimens existent en effet qui constituent des exceptions à certains traits descriptifs associés aux taxons dont ils sont pourtant déclarés instances. Par exemple, la couleur des fleurs d'orchidées du genre *Cephalanthera* est dite “rose ou blanche”, celle de l'espèce *C. Pallens*, “blanche et jaune orangé”. Or l'espèce *C. Pallens* est un taxon inférieur au genre *Cephalanthera*, ce qui signifie que les spécimens de l'espèce *C. Pallens* sont des spécimens du genre Cephalanthera. Donc, en règle générale, “par défaut”, la couleur des fleurs d'orchidées du genre *Cephalanthera* est dite “rose ou blanche” mais il existe des *exceptions* dans certains des taxons inférieurs à ce genre, telles que “blanche et jaune orangé” pour les spécimens de l'espèce *C. Pallens*.

Dans ce chapitre, nous nous intéressons à la représentation de telles exceptions à l’héritage sur des structures hiérarchiques. Dans la première partie, nous dégageons à partir d’un exemple une caractérisation plus précise de ce problème de représentation et mettons en évidence l’impossibilité de le résoudre dans le formalisme des graphes conceptuels tel que nous l’avons décrit dans le chapitre précédent (§ 2.1.). Nous exposons ensuite différentes solutions qui ont été proposées à ce problème dans d’autres formalismes de représentation et nous montrons la meilleure adéquation d'une démarche définitionnelle à notre problème particulier (§ 2.2.). Nous proposons alors une extension du modèle des graphes conceptuels simples qui s'inspire des travaux de Coupey et Fouqueré dans le cadre des logiques de descriptions [Coupey et Fouqueré, 97] (§ 2.3.).

2.1. Une limite du modèle des Graphes Conceptuels

2.1.1. Des taxinomies comportant des exceptions

2.1.1.1. Un exemple : la taxinomie des animaux contemporains

Considérons un extrait de la taxinomie des animaux contemporains, un exemple auquel font classiquement référence les travaux sur la gestion d'exceptions en représentation des connaissances (cf. § 2.2.). Plus particulièrement, intéressons-nous à la classe `éléphant`, une sous-classe de la classe `animal-contemporain`. Comme le montre la figure 2.1. suivante, elle se subdivise en deux classes : `éléphant-africain` et `éléphant-royal`. La classe `éléphant-royal-sale` est une sous-classe de la classe `éléphant-royal`.

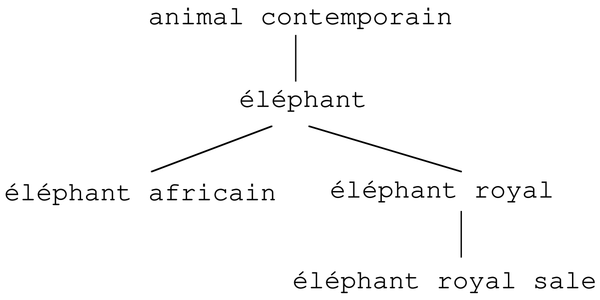


figure 2.1. Taxinomie des animaux contemporains

2.1.1.2. Représentation dans le modèle des graphes conceptuels

Définissons chacun des taxons de cette taxinomie dans le modèle des graphes conceptuels. Rappelons que la description d'un taxon est représentée par un arbre dont les types de relations conceptuelles sont les types de relations binaires `caractérisation` ou `composition`, ou des sous-types de ceux-ci, un taxon étant décrit par ses parties et sous-parties et les caractéristiques de celles-ci (cf. § 1.4.).

Nous définirons les éléphants par le fait qu'ils possèdent une trompe et sont gris, à l'exception des éléphants royaux. Ces derniers sont en effet de couleur blanche, à l'exception des éléphants royaux sales qui sont de couleur grise[2].

La figure 2.2. ci-dessous présente les définitions des types de concepts `E`, `EA`, `ER` et `ERS` qui représentent les taxons `éléphant`, `éléphant-africain`, `éléphant-royal` et `éléphant-royal-sale`.

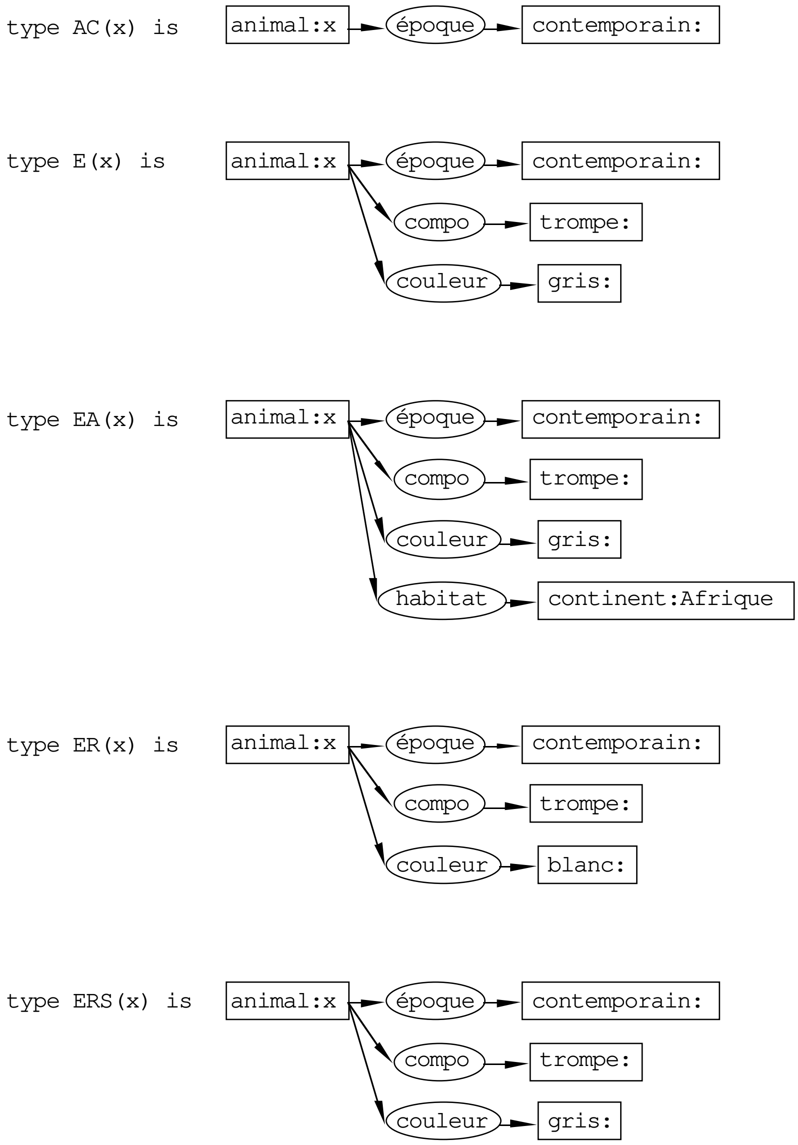


figure 2.2. Définition des taxons de la taxinomie des animaux contemporains

---

[2] Soulignons qu'il ne s'agit pas ici de donner des descriptions réalistes des animaux considérés. La faiblesse des descriptions choisies doit permettre en contrepartie une compréhension immédiate du problème de représentation qui se pose.

2.1.1.3. Échec dans la représentation de la taxinomie

Considérons maintenant la hiérarchie de généralisation sur les types ainsi définis. Elle devrait normalement représenter la taxinomie des animaux contemporains. La racine de cette hiérarchie est le type `AC` qui représente le taxon `animal-contemporain`. Le type `E` qui représente le taxon `éléphant` est inférieur à `AC`. Son graphe de définition est en effet une spécialisation de celui de `AC` : il peut être dérivé à partir de ce dernier par joint comme cela est présenté sur la figure 2.3. ci-dessous.

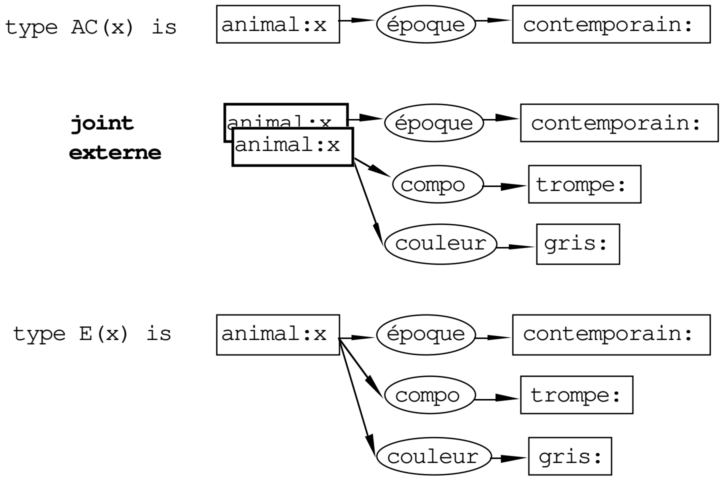


figure 2.3. Dérivation du graphe de définition de E à partir de celui de AC

De même, le type `EA` qui représente le taxon `éléphant-africain` est inférieur à `E` : son graphe de définition peut être dérivé à partir de celui de `E` par joint. Pour que la taxinomie des animaux contemporains de la figure 2.1. soit représentée par les types définis sur la figure 2.2., il faudrait encore que le graphe de définition du type `ER` qui représente le taxon `éléphant-royal` soit une spécialisation de celui de `E` et que le graphe de définition de `ERS` qui représente le taxon `éléphant-royal-sale` soit une spécialisation de celui de `ER`. Or, il apparaît clairement qu'aucune règle de spécialisation ne permet de dériver le graphe de définition de `ER` de celui de `R` ni le graphe de définition de `ERS` de celui de `ER`. Plus précisément, aucune règle de spécialisation ne permet de remplacer dans le graphe de définition de E le type de concept `gris` par `blanc`, ni dans celui de `ER` le type de concept `blanc` par `gris`. En effet, la seule opération de spécialisation qui consiste à remplacer dans un graphe conceptuel le type d'un des noeuds concepts par un autre est la restriction de concept (cf. § 1.3.4.). Mais un concept ne peut être restreint qu'en diminuant le type de son étiquette, c'est-à-dire en remplaçant son type par un sous-type, or `blanc` n'est pas un sous-type de `gris`. `ER` n'est donc pas une spécialisation de `E`, ni `ERS` une spécialisation de `ER`.

Les relations de généralisation entre les graphes construits pour décrire les taxons ne coïncident donc pas avec la structure de leur taxinomie. Comme le montre la figure 2.4., les plus petits généralisés des graphes `ER` et `ERS` sont en effet les graphes respectivement `AC` et `E`, et non pas `E` et `ER`. La hiérarchie de généralisation sur les types tels qu'ils sont définis sur la figure 2.2. ne permet donc pas de représenter la taxinomie des animaux contemporains présentée sur la figure 2.1.

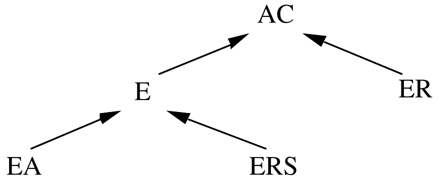


figure 2.4. Hiérarchie de généralisation sur les types représentant les taxons d'animaux contemporains

L'exemple de la taxinomie des animaux contemporains nous a permis de mettre en évidence la nécessité d'enrichir le modèle des graphes conceptuels pour représenter des taxinomies comportant des exceptions et, plus précisément, pour que la hiérarchie de généralisation sur les types de concepts définis à l'aide de graphes décrivant les taxons coïncide avec la structure de la taxonomie à représenter. Nous présentons dans la partie suivante différentes alternatives pour pallier ce problème de représentation.

2.1.2. Vers une extension du modèle des graphes conceptuels

Le problème de la gestion d'exceptions à l'héritage dans le modèle des graphes conceptuels se ramène, dans l'exemple de la taxinomie des animaux contemporains, à celui d'établir une relation de généralisation entre les graphes `ER` et `E`. En d'autres termes, il faut que le graphe `ER` soit dérivable à partir du graphe `E`. Pour cela, nous envisageons dans cette partie différentes solutions. Nous montrons que définir une nouvelle opération de spécialisation est impossible. Nous mettons alors en évidence l'alternative entre redéfinir la relation de généralisation entre graphes ou redéfinir le support de construction des graphes conceptuels.

2.1.2.1. Une nouvelle opération de spécialisation

Tels que les taxons `éléphant` et `éléphant-royal` sont décrits (cf. figure 2.2.), aucune opération de spécialisation définie dans le modèle des graphes conceptuels ne permet de dériver le graphe de définition de `ER` à partir de celui de `E` car aucune ne permet le remplacement du type de concept `gris` dans le graphe `E` par `blanc`. Supposons l'existence d'une nouvelle opération `Ops` qui permettrait une telle dérivation et considérons alors un nouveau taxon inférieur à `animal-contemporain` dont les instances possèdent une trompe et dont la couleur soit verte : le taxon `mouche-verte`. Le type de concept `MV` qui représente ce taxon est défini sur la figure 2.5. ci-dessous.

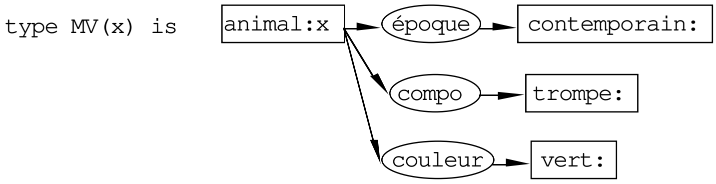


figure 2.5. Représentation du taxon `mouche-verte`

L'opération de spécialisation `OpS` permettrait de remplacer le type de concept `gris` dans le graphe `E` aussi bien par `vert` que par `blanc`, donc de dériver à partir de `E` aussi bien `MV` que `ER` et d'établir ainsi les relations de généralisation `ER≤E` et `MV≤E`. En d'autres termes, `mouche-verte`, comme `éléphant-royal`, serait décrit comme un taxon inférieur à `éléphant` ; rien ne permet dans la syntaxe des descriptions de ces taxons de distinguer la sémantique différente de la couleur au regard du gris dans la description du taxon père `éléphant`.

De plus, l'opération `OpS` permettant de remplacer le type de concept `gris` dans le graphe de définition de `E` par `blanc` ou `vert` permettrait, symétriquement, de remplacer `blanc` dans le graphe de définition de `ER` ou `vert` dans celui de `MV` par `gris`. Les relations de généralisation `E≤ER` et `E≤MV` seraient ainsi établies, en dérivant le graphe de définition de `E` à partir de celui de `ER` ou de `MV` à l'aide de l'opération de spécialisation `OpS`. Cela signifie que la relation de généralisation sur les graphes ne serait plus antisymétrique, donc plus une relation d'ordre. Il apparaît ainsi que la relation de généralisation entre les graphes définissant les taxons `éléphant` et `éléphant-royal` ne peut être simplement fondée sur une nouvelle règle de spécialisation.

2.1.2.2. Révision de la relation de généralisation

Une première solution pour représenter une taxinomie comportant des exceptions à l'héritage consiste à réviser les fondements de la relation de généralisation sur les graphes conceptuels de manière à ce qu'elle ne repose plus uniquement sur les opérations de spécialisation définies dans le modèle. Il s'agit de redéfinir la relation de généralisation de manière à ce qu'elle puisse être établie entre deux graphes tels que ceux définissant `E` et `ER`, alors qu'aucune opération de spécialisation ne permet de dériver `ER` à partir de `E` et que toute la description de `E` n'est pas héritée par `ER` (le `gris` devient `blanc`). Cette solution relève d'une démarche *inférentielle* qui consiste à aborder le problème des exceptions sur une structure hiérarchique par un mécanisme *d'héritage par défaut* de propriétés qui ne sont pas *strictes* et admettent des *exceptions*

2.1.2.3. Révision du support du modèle des graphes conceptuels

Une autre solution pour représenter une taxinomie comportant des exceptions consiste à préserver les opérations de spécialisation comme fondements de la relation de généralisation sur les graphes conceptuels et, pour cela, à gérer les exceptions dans la définition même des taxons. Il s'agit de faire apparaître dans les descriptions des taxons quels traits admettent des exceptions, en distinguant ceux *par défaut*, tels que le `gris` d'`éléphant`, de ceux *stricts*, tels que le `gris` d'`éléphant-royal-sale` ou le `vert` de `mouche-verte`. De la sorte, l'antisymétrie de la relation de généralisation est restaurée. De plus, pour que `mouche-verte` ne soit plus définie, comme `éléphant-royal`, comme un taxon inférieur à `éléphant`, il faut aussi pouvoir souligner le caractère *exceptionnel* d'un trait descriptif, tel que le `blanc` d'`éléphant-royal` au regard du `gris`, par défaut, d'`éléphant`. Cette solution relève d'une démarche dite définitionnelle, par opposition à celle inférentielle. Il s'agit de définir des *défauts* et des *exceptions* dans les descriptions des taxons pour ne considérer alors non pas *l'héritage par défaut* de traits descriptifs mais plutôt l'héritage de *traits par défaut* .

La partie suivante fait état des travaux sur la gestion d'exceptions sur des structures hiérarchiques. Notre présentation s'appuie sur cette dichotomie que nous venons de mettre en évidence entre démarches définitionnelle et inférentielle.

2.2. Les approches pour la gestion des exceptions à l’héritage

Cette partie est dédiée à une synthèse des travaux sur la gestion d'exceptions sur des structures hiérarchiques en représentation des connaissances. Nous présentons ces travaux selon la démarche adoptée et le formalisme de représentation utilisé. Réviser le système d'héritage associé à une structure hiérarchique relève d'une démarche inférentielle (§ 2.2.1) tandis que modifier la représentation même de cette structure hiérarchique relève d'une démarche définitionnelle (§ 2.2.2). Nous montrons l'adéquation d'une démarche définitionnelle à la spécificité de notre tâche de représentation et au formalisme des graphes conceptuels (§ 2.2.3).

2.2.1. Héritage par défaut de propriétés

Aborder la représentation de connaissances d'un point de vue inférentiel, c'est s'attacher aux capacités de déduction d'une base de connaissances. Une manière de décrire l'ensemble des déductions élémentaires entre les concepts d'un domaine est de structurer la base de connaissances en organisant les concepts sous forme hiérarchique puis de fonder un mécanisme d'héritage de propriétés sur cette structure hiérarchique. Ces hiérarchies sont encore appelées *hiérarchies d'héritage* [3] [Etherington et Reiter, 83]. Dans cette perspective, les exceptions qui apparaissent sur une hiérarchie de concepts sont gérées par le mécanisme d'héritage de propriétés associé à la hiérarchie. Un tel mécanisme est dit d'*héritage par défaut*, au sens où des exceptions peuvent bloquer l'héritage normal d'une propriété. L'héritage par défaut pour la représentation de structures hiérarchiques comportant des exceptions est le point de vue privilégié en logique des défauts ou dans les réseaux sémantiques. Plus marginalement, il a été adopté dans les logiques de descriptions.

---

[3] En anglais : *inheritance hierarchies*.

2.2.1.1. Logique des défauts

Une hiérarchie de concepts peut être représentée en logique du premier ordre par un ensemble de clauses, appelé *théorie*. Ces clauses traduisent des implications logiques entre des propositions qui désignent les concepts du domaine. Les implications entre ces propositions représentent les propriétés des concepts et la structure hiérarchique sur ces concepts. Le calcul des propriétés héritées d'un concept de la hiérarchie se ramène au calcul de l'extension de la théorie correspondante. Cependant, la logique du premier ordre ne permet pas d'exprimer les propriétés des concepts dès lors qu'elles admettent des cas exceptionnels. Plus généralement, Reiter montre que la logique du premier ordre n'est pas adaptée à la modélisation des *raisonnements par défaut* [Reiter, 78]. Or de tels raisonnements sont très répandus. Beaucoup de nos connaissances sur le monde sont en effet presque toujours vraies - à l'exception de quelques cas particuliers. Tous les oiseaux volent - à l'exception des pingouins, des autruches ; tous les éléphants sont gris - à l'exception des éléphants royaux ; etc. Il s'agit de pouvoir conclure d'un oiseau qu'il vole, à moins qu'il ne s'agisse d'une exception, c'est-à-dire en l'absence d'information précisant qu'il est un pingouin ou une autruche.

Une représentation en logique du premier ordre oblige à lister les exceptions :

$$\forall x\ \texttt{oiseau(x)} \wedge \neg\texttt{pingouin(x)} \wedge \neg\texttt{autruche(x)} \wedge \ldots \supset \texttt{vole(x)}$$

Étant donné un oiseau `tweety`, pour conclure la propriété générale `vole(tweety)`, cette représentation requiert de *vérifier* que `tweety` n'est pas une exception en évaluant la formule $\neg$`pingouin(tweety)` $\wedge$ $\neg$`autruche(tweety)` $\wedge$ `...` Or cette démarche est contraire au raisonnement par défaut qui consiste à conclure `vole(tweety)` précisément *en l'absence d'information* précisant qu'il est un pingouin ou une autruche.

Pour modéliser les raisonnements par défaut, Reiter propose une logique, dite *logique des défauts* [Reiter, 80]. Elle consiste en un enrichissement de la logique du premier ordre par des règles d'inférence, appelées *défauts*, parce qu'elles permettent de conclure des faits par défaut, c'est-à-dire en l'absence d'information contraire. Ces règles sont de la forme $\frac{\alpha:\beta}{\delta}$, où α, β et δ sont des formules du premier ordre ; α est la prémisse du défaut, δ sa conclusion et β sa justification. L'exemple précédent est reformulé à l'aide de la règle défaut suivante :

$$\frac{\texttt{Oiseau(x):Vole(x)}}{\texttt{Vole(x)}}.$$

Cette règle stipule de conclure par défaut que `vole(tweety)` si `oiseau(tweety)` et s'il est consistant de supposer `vole(tweety)`, c'est-à-dire si la formule `¬vole(tweety)` ne peut être établie. Les exceptions sont représentées par des formules du premier ordre qui concluent précisément sur `¬vole(x)` :

$$\mathtt{pingouin(x)} \supset \neg\mathtt{vole(x)},\ \mathtt{autruche(x)} \supset \neg\mathtt{vole(x)}.$$

La figure 2.6. ci-dessous illustre sur l'exemple de la taxinomie des animaux contemporains l'utilisation qui est faite de la logique des défauts par Etherington et Reiter comme sémantique formelle pour les structures hiérarchiques avec exceptions [Etherington et Reiter, 83]. Les prédicats `A`, `C`, `AC`, `E`, `EA`, `ER`, `ERS`, `Tr`, `G`, `B` et `AF` désignent respectivement `animal`, `contemporain`, `animal-contemporain`, `éléphant`, `éléphant-africain`, `éléphant-royal`, `éléphant-royal-sale`, `trompe`, `gris`, `blanc` et `africain`.

$$\begin{array}{l}
\{\mathtt{AC(x)} \supset \mathtt{A(x)}, \mathtt{AC(x)} \supset \mathtt{C(x)}, \mathtt{E(x)} \supset \mathtt{AC(x)},\\
\mathtt{EAF(x)} \supset \mathtt{E(x)}, \mathtt{ER(x)} \supset \mathtt{E(x)}, \mathtt{ERS(x)} \supset \mathtt{ER(x)},\\
\mathtt{E(x)} \supset \mathtt{Tr(x)}, \dfrac{\mathtt{E(x)}:\mathtt{G(x)}}{\mathtt{G(x)}}, \mathtt{EAF(x)} \supset \mathtt{AF(x)},\\
\dfrac{\mathtt{ER(x)}:\mathtt{B(x)}}{\mathtt{B(x)}}, \dfrac{\mathtt{ER(x)}:\neg\mathtt{G(x)} \wedge \mathtt{ERS(x)}}{\neg\mathtt{G(x)}},\\
\mathtt{ERS(x)} \supset \neg\mathtt{B(x)}\}
\end{array}$$

figure 2.6. Une théorie des défauts pour la taxinomie des animaux contemporains

Le calcul des propriétés héritées d'un concept appartenant à une hiérarchie d'héritage avec exceptions se ramène, comme en logique du premier ordre, au calcul de l'extension de la théorie qui représente la hiérarchie. Par exemple, si on introduit dans la théorie décrite ci-dessus la proposition `ER(babar)`, son extension contiendra les propositions `E(babar)` et `TR(babar)` établies par l'application du mods ponens ainsi que `B(babar)` et `G(babar)` établies par l'application de règles défauts.

Cet exemple met en évidence l'adéquation de la logique des défauts pour la représentation de hiérarchies d'héritage, dès lors que l'héritage admet des exceptions et n'est plus modélisable par des formules du premier ordre. Plus précisément, une telle hiérarchie d'héritage est représentée par une théorie des défauts, c'est-à-dire un ensemble d'implications du premier ordre et de défauts.

2.2.1.2. Réseaux Sémantiques

Un réseau sémantique est un graphe dont les noeuds figurent des concepts et les arcs des relations entre ces concepts. Sous l'appellation de réseaux sémantiques, de nombreux langages de représentation sont regroupés, qui partagent les caractéristiques énoncées ci-dessus. Nous envoyons le lecteur intéressé par un exposé de ces différents langages aux recueils d'articles de référence [Lehman, 92] et [Sowa, 91].

Il est aisé de représenter une hiérarchie de concepts d'un domaine par un réseau sémantique : les noeuds du réseau figurent les concepts du domaine et un type d'arcs est réservé à la représentation de la relation de généralité entre ces concepts. Un réseau sémantique est doté d'un mécanisme d'héritage de propriétés grâce auquel un noeud hérite des propriétés des noeuds qui le subsument.

Falhman s'est le premier intéressé à la gestion des exceptions sur une hiérarchie d'héritage représentée par un réseau sémantique [Falhman, 79]. Dans cette version d'un réseau sémantique, implantée dans le système NETL, les exceptions prennent la forme de liens d'exception (`cancel`) qui empêchent l'héritage par un concept du réseau de certaines propriétés des subsumeurs de ce concept dans la hiérarchie [Falhman, 79 ; Falhman et al, 81]. Par exemple, pour représenter la taxinomie des animaux contemporains présentée sur la figure 2.1., un lien `cancel` entre les noeuds `éléphant-royal` et `gris` empêchera `éléphant-royal` d'hériter de la propriété `gris` d'`éléphant`, en cheminant le long des liens `is-a` entre `éléphant-royal` et `éléphant` puis entre `éléphant` et `gris`. Des liens d'exception (`uncancel`) qui pointent vers un lien `cancel` sont également introduits pour, par exemple, annuler le non héritage de `gris` par `éléphant-royal-sale`.

Etherington et Reiter proposent une variante à cette représentation [Etherington et Reiter, 83]. Ils distinguent les liens d'héritage strict (`strict is-a`) des liens d'héritage par défaut (`default is-a`) et seul l'héritage par défaut peut être bloqué, par un lien `cancel`. De la même manière, ils distinguent les liens d'exception stricte (`strict cancel`) des liens d'exception par défaut (`default cancel`). C'est la formalisation que nous avons utilisée dans l'exemple de la taxinomie des animaux contemporains de la figure 2.7. ci-après.

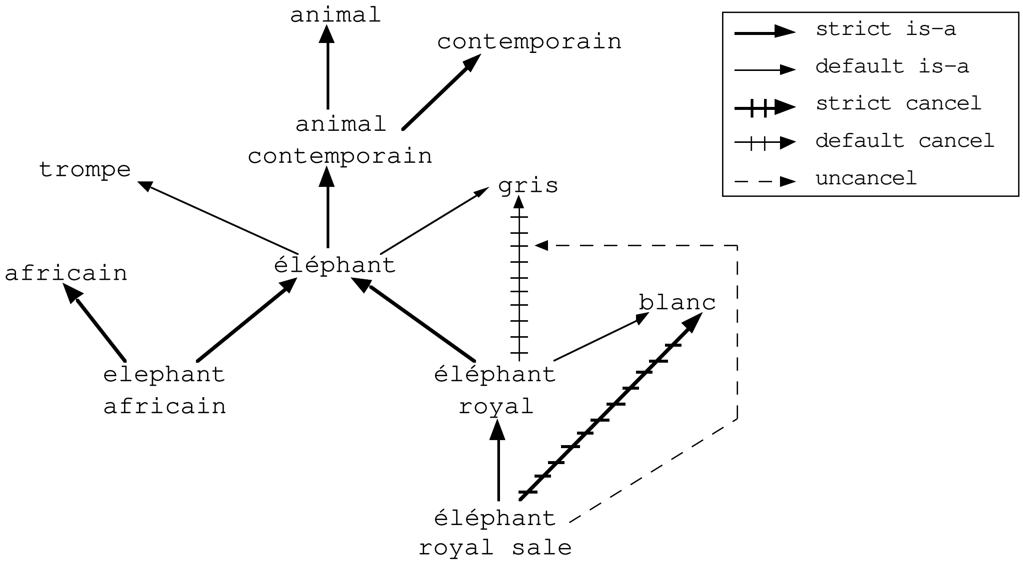


figure 2.7. Un réseau sémantique pour la taxinomie des animaux contemporains.

Le calcul des propriétés héritées d'un concept d'une hiérarchie d'héritage par défaut représentée par un réseau sémantique de type NETL est une procédure de recherche dans le réseau en cheminant le long des liens `is-a`. Celle-ci fait appel à des stratégies de propagation de marqueurs sur les noeuds du réseau, fondées sur des heuristiques d'exploration de type “plus court chemin” qui reflètent les priorités d'exploration entre différents types de liens [Falhman, 79]. C'est un calcul difficile ; dans les premiers algorithmes, décrits dans [Falhman, 79], des anomalies ont été détectées et les auteurs n'ont apporté que des solutions partielles [Fahlman et al, 81].

Pour pallier ce problème, des travaux s'attachent à la mise en évidence d'une correspondance entre les représentations de hiérarchies d'héritage à l'aide de théories des défauts et celles à l'aide de réseaux sémantiques de type NETL [Etherington et Reiter, 83], [Froidevaux, 86]. Des types de liens et leurs interprétations logiques (en logique des défauts) sont définis. Cette démarche assure de l'existence d'un algorithme correct pour le calcul des propriétés héritées par exploration du réseau sémantique. En effet, à un réseau sémantique correspond une théorie pour laquelle il s'agit de vérifier l'existence d'une et une seule extension et de la calculer. Par exemple, le réseau sémantique présenté sur la figure 2.8. correspond à la théorie des défauts présentée sur la figure 2.7. ; la même hiérarchie d'héritage y est représentée. Les liens `strict is-a` et `strict cancel` correspondent à des implications de la logique du premier ordre, les liens `default is-a` et `default cancel` à des défauts [Etherington et Reiter, 83]. Ainsi, de la même manière qu'une théorie des défauts, un réseau sémantique se prête bien à la représentation de hiérarchies d'héritage avec exceptions.

2.2.1.3. Logiques de descriptions

Les logiques de descriptions sont une famille de langages de représentation issus de KL-ONE [Brachman et Schmolze, 85]. Leur formalisme, comme celui des réseaux sémantiques, fait appel à des concepts et des relations entre concepts pour représenter un domaine. Cependant, à la différence des réseaux sémantiques, les concepts sont des descriptions structurées et munies d'une sémantique formelle. Les individus sont distincts des concepts, ils en sont les *instances*.

Dans un langage de descriptions, les niveaux de connaissances génériques et spécifiques sont distincts. On distingue ainsi deux ensembles :

- la TBox (Terminological Box) renferme la connaissance terminologique, c'est-à-dire les concepts du domaine, leurs définitions et leur ordonnancement par la relation de subsomption ; elle est donc organisée en une hiérarchie de concepts ;

- la ABox (Assertional Box) renferme la connaissance factuelle, c'est-à-dire les assertions relatives aux individus du domaine.

Notre propos n'est pas ici de présenter en détail les logiques de descriptions mais de montrer comment les exceptions y sont gérées. Pour une présentation plus approfondie, nous renvoyons le lecteur intéressé à l'article de synthèse [Napoli et Volle, 93].

Des travaux de représentation des connaissances en logiques de descriptions s'intéressent aux exceptions dans l'héritage de propriétés non pas sur la hiérarchie des concepts mais au niveau des individus du domaine [Baader et Hollunder, 92] [Quantz et Royer, 92]. Le point de départ de ces travaux est le mécanisme de classification d'objet de la ABox d'un langage de descriptions, c'est-à-dire l'inférence qui consiste à établir si un objet o est instance d'un concept C de la TBox du langage, ou, en d'autres termes, si un fait $C(o)$ est conséquence logique de l'ensemble des faits d'une ABox A, pour une TBox T donnée. Formellement, il s'agit de vérifier si $A \models^{T} C(o)$. Notons que le terme instanciation est utilisé dans [Baader et Hollunder, 92] plutôt que classification d'objet.

Les exceptions sont donc envisagées au niveau spécifique : un objet instance d'un concept C_1 peut ne pas posséder toutes les propriétés normalement héritées par les instances de C_1 en n'étant pas, exceptionnellement, instance d'un concept C_2. Plus précisément, l'idée est d'introduire des défauts au niveau de règles de la forme C_1 -> C_2 qui stipulent que toute instance du concept C_1 est aussi instance du concept C_2.

Baader et Hollunder remplacent les règles de la forme C_1 -> C_2 de la logique du premier ordre par des défauts de la logique de Reiter et définissent une *théorie des défauts terminologique* (A, D), où A est la ABox du langage et D un ensemble de défauts.

Formellement, un objet o est instance d'un concept C de la TBox T si $(A, D) \models^{T} C(o)$ [Baader et Hollunder, 92].

Quantz et Royer ajoutent aux côtés des règles de la logique du premier ordre un ensemble Δ de défauts de la forme $C_1 \sim> C_2$ qui expriment le fait que si un objet est instance du concept C_1 alors il est aussi instance du concept C_2, à moins que cela ne soit en conflit avec d'autres faits établis de la ABox. Plus formellement, un objet o instance d'un concept C_1 est instance d'un concept C_2 (et hérite donc de ses propriétés) si $A \cup \{C_1(o)\} \models_{\Delta} C_2(o)$, où A est la ABox du langage [Quantz et Royer, 92].

D'autres travaux de représentation des connaissances en logiques de descriptions s'intéressent à la gestion des exceptions sur les hiérarchies de concepts des TBox mais le point de vue adopté est non plus inférentiel mais définitionnel. Ils font l'objet de la partie suivante.

2.2.2. Définition de propriétés par défaut

Le point de vue définitionnel en représentation des connaissances s'oppose à celui inférentiel exposé dans la partie précédente en ce que la tâche de modélisation d'un domaine est abordée non plus sous l'angle des capacités d'inférence sur la base de connaissances qui modélise le domaine mais plutôt sous celui des définitions des concepts du domaine. En particulier dans les logiques de descriptions, la différence de point de vue est évidente : l'héritage de propriétés par les instances de concepts s'adresse à la partie assertionnelle du langage (ABox), la définition des concepts à la partie terminologique (TBox).

D'un point de vue définitionnel, des exceptions sur une hiérarchie d'héritage sont appréhendées dans la définition même des concepts. Il s'agit non plus d'un *héritage par défaut* de propriétés qui admettent des exceptions mais de l'héritage de *propriétés par défaut*. Le caractère exceptionnel de telles propriétés est appréhendé dans la définition même des concepts et l'héritage est normal.

2.2.2.1. Le langage DTL

Le langage DTL proposé par Padgham appartient à la famille des logiques de descriptions [Padgham, 89]. Dans ce langage, un concept est représenté par deux noeuds de la hiérarchie de concepts de la TBox :

- un noeud-noyau *(core)* auquel sont attachées les propriétés strictes du concept ;

- un noeud-défaut *(default)* auquel sont attachées les propriétés typiques du concept.

Formellement, un concept est représenté par un couple d'éléments de la TBox de DTL : un concept-noyau C_C et un concept-défaut C_D tels que C_C subsume C_D : $C_D \supseteq C_C$.

Les propriétés qui définissent le noeud-noyau d'un concept sont des conditions nécessaires ; toutes les instances du concept en héritent. Le noeud-défaut du concept est défini par ces propriétés strictes dont il hérite du noeud-noyau et par des propriétés par défaut qui peuvent ne pas être héritées par toutes les instances du concept.

Une relation de *subsomption par défaut* est définie aux côtés de la relation de subsomption stricte pour rendre compte de la relation structurelle entre un concept-noyau et le concept-défaut associé [Padgham et Zhang, 93]. La classification de concept, c'est-à-dire l'introduction d'un nouveau concept dans la TBox, fait appel à un algorithme standard de recherche des fils et pères immédiats qu'aurait le concept dans la hiérarchie de la TBox. Ce sont respectivement les concepts les plus spécifiques qui subsument le concept à classer et les concepts les plus généraux subsumés par ce concept. Un algorithme de référence est celui utilisé dans KL-ONE [Schmolze et Lipkis, 83]. Selon que le concept à classer est un concept-noyau ou un concept-défaut, c'est à la relation de subsomption stricte ou à celle par défaut qu'il est fait appel dans l'algorithme. La définition des concepts repose donc pour partie sur un mécanisme d'*héritage par défaut*, la démarche utilisée n'est pas seulement définitionnelle mais pour partie inférentielle.

Une application au diagnostic médical révèle une structure particulière de la base de connaissances construite à l'aide d'un tel système terminologique. Dans cette application, presque toutes les relations d'héritage entre concepts sont entre concepts-noyaux et la plupart des propriétés sont attachées aux concepts-défauts plutôt qu'aux concepts-noyaux [Padgham et Zhang, 93]. C'est l'hypothèse que nous avons faite pour représenter sur la figure 2.8. ci-dessous la taxinomie des animaux contemporains dans le langage DTL.

$$\mathtt{AC}_C \equiv \mathtt{A}_C \cap \mathtt{C}_C$$

$$\mathtt{E}_C \equiv \mathtt{AC}_C \cap \mathtt{Tr}_C \qquad \mathtt{E}_D \equiv \mathtt{E}_C \cap \mathtt{Gr}_C$$

$$\mathtt{EAf}_C \equiv \mathtt{E}_C \cap \mathtt{Af}_C$$

$$\mathtt{ER}_C \supseteq \mathtt{E}_C \qquad \mathtt{ER}_D \equiv \mathtt{ER}_C \cap \mathtt{Bl}_C$$

$$\mathtt{ERS}_C \supseteq \mathtt{ER}_C \qquad \mathtt{ERS}_D \equiv \mathtt{ERS}_C \cap \mathtt{Gr}_C$$

figure 2.8. TBox de DTL représentant la taxinomie des animaux contemporains

La classification d'objets, c'est-à-dire l'identification de la classe à laquelle appartient un objet, combine classification de concept et héritage par défaut de propriétés. Padgham et Nebel définissent la notion de *classification par défaut* [Padgham et Nebel, 93]. Elle consiste à établir les classes auxquelles appartient *probablement* un objet, eu égard à la définition des concepts noyaux et défauts qui la représentent. La classification du concept correspondant à la description d'un objet `o` fournit un premier ensemble de concepts dont `o` est reconnu instance ; l'héritage par défaut des propriétés attachées à ces concepts permet de raffiner la description de `o` qui est de nouveau classée. La classification d'objets procède donc d'un cycle qui alterne classification de concept et héritage par défaut de propriétés.

Dans cette approche, l'occurrence d'exceptions sur une hiérarchie d'héritage est pour partie gérée dans une perspective définitionnelle en distinguant une partie *défaut*, ensemble de propriétés qui admettent des exceptions, du *noyau* de la définition d'un concept de la hiérarchie. Cependant, la classification de concepts et la classification d'objets font toutes deux appel à un mécanisme d'héritage par défaut de propriétés. C'est donc finalement le point de vue inférentiel qui prime encore dans cette approche.

2.2.2.2. Le langage $\mathcal{AL}_{\delta\varepsilon}$

$\mathcal{AL}_{\delta\varepsilon}$ est un langage de représentation de la famille des logiques de descriptions [Coupey et Fouqueré, 97]. L'originalité de ce langage, par rapport aux autres logiques de descriptions et plus généralement à l'ensemble des formalismes dans lesquels des structures hiérarchiques avec exceptions sont représentées, réside en ce que le point de vue adopté pour la représentation des exceptions est purement définitionnel. Une hiérarchie d'héritage représentée dans le langage $\mathcal{AL}_{\delta\varepsilon}$ fait appel à un mécanisme d'héritage normal des propriétés et c'est dans la définition même des concepts de la hiérarchie que sont gérées les exceptions. Des propriétés par défaut sont envisagées comme des propriétés à part entière des concepts, et un concept est représenté par un unique concept de la TBox.

Plus précisément, le langage $\mathcal{AL}_{\delta\varepsilon}$ est doté de deux connecteurs unaires δ et ε qui permettent de définir dans la TBox des concepts avec des propriétés par défaut. Appliqués à un concept `C`, δ`C` et $\mathtt{C}^{\varepsilon}$ signifient respectivement “`C` par défaut” et “exception sur `C`”.

Dans l'exemple de la taxinomie des animaux contemporains, `gris` est ainsi une propriété par défaut d'`éléphant`, puisque certaines des instances de ce taxon ne sont pas de couleur grise ; en particulier, la définition d'`éléphant-royal`, dont les instances sont typiquement de couleur blanche, comporte une exception de la propriété `gris`. Cette exception de `gris` est elle-même une propriété par défaut d'`éléphant-royal` ; la définition d'`éléphant-royal-sale`, dont les instances sont de couleur grise, comporte en effet une exception de l'exception de `gris`. Formellement, soient `E`, `ER`, `ERS` et `Gr` les termes de la TBox de $\mathcal{AL}_{\delta\varepsilon}$ qui représentent les concepts `éléphant`, `éléphant-royal`, `éléphant-royal-sale` et `gris`; alors $\delta\mathtt{Gr}$, $\delta((\delta\mathtt{Gr})^{\varepsilon})$ et $(\delta((\delta\mathtt{Gr})^{\varepsilon}))^{\varepsilon}$ interviennent dans la définition respectivement de `E`, `ER` et `ERS`. Les définitions complètes des taxons qui nous intéressent sont représentées sur la figure 2.9. ci-dessous.

$$\mathtt{AC} \equiv \mathtt{A} \cap \mathtt{C}$$

$$\mathtt{E} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta\mathtt{Gr}$$

$$\mathtt{EA} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \mathtt{Gr}$$

$$\mathtt{ER} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta((\delta\mathtt{Gr})^{\varepsilon}) \cap \delta\mathtt{Bl}$$

$$\mathtt{ERS} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap (\delta((\delta\mathtt{Gr})^{\varepsilon}))^{\varepsilon} \cap (\delta\mathtt{Bl})^{\varepsilon}$$

figure 2.9. TBox de $\mathcal{AL}_{\delta\varepsilon}$ représentant la taxinomie des animaux contemporains

Les connecteurs δ et ε possèdent des propriétés de simplification d'écriture. Intuitivement, elles expriment le fait que les chaînes de défauts sont supprimables et que l'exception d'une propriété est implicitement l'exception d'une propriété par défaut. Ces simplifications d'écriture sont énoncées comme suit :

$$\delta(\delta C) = \delta C$$

$$(\delta C)^{\varepsilon} = C^{\varepsilon}$$

Sur la figure 2.10. ci-dessous, les définitions des termes qui représentent la taxinomie des animaux contemporains sont simplifiées en appliquant ces propriétés des connecteurs.

$$\mathtt{AC} \equiv \mathtt{A} \cap \mathtt{C}$$

$$\mathtt{E} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta\mathtt{Gr}$$

$$\mathtt{EA} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \mathtt{Gr}$$

$$\mathtt{ER} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta(\mathtt{Gr}^{\varepsilon}) \cap \delta\mathtt{Bl}$$

$$\mathtt{ERS} \equiv \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap (\mathtt{Gr}^{\varepsilon})^{\varepsilon} \cap \mathtt{Bl}^{\varepsilon}$$

figure 2.10. TBox simplifiée représentant la taxinomie des animaux contemporains

Une autre propriété du connecteur δ est sa distributivité par rapport au connecteur $\cap$; un défaut sur une conjonction de concept est la conjonction de ces concepts par défaut :

$$\delta(C_1 \cap C_2) = \delta C_1 \cap \delta C_2 .$$

Les concepts d'une TBox sont partiellement ordonnés par la relation de *subsomption descriptive* $\subset$: $C \subset D$ ssi $C \cap D = C$. Pour tout concept C, les relations de subsomption entre C et δC et entre C^{ε} et δC sont caractérisées par les propriétés ci-dessous de δ et ε :

$$C \cap \delta C = \delta C$$

$$C^{\varepsilon} \cap \delta C = C^{\varepsilon}$$

Reprenons l'exemple de la taxinomie des animaux contemporains. L'ensemble des propriétés décrites ci-dessus assurent les relations de subsomption attendues entre les concepts définis sur la figure 2.11. : entre `AC` et `E`, `E` et `EA`, `E` et `ER`, `E` et `ERS`, `ER` et `ERS`. Détaillons la démonstration de celles problématiques, entre `E` et `ER` d'une part et entre `ER` et `ERS` d'autre part. Il suffit de montrer que $\mathtt{ER} \cap \mathtt{E} = \mathtt{ER}$ et $\mathtt{ERS} \cap \mathtt{ER} = \mathtt{ERS}$, eu égard à la définition de la relation de subsomption :

$$\begin{aligned} ER \cap E &= \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta(\mathtt{Gr}^{\varepsilon}) \cap \delta\mathtt{Gr} = \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta(\mathtt{Gr}^{\varepsilon}) \cap \delta(\delta\mathtt{Gr}) \\ &= \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta(\mathtt{Gr}^{\varepsilon} \cap \delta\mathtt{Gr}) = \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap \delta(\mathtt{Gr}^{\varepsilon}) = \mathtt{ER} \end{aligned}$$

$$\begin{aligned} \mathtt{ERS} \cap \mathtt{ER} &= \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap (\mathtt{Gr}^{\varepsilon})^{\varepsilon} \cap \mathtt{Bl}^{\varepsilon} \cap \delta(\mathtt{Gr}^{\varepsilon}) \\ &= \mathtt{A} \cap \mathtt{C} \cap \mathtt{Tr} \cap (\mathtt{Gr}^{\varepsilon})^{\varepsilon} \cap \mathtt{Bl}^{\varepsilon} = \mathtt{ERS} \end{aligned}$$

La mise en évidence de ces relations de subsomption entre les termes `AC`, `E`, `EA`, `ER` et `ERS` permet de simplifier encore les définitions de ces termes qui représentent la taxinomie des animaux contemporains en intégrant dans la définition d'un terme les termes les plus spécifiques qui le subsument. Le terme `AC` qui représente le taxon `animal-contemporain` est utilisé dans la définition du terme `E` qui représente `éléphant`, `E` est lui-même utilisé dans la définition du terme `ER` qui représente le taxon `éléphant-royal`, etc. La figure 2.11. ci-dessous rassemble l'ensemble des définitions que nous obtenons ainsi des termes utilisés pour représenter la taxinomie des animaux contemporains.

$$\mathtt{AC} \equiv \mathtt{A} \cap \mathtt{C}$$

$$\mathtt{E} \equiv \mathtt{AC} \cap \mathtt{Tr} \cap \delta\mathtt{Gr}$$

$$\mathtt{EA} \equiv \mathtt{E} \cap \mathtt{Gr}$$

$$\mathtt{ER} \equiv \mathtt{E} \cap \delta(\mathtt{Gr}^{\varepsilon}) \cap \delta\mathtt{Bl}$$

$$\mathtt{ERS} \equiv \mathtt{ER} \cap (\mathtt{Gr}^{\varepsilon})^{\varepsilon} \cap \mathtt{Bl}^{\varepsilon}$$

figure 2.11. TBox simplifiée représentant la taxonomie des animaux contemporains

Cet exemple de la taxinomie des animaux contemporains met en évidence l'adéquation du langage $\mathcal{AL}_{\delta\varepsilon}$ pour la représentation de hiérarchies de concepts avec exceptions.

Le processus de classification de concept associé à $\mathcal{AL}_{\delta\varepsilon}$ ne dépend que de la relation de subsomption décrite ci-dessus, il utilise donc un algorithme standard de classification [Coupey et Fouqueré, 94a]. Les concepts les plus spécifiques qui subsument le concept *C* à classer et ceux les plus généraux subsumés par *C* sont recherchés en ne faisant appel qu'à la relation de subsomption standard décrite ci-dessus. En d'autres termes, aucune relation de subsomption par défaut, aucun mécanisme de classification par défaut ne sont utilisés.

La classification d'objet est désignée dans [Coupey et Fouqueré, 94b] par l'expression “test d'instance” . Le test d'instance se ramène normalement à la classification du concept correspondant à la description de l'objet pour lequel on recherche les concepts dont il est instance. Il est raffiné dans le langage $\mathcal{AL}_{\delta\varepsilon}$ par la distinction entre instances sûres et instances probables et entre instances typiques et instances exceptionnelles.

Le calcul des propriétés héritées d'un concept dans le langage $\mathcal{AL}_{\delta\varepsilon}$ ne fait appel à aucun mécanisme de raisonnement par défaut et n'intervient pas dans la classification d'objet [Coupey et Fouqueré, 94b]. La démarche utilisée est purement définitionnelle.

2.2.3. Discussion

L'étude des principales approches pour l'introduction de défauts et exceptions dans différents formalismes de représentation met en évidence deux points de vue alternatifs ou complémentaires pour la gestion des exceptions sur une hiérarchie de concepts : l'héritage par défaut de propriétés et la définition de propriétés par défaut. Le point de vue de l'héritage par défaut relève d'une démarche inférentielle qui consiste à prendre en compte les exceptions sur une taxinomie dans les mécanismes de classification de ses concepts. Le point de vue de la définition de propriétés par défaut relève d'une démarche définitionnelle qui consiste au contraire à représenter les exceptions existant sur une taxinomie dans la définition même de ses concepts, les mécanismes de classification restant inchangés.

L'intérêt d'une démarche définitionnelle réside dans la capacité étendue de définir des concepts et en particulier des concepts naturels tels que ceux des taxinomies naturelles. Or notre but est précisément de définir les concepts d'un domaine et les relations entre eux et d'expliciter ainsi les primitives de langage nécessaires à la description du domaine. En effet, nous nous intéressons à établir des définitions de taxons dans la perspective de les utiliser ensuite pour permettre l'identification des spécimens, c'est-à-dire la détermination, pour un spécimen donné, de l'espèce dont il est une instance (cf. § 1.1.2.). Pour ce faire, nous nous attachons à expliciter un langage de description des taxons à même de discriminer ceux d'un même rang, comme par exemple les différents genres de la famille des orchidées, les différentes espèces d'un même genre d'orchidées (cf. § 1.4.). De ce point de vue, une démarche définitionnelle est mieux adaptée à notre tâche d'explicitation. Pour représenter une taxinomie comportant des exceptions, c'est donc le support sur lequel sont construits les graphes conceptuels qui décrivent les taxons que nous nous proposons de réviser - plutôt que la relation de généralisation entre graphes (cf. § 2.1.2.).

Une théorie des défauts ou un réseau sémantique permettent de représenter une hiérarchie de concepts *atomiques*, de sorte qu'un concept qui définit par exemple un taxon a le même statut qu'un concept qui définit une propriété de ce taxon : un noeud du réseau sémantique, une proposition logique. La description d'un taxon est obtenue par le calcul des propriétés dont il hérite dans la hiérarchie, de sorte que le point de vue inférentiel est naturel à ces formalismes. L'adoption d'une démarche définitionnelle pour la gestion des exceptions sur des structures hiérarchiques va au contraire de paire avec le choix d'un formalisme qui permette d'élaborer des descriptions qui *définissent* les concepts du domaine et de munir ces définitions d'une relation d'ordre qui représente la taxinomie du domaine. La description associée à un taxon est alors sa définition. Ces caractéristiques sont celles des logiques de descriptions ; elles sont aussi celles du modèle des graphes conceptuels.

Nous proposons dans la partie suivante d'adopter la démarche définitionnelle proposée par Coupey et Fouqueré, pour enrichir le modèle des graphes conceptuels afin de permettre la représentation de taxinomies comportant des exceptions dans ce formalisme.

2.3. Une extension du modèle des Graphes Conceptuels pour la gestion des défauts et exceptions à l'héritage

Nous exposons dans cette partie notre contribution à la gestion d'exceptions sur une taxinomie représentée dans le modèle des graphes conceptuels. Notre approche s'inspire des travaux de Coupey et Fouqueré menés dans le cadre du langage de descriptions $\mathcal{AL}_{\delta\varepsilon}$ (cf. § 2.2.2.2.). Nous adoptons comme eux une démarche définitionnelle pour représenter explicitement dans la définition des concepts le caractère par défaut de certains traits descriptifs et les exceptions à ceux-ci [Faron et Ganascia, 97a]. Ces nouvelles capacités définitionnelles sont fondées sur un enrichissement du support sur lequel sont construits les graphes conceptuels.

2.3.1. Enrichissement du support S

2.3.1.1. Définition de $S_{\delta\varepsilon}$

Pour gérer défauts et exceptions dans le modèle des graphes conceptuels simples, nous proposons un enrichissement du support S par l'introduction, sur l'ensemble T_C des types de concepts, de deux connecteurs unaires δ et ε ayant la même sémantique que ceux du langage de descriptions $\mathcal{AL}_{\delta\varepsilon}$ [Coupey et Fouqueré, 97]. Étant donné un type de concept t, δt et εt sont des types de concept qui signifient respectivement "t par défaut" et "exception sur t". Le terme "exception" est pris ici au sens premier d'une "action par laquelle on excepte" [Le Robert, 94] et ce qui peut être excepté est précisément ce qui est "par défaut".

Définition 2.1. [$S_{\delta\varepsilon}$] : *$S_{\delta\varepsilon}$ est un support S (cf. définition 1.3.) dont l'ensemble des types de concepts T_C est muni de deux connecteurs unaires δ et ε qui vérifient :*

(i) pour tout type t élément de T_C, $t \leq \delta t$

(ii) pour tout type t élément de T_C, $\varepsilon t \leq \delta t$

(iii) pour tout type t_1 et tout type t_2 éléments de T_C, si $t_1 \leq t_2$ alors $\delta t_1 \leq \delta t_2$

(iv) pour tout type t élément de T_C, $\delta(\delta t) = \delta t$

(v) pour tout type t élément de T_C, $\varepsilon(\delta t) = \varepsilon t$

Les hypothèses *(i)* et *(ii)* caractérisent les relations d'ordre entre les types de concepts t et δt, d'une part, et les types de concepts εt et δt, d'autre part : t et εt sont des sous-types de δt. Intuitivement, cela rend compte du fait que sous la désignation de, par exemple, “gris par défaut” sont entendus des “gris” ou des “exceptions sur gris”.

L'hypothèse *(iii)* indique que deux types de concepts qui représentent des traits descriptifs par défaut sont comparables si les types de concepts qui représentent les traits stricts correspondants le sont eux-mêmes.

Les hypothèses *(iv)* et *(v)* consistent en des simplifications d'écriture. D'après l'hypothèse *(iv)*, les chaînes de connecteurs δ peuvent être supprimées : quelle que soit leur longueur, elles expriment toutes le caractère “par défaut” d'un “type de base”. D'après l'hypothèse *(v)*, appliquer le connecteur ε à un type par défaut revient à l'appliquer sur le type de base correspondant. Intuitivement, une exception (action) est représentée par l'application du connecteur ε au type de concept *à excepter* et seul peut être excepté ce qui est par défaut.

Remarque 2.1. : D'après les propriétés *(i)* et *(ii)*, un sommet concept d'étiquette $(\delta t, m)$ est plus général qu'un sommet d'étiquette (t, m) ou $(\varepsilon t, m)$. Remplacer δt par t ou εt dans l'étiquette d'un concept, c'est opérer une restriction de concept.

Remarque 2.2. : D'après les propriétés *(ii)* et *(v)*, appliquer le connecteur ε au type δt de l'étiquette $(\delta t, m)$ d'un sommet concept revient à restreindre ce concept en remplaçant dans son étiquette le type δt par son sous-type εt.

Remarque 2.3. : L'ensemble des types de concepts muni des connecteurs δ et ε n'est pas un ensemble tri-partitionné de la forme ($T_C + \delta T_C + \varepsilon T_C$), avec T_C l'ensemble des “types de base”. Les connecteurs δ et ε ne s'appliquent en effet pas seulement sur les “types de base”. Les hypothèses *(iv)* et *(v)* ne permettent pas de ramener toutes les écritures à l'application d'un connecteur sur un “type de base”. $\delta\varepsilon t$ ne peut être simplifié, ni $\varepsilon\varepsilon t$, ni plus généralement, aucune chaîne d'exceptions.

Propriété 2.1. $\forall t \in T_c, \varepsilon\varepsilon t \leq \delta\varepsilon$.

Démonstration. Cela est immédiat en posant $t' = \varepsilon t$ et en appliquant l'hypothèse *(ii)* de la définition 2.1.

Propriété 2.2. $\forall t \in T_c, \delta\varepsilon t \leq \delta t$.

Démonstration. D'après l'hypothèse *(iv)* de la définition 2.1, $\delta t = \delta\delta t$. On se ramène ainsi à montrer que $\delta\varepsilon t \leq \delta\delta t$. D'après l'hypothèse *(iii)*, si $\varepsilon t \leq \delta t$ alors $\delta\varepsilon t \leq \delta\delta t$. Or l'hypothèse *(ii)* établit que $\varepsilon t \leq \delta t$ donc $\delta\varepsilon t \leq \delta\delta t$ donc $\delta\varepsilon t \leq \delta t$.

Afin de justifier les propriétés que nous donnons dans la définition de $S_{\delta\varepsilon}$ aux connecteurs δ et ε, reprenons l'exemple de la taxinomie des animaux contemporains. Le support $S_{\delta\varepsilon}$ permet de construire des descriptions de ses taxons dans lesquelles sont explicitement expimés des exceptions et des défauts. L'ensemble des hypothèses posées dans la définition de $S_{\delta\varepsilon}$ doit permettre d'établir des relations de généralisation entre graphes de définition telles qu'elles représentent la taxinomie considérée.

a) Définition de nouveaux types de concepts

La figure 2.1.2 ci-dessous présente les définitions de taxons d'animaux contemporains qui peuvent être construites sur $S_{\delta\varepsilon}$.

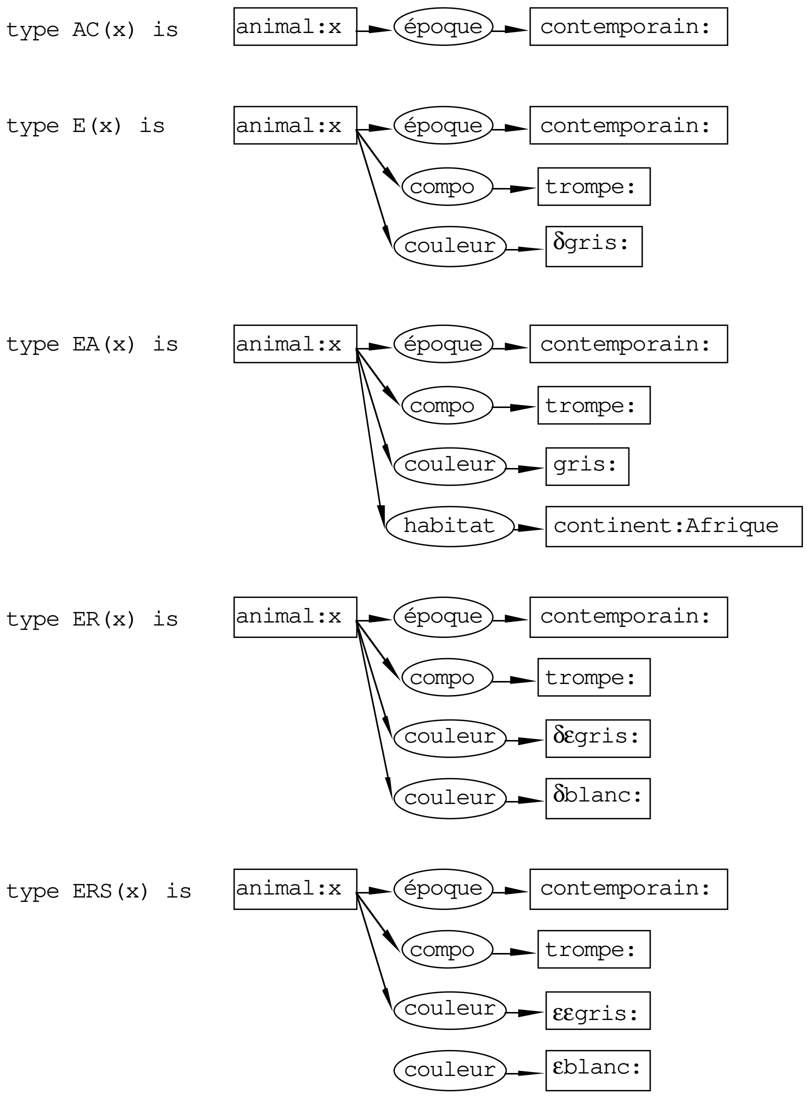


figure 2.12. Taxinomie des animaux contemporains : nouvelles définitions de taxons.

Dans le graphe de définition du type `E` qui représente le taxon `éléphant`, le type δ`gris` exprime le “gris par défaut”. Certaines instances - exceptionnelles - d’`éléphant`, telles que celles d’`éléphant-royal`, ont en effet pour couleur non pas le gris mais le blanc. De même, dans le graphe de définition du type `ER` qui représente le taxon `éléphant-royal`, nous utilisons le type δ`blanc` pour exprimer “blanc par défaut”, certaines des instances exceptionnelles d'`éléphant-royal`, telles que celles d'`éléphant-royal-sale`, ayant pour couleur non pas le blanc mais le gris.

Poursuivons sur cet exemple. Il n’est pas suffisant d’exprimer dans la description associée au taxon `éléphant-royal` que la couleur est, par défaut, le blanc ; il faut encore préciser que cette couleur est une exception au gris par défaut du taxon père `éléphant`. C’est ce qui distingue la définition d'`éléphant-royal` de celle de `mouche-verte`, par exemple : celle-ci, contrairement à celle-là, ne constitue pas une exception au `gris` d’`éléphant`. Dans le graphe de définition du type `ER`, nous mettons donc le sommet concept [`animal`:] en relation `couleur` non seulement avec [δ`blanc:`] mais aussi avec [δε`gris`:]. L'exception du gris est ainsi exprimée par l'application du connecteur ε au type `gris`. L'application du connecteur δ au type ε`gris` ainsi construit exprime en outre le caractère par défaut de cette exception, certaines instances d'`éléphant-royal` telles que celles d'`éléphant-royal-sale` y faisant elles-mêmes exception. Quant au graphe de définition du type `ERS` qui représente ce dernier taxon, son sommet concept [`animal`:] est en relation `couleur` avec [ε`blanc`:] et [εε`gris`:] : il fait exception au blanc et à l'exception sur le gris.

b) Propriétés des connecteurs δ et ε et relations de généralisation

Sur l'exemple de la taxinomie des animaux contemporains, à partir des types de concepts que nous venons de définir, nous pouvons maintenant justifier les propriétés données aux connecteurs δ et ε dans la définition de $S_{\delta\varepsilon}$ (définition 2.1.). Nous allons montrer pour cela qu'elles assurent les relations de généralisation entre types attendues, celles qui permettent de représenter la taxinomie considérée.

Les relations de généralisation entre les graphes de définition des types `AC` et `E` qui représentent les taxons `animal-contemporain` et `éléphant` d'une part, et entre ceux des types `E` et `EA` qui représentent `éléphant` et `éléphant-africain` d'autre part sont conservées. En effet, le graphe de définition de `E` est dérivable à partir de celui de `AC` par la séquence de deux joints externes, (1) entre le graphe de définition du type `AC` et le graphe [`animal`:]->(`compo`)->[`trompe`:] puis (2) entre le graphe résultat et le graphe [`animal`:]->(`couleur`)->[δ`gris`:]. La relation `E`≤`AC` est donc bien conservée.

Quant au nouveau graphe de définition du type `EA`, la figure 2.13. ci-après montre qu'il est maintenant dérivable à partir du graphe de définition du type `E` par la séquence d'une restriction de concept et d'un joint externe. La diminution du type δ`gris` remplacé par son sous-type `gris` est une restriction du sommet concept [δ`gris`:] du graphe de définition de `E` (cf. Remarque 2.1.). Un joint est ensuite opéré entre le graphe résultat et le graphe [`animal`:]->(`habitat`)->[`continent:Afrique`]. La relation `EA≤E` est donc également bien conservée.

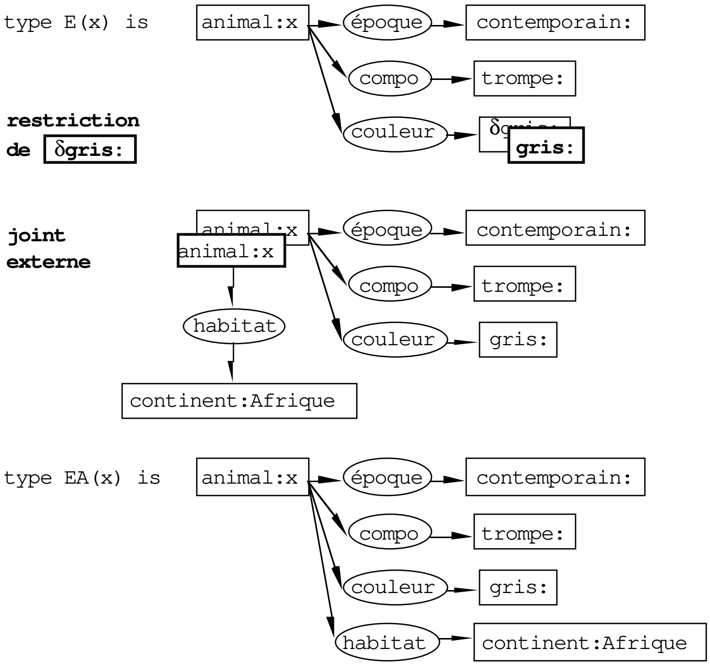


figure 2.13. Dérivation du graphe de définition de `EA` à partir de celui de `E`.

Regardons maintenant ce qu'il en est des relations de généralisation entre les types `E` et `ER` qui représentent `éléphant` et `éléphant-royal` et entre les types `ER` et `ERS` qui représentent `éléphant-royal` et `éléphant-royal-sale`. C'est en effet l'organisation hiérarchique de ces taxons qui était impossible à représenter en construisant les graphes de définition des types `E`, `ER` et `ERS` sur un support `S` (cf. § 2.1.1.).

La figure 2.14. ci-contre montre que le graphe de définition de `ER` est maintenant dérivable à partir de celui de `E` par la séquence (1) d'une restriction de concept : diminution de l'étiquette du sommet [δ`gris`:] en remplaçant le type δ`gris` par son sous-type δε`gris` (cf. propriété 2.2.) ; et (2) d'un joint externe entre le graphe résultat et le graphe [`animal`:]->(`couleur`)->[δ`blanc`:].

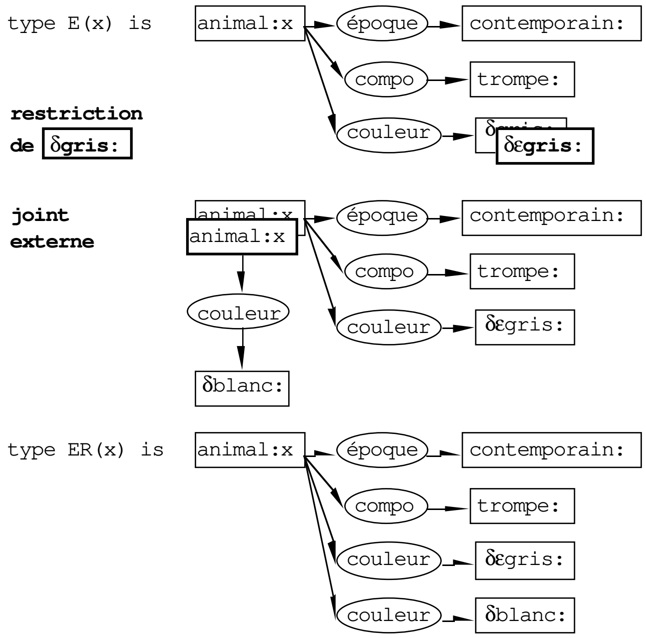


figure 2.14. Dérivation du graphe de définition de `ER` à partir de celui de `E`.

La figure 2.15. ci-dessous montre que le graphe de définition de `ERS` est maintenant dérivable à partir de celui de `ER` par la séquence de deux restrictions de concept : (1) l'étiquette du sommet [δε`gris`:] est diminuée en remplaçant δε`gris` par son sous-type εε`gris` (cf. propriété 2.1.) et (2) l'étiquette du sommet [δ`blanc`:] est diminuée en remplaçant δ`blanc` par son sous-type ε`blanc`.

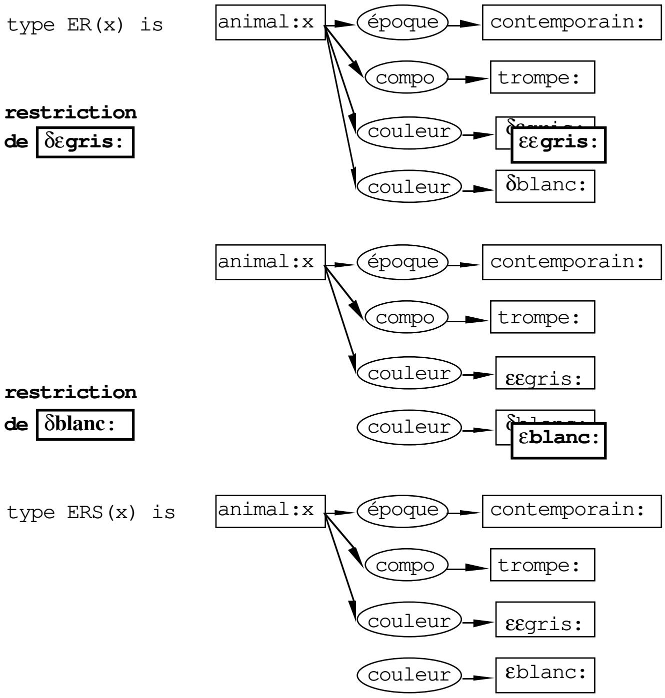


figure 2.15. Dérivation du graphe de définition de `ERS` à partir de celui de `ER`.

Construits sur le support $S_{\delta\varepsilon}$, le graphe de définition de ER est donc maintenant une spécialisation de celui de E et celui de ERS une spécialisation de celui de ER. Ces nouvelles définitions permettent donc d'établir les relations attendues ER≤E et ERS≤ER.

Ainsi, les propriétés des connecteurs δ et ε établies dans la définition 2.1. du support $S_{\delta\varepsilon}$ permettent de construire de nouvelles définitions de types telles que les relations de généralisation entre, par exemple, les types qui représentent les taxons de la taxinomie des animaux contemporains correspondent à l'organisation hiérarchique de ces taxons. En d'autres termes, la hiérarchie de généralisation sur les types de concepts ainsi définis coïncide bien avec la structure de la taxinomie considérée et donc la représente.

2.3.1.2. Le processus d'exception par spécialisation

En considérant maintenant la *construction* de la représentation de la taxinomie des animaux contemporains et en généralisant cet exemple, nous pouvons maintenant décrire le processus d'exception. Les définitions de types étant construites sur $S_{\delta\varepsilon}$, l'exception d'une partie de la description d'un taxon dans la description d'un taxon de rang inférieur s'exprime en termes d'une séquence d'opérations de spécialisation. Plus précisément, nous distinguons deux phases successives, à partir d'une phase initiale.

0. Un défaut est représenté par l'application du connecteur δ à un type de concept. C'est donc une opération de généralisation, par augmentation de concept. En effet, eu égard à la relation d'ordre entre δt et t, un sommet concept d'étiquette (δt,m) est plus général qu'un sommet d'étiquette (t,m). Il peut donc être augmenté en remplaçant dans son étiquette t par δt.

1. Une exception (action) est représentée par l'application du connecteur ε au type de concept *à excepter*. Elle n'a donc pour objet que ce qui peut être excepté, c'est-à-dire un défaut. L'exception d'une propriété est donc une opération de spécialisation par restriction de concept. En effet, eu égard à la relation d'ordre entre δt et εt, un sommet concept d'étiquette (δt,m) est plus général qu'un sommet d'étiquette (εt,m). Il peut donc être restreint en remplaçant dans son étiquette δt par εt. Dans notre exemple, ce sont le concept [δgris:] du graphe de définition du type E et les concepts [δεgris:] et [δblanc:] du graphe de définition du type ER qui sont restreints (figures 2.15 et 2.16).

2. L'action d'exception peut être accompagnée par la précision de ce qui exclut, responsable de l'exception. Cette étape consiste encore en une opération de spécialisation, par joint externe avec un graphe décrivant ce qui exclut. Dans notre exemple, il s'agit du joint entre le graphe de définition de ER et [animal:x]->(couleur)->[δblanc:] (figure 2.15).

Il apparaît ainsi que le mécanisme de définition de types du modèle des graphes conceptuels, fondé sur les opérations de spécialisation de graphe rendent bien compte du processus d'exception : une exception est la diminution d'un type de concept par défaut qui est excepté, accompagnée éventuellement de la précision de ce qui fait exception par un joint.

2.3.1.3. $S_{\delta\varepsilon}$ et $\mathcal{AL}_{\delta\varepsilon}$

L'ensemble constitué d'un support et d'un mécanisme de définition de types dans le modèle des graphes conceptuels est comparable à la TBox d'un langage de descriptions. En particulier, $S_{\delta\varepsilon}$ et $\mathcal{AL}_{\delta\varepsilon}$ sont comparables, qui permettent tous deux de définir des défauts et des exceptions sur des types de concepts ou des termes définis. Et en effet, les propriétés des connecteurs δ et ε de $S_{\delta\varepsilon}$ et $\mathcal{AL}_{\delta\varepsilon}$ sont similaires.

Considérons les hypothèses posées dans la définition 2.1. du support $S_{\delta\varepsilon}$. Seule l'hypothèse *(iii)* ne se retrouve pas dans le cadre de $\mathcal{AL}_{\delta\varepsilon}$. Sa justification apparaît dans la démonstration de la propriété 2.2, pour établir que $\delta\varepsilon t \leq \delta\delta t$. Dans $\mathcal{AL}_{\delta\varepsilon}$, c'est la distributivité du connecteur $\cap$ par rapport au connecteur δ qui permet d'établir ce résultat.

Les hypothèses de simplification *(iv)* et *(v)* sont identiques à celles faites dans le cadre de la TBox de $\mathcal{AL}_{\delta\varepsilon}$. Les hypothèses (i) et (ii) se déduisent, dans $\mathcal{AL}_{\delta\varepsilon}$, des hypothèses $\delta t \cap t = t$ et $\delta t \cap \varepsilon t = \varepsilon t$. Cependant, le modèle des graphes conceptuels permet de les justifier intuitivement par leur traduction en termes d'opérations de spécialisation (cf. remarques 2.1 et 2.2). Comme nous l'avons décrit dans la partie précédente, c'est le processus même d'exception qui peut ainsi être traduit de manière intelligible en termes de restriction de concept et de joint (cf. § 2.3.1.2.) - tandis que la notion *d'opération* de spécialisation est absente dans les logiques de descriptions.

Une autre différence qui apparaît entre $S_{\delta\varepsilon}$ et $\mathcal{AL}_{\delta\varepsilon}$ est liée au mécanisme de définition d'un type de concept dans $S_{\delta\varepsilon}$ ou d'un concept dans $\mathcal{AL}_{\delta\varepsilon}$. Dans cette partie, nous avons limité l'application des connecteurs δ et ε à des types de concepts atomiques. Nous définissons dans la partie suivante l'application d'un défaut à un type défini, comme cela est proposé dans $\mathcal{AL}_{\delta\varepsilon}$.

2.3.2. Types de concept par défaut définis

Dans cette partie, nous revenons sur le mécanisme de définition de types de concepts (cf. § 1.3.3.). Nous étendons ce mécanisme au cas où le type à définir est un type par défaut (§ 2.3.2.1.). Nous comparons ensuite la définition d'un type par défaut sur $S_{\delta\varepsilon}$ et celle d'un concept par défaut sur $\mathcal{AL}_{\delta\varepsilon}$ (§ 2.3.2.2.).

2.3.2.1. Définition et insertion de types par défaut

a) Définition d'un type par défaut

Nous avons présenté dans le chapitre 1 de cette thèse le mécanisme de définition de type de concept sur S tel qu'il est proposé dans [Leclère, 95] (cf. définition 1.5.). Ce mécanisme s'applique encore pour définir sur $S_{\delta\varepsilon}$ des “types de base”. Nous proposons de l'étendre à la définition sur $S_{\delta\varepsilon}$ de types de concepts par défaut. Dans la définition 2.2. ci-dessous, le genre et la différence d'un type par défaut sont définis à partir du genre et de la différence du type de base correspondant.

Définition 2.2. [Définition d'un type par défaut, Genre et Différence] :

Soit `t` *un type de concept d'un support* $S_{\delta\varepsilon}$ *défini par* `type t(x) is G` *; alors* `δt` *est défini par* `type δt(x) is G'`*, avec :*

genre(`δt`*) = genre(*`t`*) ;*

la différence `G'` *est obtenue à partir de celle de* `t`*, en appliquant aux types des étiquettes de tous les sommets concepts de* `G`*, sauf sa tête, le connecteur δ.*

Reprenons l'exemple de la taxinomie des animaux contemporains et imaginons un nouveau taxon `noname` défini par le type de concept `N`. Comme le montre la figure 2.16 ci-dessous, le type `N` a pour genre le type `AC`, c'est-à-dire le même genre que le type `E` (cf. figure 2.12). Son graphe différence est obtenu à partir de celui du type `E` en appliquant le connecteur δ aux types `trompe` et `gris`. Par définition, le type de concept `N` est le type `δE`.

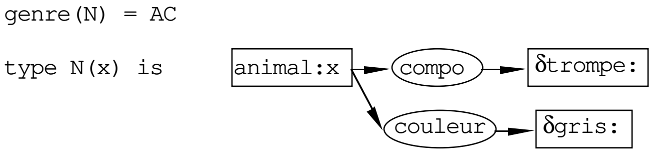


figure 2.16. Définition du type de concept `δE`.

b) Classification d'un défaut défini

D'après le protocole d'insertion établi par Leclère (cf. § 1.3.1.4.), tout type de concept est introduit juste sous son plus petit généralisé et au-dessus de son plus grand spécialisé. Dans le cas d'un type de concept défini, on peut restreindre la recherche de son plus petit généralisé aux types plus spécifiques que son genre. Tout type de concept défini est en effet une spécialisation de son genre.

Dans le cas particulier où le type défini est un défaut δt, celui-ci est introduit dans le support sous le genre de t, qui est aussi le sien. On peut en outre préciser sa place au-dessus de t, le graphe différence de t étant dérivable à partir de celui de δt par la séquence des restrictions de ses sommets concepts, par diminution de leurs types δt_i remplacés par t_i. Dans l'exemple, le type de concept δE est donc introduit sous le type `AC` et au-dessus du type `E`.

2.3.2.2. $S_{\delta\varepsilon}$ et $\mathcal{AL}_{\delta\varepsilon}$

La définition d'un défaut δt sur $\mathcal{AL}_{\delta\varepsilon}$ consiste, nous l'avons vu, à appliquer le connecteur δ sur tous les types de concepts dont la conjonction définit t (cf. § 2.2.2.2.). En effet, la notion de genre et de différence n'existe pas dans les logiques de description, un terme étant défini par conjonction de termes. Le mécanisme de définition d'un défaut fait simplement appel à la propriété de distributivité du connecteur δ par rapport au connecteur $\sqcap$. Par exemple, dans la définition du défaut δE dans $\mathcal{AL}_{\delta\varepsilon}$, les défauts porteraient sur `trompe`, sur `gris` mais aussi sur `AC`, c'est-à-dire sur `C` et sur `A`.

Au contraire, la définition d'un défaut δt sur $S_{\delta\varepsilon}$ consiste à appliquer le connecteur δ sur tous les types de concepts qui apparaissent dans la différence de δt *sauf sur le type du sommet tête*, c'est-à-dire le genre de δt. Intuitivement, cela exprime bien l'idée qu'un défaut δt est défini par rapport à un type strict t et plus précisément que les défauts qui participent à la définition de δt ne portent que sur les concepts propres à t, et pas sur ceux de son genre. En d'autres termes, les défauts ne portent que sur les types de la différence de t, qui différencient t de son genre. Par exemple, dans la définition du type δE, les défauts portent sur les types `trompe` et `gris` et non pas sur son genre `AC`, c'est-à-dire pas sur la description héritée d'`AC`, qui n'est pas propre à `E`.

2.4. Conclusion

Eu égard les similarités entre la TBox d'un langage de descriptions et le support du modèle des graphes conceptuels enrichi d'un mécanisme de définition de types, nous avons proposé une méthode de représentation de défauts et exceptions dans le modèle des graphes conceptuels qui étend la démarche définitionnelle établie par Coupey et Fouqueré dans le langage de descriptions $\mathcal{AL}_{\delta\varepsilon}$. Plus précisément, nous avons défini le support $S_{\delta\varepsilon}$ enrichi de deux connecteurs dont la sémantique et les propriétés sont comparables à celles des connecteurs de $\mathcal{AL}_{\delta\varepsilon}$. Les différences qui apparaissent entre $S_{\delta\varepsilon}$ et $\mathcal{AL}_{\delta\varepsilon}$ trouvent leur origine pour partie dans la nature des mécanismes de définition de type associés à ces deux formalismes. Les opérations de spécialisation de graphes définies sur $S_{\delta\varepsilon}$ permettent en outre de bien rendre compte du processus d'exception.

A l'issue de ce chapitre, nous sommes maintenant à même de représenter des taxinomies comportant des exceptions dans le formalisme des graphes conceptuels. Les bases de connaissances taxinomiques que nous pouvons ainsi produire sont telles que nous les avons décrites dans le chapitre 1 pour le cas de taxinomies “strictes” ; seul diffère le support $S_{\delta\varepsilon}$ sur lequel sont construits les graphes décrivant les taxons. Dans le chapitre suivant, nous décrivons les méthodes et algorithmes que nous mettons en oeuvre pour raisonner avec de telles connaissances. Il s'agit plus précisément d'acquérir des connaissances taxinomiques et d'identifier les objets du domaine à l'aide de ces connaissances.

Chapitre 3

Acquisition de connaissances taxinomiques

Dans ce chapitre, nous commençons par décrire le principe général de l'acquisition d'une taxinomie dont la structure est connue ; c'est une démarche descendante, qui procède par spécialisation, à l’inverse de celle habituelle de classification, qui procède par généralisation (§ 3.1.). Le principe de notre méthode est dérivé de celui proposé dans [Aïmeur et Ganascia, 93] ; il est basé sur l'utilisation des opérations de spécialisation définies dans le modèle des graphes conceptuels. Nous avons adapté cette méthode aux choix que nous avons fait en matière de représentation des connaissances. En particulier, nous l'avons étendue afin de pouvoir acquérir des descriptions de taxons comportant des exceptions. L'acquisition de la description d'un taxon se déroule en deux étapes : l'acquisition d'une description initiale à partir de celle du taxon père (§ 3.2.) puis le raffinement de cette description pour la discriminer de celles des taxons frères (§ 3.3.). Nous présentons enfin un algorithme de construction d'une hiérarchie de types représentant la taxinomie (§ 3.4.).

3.1. Principe général

3.1.1. Une démarche descendante

En Intelligence Artificielle, la construction automatique de classifications fait l’objet de travaux menés dans le domaine du *regroupement conceptuel*. Étant donné un ensemble de descriptions d'objets d'un domaine, la problématique du regroupement conceptuel est de regrouper ces descriptions selon leurs similarités pour former des concepts. La description d’un concept est de fait une généralisation des descriptions des objets que regroupe le concept. Les concepts ainsi formés sont partiellement ordonnés par la subsomption de sorte qu'une organisation hiérarchique du domaine apparaît. Le regroupement conceptuel s'intéresse donc à l'explicitation d'une taxinomie de concepts à partir d'un ensemble de descriptions d'objets.

Le problème de l'acquisition d'une taxinomie auquel nous nous intéressons est inverse de celui-ci. Étant donné un ensemble de concepts dont on connaît la taxinomie, c'est-à-dire les relations de subsomption qu'ils entretiennent, on cherche à acquérir cette taxinomie, c’est-à-dire à élaborer les descriptions de ses concepts. Plus précisément, nous nous intéressons à rendre compte de la taxinomie d'un ensemble de concepts en explicitant des descriptions de ces concepts dont les relations de généralisation rendent compte des relations de subsomption entre les concepts. En d'autres termes, alors que le regroupement conceptuel s'intéresse à l'explicitation d'une taxinomie qui organise un ensemble de descriptions d'objets, à l'inverse, notre problème est ici l'explicitation de descriptions de concepts telles que les relations de généralisation entre descriptions rendent compte de la taxinomie donnée des concepts.

La démarche que nous adoptons est de fait inverse de celle en regroupement conceptuel. Pour expliciter une taxinomie de concepts à partir d'un ensemble de descriptions d'objets, on *généralise* ces descriptions, selon une démarche *ascendante*, pour construire une hiérarchie de concepts. Pour expliciter des descriptions de concepts à partir de la taxinomie de ces concepts, nous les explicitons à partir de la racine de la taxinomie, selon une démarche *descendante*. Les descriptions de concepts sont explicitées les unes après les autres, par *spécialisation* des descriptions de leurs concepts pères. Nous proposons une méthode d'acquisition de connaissances taxinomiques dans le formalisme des graphes conceptuels qui est fondée sur les opérations de spécialisation définies dans ce formalisme. Elle est dérivée de celle proposée par Aïmeur et Ganascia dans METIS [Aïmeur et Ganascia, 93] [Aïmeur, 94]. Nous décrivons brièvement le principe de celle-ci dans la partie suivante.

3.1.2. La méthode d’acquisition de METIS

La méthode d'acquisition proposée dans METIS consiste en un processus d'acquisition incrémental d’une taxinomie. Les taxons sont décrits un à un, sous la forme d'un graphe conceptuel ; ils héritent chacun de la description de leur père dans la taxinomie puis cette description est raffinée en interrogeant l’expert du domaine. Plus précisément, le graphe qui décrit un taxon est construit à partir de celui décrivant le taxon père par *spécialisation*, à l'aide des opérations de spécialisation telles qu'elles sont définies dans le formalisme des graphes conceptuels. L'objectif est d'expliciter des descriptions de taxons *discriminantes*, dans le but de les utiliser dans des procédures d'identification de taxons. C'est la discrimination des taxons qui guide l'interrogation de l'expert tout au long du processus d'acquisition [Aïmeur et Ganascia, 93].

3.1.2.1. Initialisation

La phase d'initialisation du processus d'acquisition d'une taxinomie consiste à expliciter une description générique des objets du domaine. Cette description est construite par l'expert du domaine sous la forme d'un graphe conceptuel appelé “graphe primitif”. Formellement, il s'agit de définir une base canonique : le graphe primitif est déclaré canonique, c’est l’ancêtre commun à toutes les descriptions de taxons. De fait, toutes les descriptions de taxons sont construites par dérivation à partir de celui-là. Reprenons par exemple la taxinomie des orchidées décrite dans le chapitre 1 de cette thèse : la description minimale que l'on puisse faire de la famille des *Orchidaceae* est le graphe réduit à l'unique noeud concept `[orchidale:]`. Une telle description est augmentée incrémentalement (par joint) chaque fois que de nouveaux concepts doivent être introduits pour décrire un taxon nouvellement acquis.

A ce stade, le support `S` qui représente la terminologie du domaine est réduit aux seuls types de concepts et de relations nécessaires à la construction du graphe primitif. Le support est enrichi au fur et à mesure de l'acquisition de nouveaux taxons, chaque fois que de nouveaux types sont nécessaires à la description des taxons nouvellement acquis.

3.1.2.2. Complétion des descriptions des taxons

Chaque taxon nouvellement acquis est décrit à partir de la description de son père dans la taxinomie. Il hérite de celle-ci : sa description initiale est un graphe identique à celui qui décrit son père. Ce graphe est ensuite spécialisé au fur et à mesure de l'interrogation de l'expert pour préciser la description du taxon. Plus précisément, cette étape dite de complétion [Aïmeur, 94] consiste en une séquence de restrictions des types de concepts des arcs conceptuels du graphe de description qui représentent des traits caractéristiques génériques (cf. § 1.4.1). Les nouveaux types de concepts utilisés pour ces restrictions sont introduits dans le support `S` de sorte que celui-ci s’enrichit au fur et à mesure de l’acquisition de nouveaux taxons. Considérons par exemple à nouveau la taxinomie des orchidées. Le graphe de description de la tribu des *Vandeae* est construit à partir de celui de la sous-famille des *Orchidioideae* par restriction du type TEXTURE, remplacé par le type

DUR qui lui est inférieur (cf. figure 1.16.) ; DUR est donc introduit dans $\mathtt{T}_C$ sous TEXTURE. De tels traits caractéristiques génériques dans la description d'un taxon sont hérités du père de celui-ci ; ils sont restreints dans la description de tous les taxons frères de sorte qu'ils permettent de discriminer ces derniers par comparaison de leurs descriptions.

3.1.2.3. Extraction de traits descriptifs pertinents

Chaque taxon acquis est discriminé de ses frères déjà présents dans la taxinomie en recherchant un trait descriptif (ou une conjonction de traits) qui permette de distinguer sa description de celles de ses frères et qui soit jugé pertinent par l'expert. C'est l'étape dite d'extraction de traits pertinents [Aïmeur, 94]. Par exemple, la tribu des *Neottieae* étant acquise, l'acquisition de sa soeur, la tribu des *Vandeae*, est suivie d'une phase de discrimination de ces deux taxons. La comparaison de leurs graphes de description conduit à l'extraction de la texture des pollinies comme trait descriptif discriminant : elle est molle chez les Neottieae et dure chez les Vandeae. Formellement, les types MOU et DUR étant incomparables, les spécialisations du graphe [POLLINIE:] -> (DE_TEXTURE) -> [TEXTURE:] par restriction du type TEXTURE sont des sous-graphes des descriptions des tribus *Neottieae* et *Vandeae* qui les rendent elles-mêmes incomparables (cf. § 1.4.2.). L'expert est alors interrogé quant à la pertinence du trait descriptif extrait automatiquement des descriptions des taxons à discriminer. Il approuve ou non le trait descriptif qui lui est proposé ; dans la négative, un autre trait descriptif discriminant est recherché.

Dans le cas où un trait descriptif a été précédemment retenu pour discriminer un ensemble de taxons, lors de l'acquisition d'un nouveau taxon frère de ces derniers, il faudra d'abord vérifier si ce trait permet encore de discriminer le nouveau taxon de ses frères. Considérons par exemple l'acquisition de la description de la tribu des *Epidendreae*, une fois acquises les descriptions des tribus soeurs des *Vandeae* et des *Neottieae*. Après la phase de complétion du graphe de description des *Epidendreae* consistant à restreindre le type TEXTURE par le type DEMI-MOU, la texture des pollinies précédemment retenue pour discriminer les descriptions des *Vandeae* et des *Neottieae* apparaît encore pertinente pour discriminer les *Epidendreae* de ces derniers. Dans le cas contraire, il faut rechercher un nouveau trait descriptif pertinent pour discriminer le dernier taxon acquis de ses frères ; ce trait sera mis en conjonction avec le ou les traits précédemment retenus pour discriminer ses frères entre eux.

En cas d'échec de la tentative de discrimination automatique du dernier taxon décrit, l'expert est requis pour augmenter la description de ce taxon, ainsi que celle de son père et celles de ses frères. Leurs graphes de description sont spécialisés par joint avec un graphe représentant un nouveau trait descriptif générique. La restriction du nouveau type de concept introduit sur les frères doit alors permettre de discriminer ces derniers. Le graphe de description d'un taxon n'est donc pas construit “une fois pour toutes”. Au contraire, il est raffiné par spécialisation chaque fois que cela est nécessaire, au moment de l'introduction d'un taxon frère ou fils, pour le discriminer de ce frère ou discriminer ses fils.

3.1.2.4. Raffinement des descriptions

L'objectif d'une description étant de discriminer un taxon de ses frères, les traits descriptifs du père d'un taxon suffisent à le décrire tant que celui-ci n'a pas de frère dans la taxinomie en cours de construction. De fait, le graphe de définition du type qui le représente est une copie du graphe de définition du type représentant son père. Revenons par exemple à l'acquisition des tribus des *Neottieae* et des *Vandeae*. Si la tribu des *Neottieae* est la première à être acquise, sa description est alors celle de son père, la sous-famille des *Orchidioideae*. La texture des pollinies n'apparaît pas encore dans les descriptions de ces deux taxons ; aucun trait descriptif générique n'a donc à être précisé lors de la phase de complétion. Ce n'est qu'ultérieurement que les descriptions de la tribu des *Neottieae* et la sous-famille des *Orchidioideae* seront précisées, lors de l'introduction de la tribu des *Vandeae*, pour que les deux tribus soient discriminées.

Lors de l'acquisition d'un taxon `T` qui, à l'issue de la phase de complétion de sa description, ne peut être discriminé de ses frères déjà acquis, la phase d'extraction de traits pertinents échoue. La description du taxon `T` ainsi que celles de ses frères et celle de leur père doivent alors être précisées. Pour cela, l'expert est interrogé et fournit un nouveau trait descriptif générique. Un joint est alors opéré entre les graphes de description des taxons frères et père et le graphe représentant le nouveau trait descriptif. Par exemple, lors de l'acquisition de la tribu des *Vandeae*, son graphe de description est identique à celui de la tribu soeur des *Neottieae* ; la phase d'extraction de traits discriminants échoue donc. C'est alors que l'expert est interrogé pour raffiner la description de ces taxons ; un joint est opéré entre les graphes de description de ces deux taxons et de leur père et le graphe [POLLINIE:]→(DE_TEXTURE)→[TEXTURE:] représentant le trait descriptif générique indiqué par l'expert. En revenant à la phase de complétion, ce trait descriptif est spécialisé dans les descriptions des deux taxons frères par restriction du type TEXTURE de sorte que la phase d'extraction de traits pertinents réussit, comme nous l'avons décrit ci-dessus.

Soulignons que le joint avec un nouveau trait descriptif est opéré non seulement avec les descriptions des taxons frères mais aussi avec la description du père de ces derniers. Cela permet, au moment de l'acquisition ultérieure d'un nouveau taxon frère, que ce dernier hérite du trait descriptif générique qui a servi à discriminer ses frères, sans devoir interroger inutilement l'expert. Par exemple, considérons l'acquisition de la tribu des *Epidendreae* dans un état de la base où les tribus soeurs des *Neottieae* et des *Vandeae* sont acquises, discriminées par la texture des pollinies qui apparaît comme trait générique dans la description de leur père, la sous-famille des *Orchidioideae*. En héritant de la description des *Orchidioideae*, la tribu des *Epidendreae* hérite en particulier du trait générique de la texture des pollinies ; ce trait est spécialisé dans la phase de complétion. La phase d'extraction de traits pertinents commence alors par un test de discrimination automatique entre ce dernier taxon acquis et les tribus des *Neottieae* et des *Vandeae*. Comme nous l'avons expliqué précédemment, il s'agit de déterminer si les traits pertinents retenus pour discriminer les frères du taxon nouvellement acquis permettent également de discriminer ce dernier de ses frères ou s'il faut rechercher un nouveau trait (cf. § 3.1.2.3).

3.1.2.5. Schéma général du processus d'acquisition

La figure 3.1. ci-dessous résume les différentes phases du processus d'acquisition que nous venons de décrire. L'ensemble de ces phases est itéré pour chaque taxon.

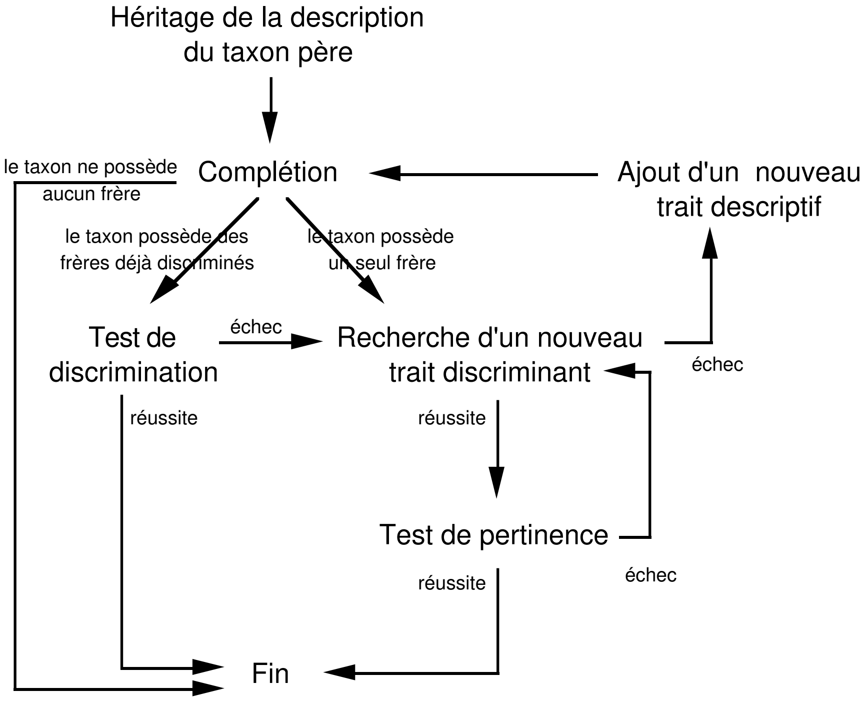


figure 3.1. Les différentes phases du processus d'acquisition d'un taxon

Les principales étapes de ce processus d'acquisition de connaissances consistent chacune en une opération sur les graphes conceptuels qui décrivent les taxons. La phase d'héritage d'une description initiale fait appel à l’opération de copie de graphe qui est un cas particulier de restriction de concept ; celle de complétion fait appel à la restriction de concept ; celle d'ajout d'un nouveau trait descriptif au joint. Le test de discrimination et la recherche d'un nouveau trait discriminant consistent en des appariements de graphes.

Cependant, les différentes étapes de la méthode d'acquisition proposée dans METIS sont définies pour des descriptions de taxons qui font appel au formalisme des graphes conceptuels simples. Or nous nous intéressons à des taxinomies susceptibles de comporter des exceptions. Il s'agit donc de modifier cette méthode d'acquisition afin de pouvoir acquérir des descriptions de taxons représentées dans l'extension du formalisme des graphes conceptuels que nous avons proposée dans le chapitre 2 de cette thèse. Dans la partie suivante, nous revenons sur la phase d'acquisition d'une description initiale d'un taxon et nous détaillons le cas de l’acquisition d’une description comportant des exceptions.

3.2. Acquisition de la description initiale d'un taxon

3.2.1. Principe

Le principe de l'acquisition d'une taxinomie dont on connaît la structure est d'élaborer des descriptions de taxons qui rendent compte de cette structure taxinomique. En particulier, toute description de taxon doit être plus spécifique que celle du taxon père. C'est pourquoi le processus d'acquisition est descendant, la description d'un taxon étant construite à partir de celle de son père (cf. § 1.4.2.). Formellement, dans le modèle des graphes conceptuels, l’acquisition d’une description initiale d’un taxon consiste en la spécialisation du graphe de description du taxon père. Cette étape de construction d'une description initiale d'un taxon correspond dans METIS à l'héritage de la description du taxon père par copie de son graphe de description puis à la complétion de ce dernier par restriction de concept (cf. figure 3.1.).

Cependant, dès lors que le taxon dont on cherche une description comporte des traits descriptifs exceptionnels par rapport à ceux de son père, la description initiale du taxon ne peut être une simple copie de celle de son père. Rappelons en effet que, selon la démarche définitionnelle que nous adoptons pour représenter l'occurrence d'exceptions sur une hiérarchie de concepts, défauts et exceptions sont représentés dans la définition même des taxons (cf. § 2.3.). Aussi, en amont de la phase de complétion, il est nécessaire de définir une phase de description initiale d'un taxon, basée sur un processus d'exception des traits descriptifs du père du taxon dont celui-ci hériterait normalement. Plus précisément, il s'agit de faire apparaître à la fois dans la description du taxon père le caractère par défaut des traits descriptifs dont n'hérite pas son fils et dans la description de ce dernier l'exception des propriétés par défaut de son père. Nous détaillons ce processus d'exception dans la partie suivante.

3.2.2. Acquisition d'une description comportant des défauts et exceptions

L'exception de traits descriptifs entre deux taxons, l'un n'héritant pas de certains traits descriptifs de l'autre qui le subsume, est représentée dans la description du premier par l'application du connecteur ε au type représentant la propriété exceptée. En même temps, le caractère par défaut de la propriété exceptée doit apparaître dans la description du second taxon, par l'application du connecteur δ au même type (cf. § 2.3.3.2.). Rappelons en effet que dans la négative, le graphe de description de celui-ci ne serait pas une généralisation de celui-là, de sorte que ces deux graphes ne décriraient pas deux taxons dont l'un subsume l'autre. Plus précisément, le passage d'un graphe comportant un noeud concept de type δt au graphe identique sauf en ce noeud dont le type est εt est une opération de spécialisation de graphe par restriction de concept, le type δt étant remplacé par le type εt qui lui est inférieur.

Le caractère strict ou par défaut des traits descriptifs d'un taxon n'est pas demandé à l'expert au moment de l'acquisition du taxon. Au contraire, tous les traits descriptifs du taxon étant initialement considérés comme stricts, le caractère par défaut de certains traits d'un taxon est mis en évidence automatiquement, sans devoir interroger l'expert. Plus précisément, les traits descriptifs d'un taxon sont mis à jour et déclarés par défaut ultérieurement, au moment où des exceptions à l'héritage sont mises en évidence lors de l'acquisition d'un taxon fils ou d'un descendant qui n'hérite pas des traits descriptifs de son père. Cela apparaît dans la phase de complétion où l'expert échoue en voulant remplacer un type par un autre qui ne lui est pas inférieur. C'est par exemple le cas si on cherche à remplacer le GRIS des éléphants par le BLANC des éléphants royaux.

De fait, la procédure d'exception consiste à généraliser la description du taxon père en appliquant le connecteur δ sur le type représentant la propriété qui est exceptée. Si ce taxon a lui-même hérité de cette propriété, la description de son propre père doit être généralisée de la même manière. Plus généralement, les descriptions de tous les ancêtres possédant cette propriété doivent être mises à jour. Il s'agit ensuite de donner comme description initiale au taxon en cours d'acquisition une copie du graphe de description de son père ainsi révisé puis de spécialiser automatiquement ce graphe par application du connecteur ε au type représentant la propriété exceptée. La précision de la propriété exceptionnelle du taxon au regard de celle de son père qui est exceptée consiste alors en un joint avec le graphe représentant ce nouveau trait descriptif.

Soit `T` le taxon dont on souhaite acquérir une description `G` et soit `TRAIT` le sous-graphe de la description de son père représentant le trait descriptif dont `T` n'hérite pas. Soit `L` la liste des taxons qui subsument `T` et qui possèdent le trait descriptif `TRAIT` ; `L` contient au moins le père de `T`. L'algorithme général d'acquisition d'une description initiale d'un taxon comportant une exception est décrit sur la table 3.1. ci-dessous ; il comprend la mise à jour automatique des descriptions des taxons supérieurs à T dont une propriété est exceptée. Dans le cas de l'acquisition de la description d'un taxon `T` comportant plusieurs exceptions, l'algorithme est itéré pour chacune d'elles.

1. Généralisation des graphes de description des taxons de L par application du connecteur d au type t de trait représentant la propriété exceptée.
2. G := copie du graphe de description du père de T.
3. Spécialisation de G : restriction de son type dt en et.
4. Spécialisation de G : joint entre G et un graphe identique à `TRAIT` sauf en t remplacé par le type t' représentant la nouvelle propriété.

table 3.1. Acquisition de la description initiale d'un taxon comportant des exceptions

Une fois acquise la description initiale d'un taxon, celle-ci est ensuite raffinée par complétion (cf. figure 3.1.). Les nouveaux types de concepts utilisés dans cette phase de complétion sont introduits dans le support `S` de sorte que celui-ci s'enrichit au fur et à mesure de l'acquisition de nouveaux taxons.

Il s'agit ensuite de vérifier que la description du taxon ainsi élaborée est discriminée de celles de ses frères déjà présents dans la taxinomie acquise et dans la négative de la raffiner encore, ainsi que celle de ses frères, afin de les discriminer. Nous détaillons cette étape de discrimination des taxons dans la partie suivante.

3.3. Acquisition de descriptions de taxons discriminantes

3.3.1. Principe

La méthode d'acquisition de connaissances que nous proposons est basée, comme dans METIS, sur la discrimination de chaque taxon nouvellement acquis. En effet, notre objectif est d'expliciter un langage de description destiné à accéder ultérieurement aux connaissances acquises, notamment à l'aide de procédures d'identification, c'est-à-dire de reconnaissance du taxon auquel appartient un objet particulier (cf. § 3.5.). Une des difficultés pour construire un langage de description est le choix des concepts du langage. Notre ambition n'est pas d'expliciter un langage universel mais au contraire de le définir en fonction de notre but, l'identification d'objets. Le langage de description explicité au fur et à mesure de l'acquisition des taxons est, de fait, un langage discriminant, c'est-à-dire qui permet la discrimination des taxons acquis. L'explicitation de ce langage ne consiste pas simplement en l'acquisition *de* descriptions de taxons pour en déduire ensuite des traits discriminants, encore faut-il déterminer *les* descriptions qui permettent effectivement de discriminer les taxons. A l'issue de la phase de complétion, tout taxon nouvellement acquis est donc discriminé de ses frères déjà présents dans la base de connaissances. Chaque fois que la discrimination de taxons frères échoue, le langage de description est enrichi afin de raffiner les descriptions de ces taxons pour les discriminer.

Les taxons étant décrits à l'aide de graphes conceptuels, la discrimination de taxons se ramène à l'appariement de ces graphes. Nous exposons dans la partie suivante le problème de l'appariement de descriptions structurées et la représentation des graphes de description des taxons que nous avons adoptée pour pallier la complexité de leur appariement. Nous détaillons ensuite la phase de discrimination des taxons. Nous précisons enfin les particularités de la discrimination de graphes comportant des types défauts et exceptions.

3.3.2. Appariement de descriptions structurées

La discrimination de taxons décrits par des graphes conceptuels se ramène à la recherche d'un appariement de descriptions structurées : deux graphes sont discriminés s'il ne peuvent pas être appariés. Or le problème de l'appariement de descriptions structurées s'apparente à l'isomorphisme de sous-graphes qui est un problème NP-complet [Haussler, 89]. La complexité de l'appariement de descriptions structurées croît de manière exponentielle avec le degré de structuration des descriptions [Zucker, 96]. Aussi, le contrôle de la complexité des algorithmes d'appariement impose d'introduire des restrictions dans les descriptions considérées, pour limiter le nombre d'appariements.

Dans le cadre de la généralisation de graphes conceptuels, une solution pour réduire la complexité de l'appariement de graphes a été proposée par Mineau dans la MSG [Mineau, 90]. C'est aussi celle adoptée par Bournaud dans COING [Bournaud, 96]. Elle consiste à représenter un graphe conceptuel par un ensemble $\{[C_i] \rightarrow (R_i) \rightarrow [C'_i],\ i \in I\}$ d'arcs indépendants et à restreindre alors l'appariement de deux graphes à des appariements partiels entre les arcs conceptuels des deux graphes. Rappelons que dans le cas qui est le notre où les relations qui interviennent dans les graphes sont toutes binaires et où les concepts C_i dont le type est inférieur à COMPOSANT sont tous distincts (incomparables), la décomposition des graphes est directe et sans perte d'information. Un graphe conceptuel décomposé de la sorte peut en effet être reconstruit en effectuant des joints entre les arcs qui le composent sur les concepts de types inférieurs à COMPOSANT (cf. § 1.4.1.).

Le gain en complexité de la décomposition des graphes en ensembles d'arcs conceptuels est évident. Une seule mise en correspondance est possible entre deux arcs $[C_i] \rightarrow (R_i) \rightarrow [C'_i]$ et $[C_j] \rightarrow (R_j) \rightarrow [C'_j]$: celle qui fait correspondre C_i avec C_j, R_i avec R_j et C'_i avec C'_j. Par conséquent, le nombre de mises en correspondances possibles entre deux graphes décomposés respectivement en n_i et n_j arcs ($n_i \geq n_j$) n'est plus que :

$$n_i * n_j \text{ au lieu de } n_i^{n_j}$$

pour la recherche d'appariements "plusieurs à un",

$$\sum_{n_2-n_1+1}^{n_2} i = \frac{n_2(n_2+1)-(n_2-n_1)(n_2-n_1+1)}{2} \text{ au lieu de } \prod_{n_2-n_1+1}^{n_2} i = \frac{n_1!}{(n_2-n_1)!}$$

pour la recherche d'appariements "un à un".

Nous définissons l'appariement de graphes conceptuels décomposés en ensembles d'arcs conceptuels en définissant d'abord l'appariement d'arcs conceptuels.

Définition 3.1. [Appariement d'arcs] : *Soit A un ensemble d'arcs conceptuels et soit* $f : A \rightarrow A\ ;\ a_i \mapsto f(a_i) = a_j$, *avec :*

$$a_i = [t_i : r_i] \rightarrow (R_i) \rightarrow [t'_i : r'_i] \text{ et } a_j = [t_j : r_j] \rightarrow (R_j) \rightarrow [t'_j : r'_j].$$

f est un appariement d'arcs conceptuels si :

$$t_i \leq t_j,\ r_i \leq r_j,\ R_i \leq R_j,\ t'_i \leq t'_j \text{ et } r'_i \leq r'_j$$

ou bien si :

$$t_i \geq t_j,\ r_i \geq r_j,\ R_i \geq R_j,\ t'_i \geq t'_j \text{ et } r'_i \geq r'_j.$$

Apparier deux arcs conceptuels a_i *et* a_j*, c'est trouver un appariement f tel que* $f(a_i) = a_j$.

Définition 3.2. [Appariement de graphes conceptuels] : *Soient* A^n *et* A^m *deux ensembles de graphes conceptuels représentés par des ensembles d'arcs de cardinalités respectives n et m,* $n \leq m$. *Soit* $F_{n,m} : A^n \rightarrow A^m$; G_n a $F_{n,m}(G_n) = G_m$, *avec* $G_n = \{a_{n1}, a_{n2}, ... a_{nn}\}$ *et* $G_m = \{a_{m1}, a_{m2}, ... a_{mm}\}$.

$F_{n,m}$ *est un appariement de graphes s'il existe un appariement d'arcs f tel que:*

$$\forall a_{ni} \in G_n, f(a_{ni}) \in G_m.$$

$F_{n,m}$ *est un appariement "un à un" si :*

$$\forall a_{ni}, a_{nj} \in G_n, a_{ni} \neq a_{nj}, f(a_{ni}) \neq f(a_{nj})$$

Apparier deux graphes conceptuels G_n *et* G_m, *c'est trouver un appariement* $F_{n,m}$ *tel que:*

$$F_{n,m}(G_n) = G_m.$$

Soulignons, qu'eu égard aux restrictions que nous imposons aux graphes de description des taxons, il existe au plus un appariement entre de tels graphes et qu'il s'agit d'un appariement un à un. En effet, les types des concepts C_i des arcs $[C_i] \rightarrow (R_i) \rightarrow [C'_i]$ d'un graphe G décrivant un taxon sont tous incomparables. De ce fait, étant donnés deux graphes G_1 et G_2, il existe au plus un concept C_{2j} de G_2 de type comparable à celui d'un concept C_{1i} de G_1 et donc au plus un arc $[C_{1i}] \rightarrow (R_{1i}) \rightarrow [C'_{1i}]$ de G_1 susceptible de s'apparier avec un arc $[C_{2j}] \rightarrow (R_{2j}) \rightarrow [C'_{2j}]$ de G_2.

3.3.3. Discrimination des descriptions de taxons

3.3.3.1. Définition

A l'aide de la notion d'appariement de graphes que nous venons d'introduire, nous définissons la discrimination de graphes conceptuels comme suit :

Définition 3.3. [Discrimination de graphes conceptuels] : *Soient deux graphes conceptuels représentés par deux ensembles d'arcs de cardinalités respectives n et m,* $n \leq m$: $G_n = \{a_{n1}, a_{n2}, ... a_{nn}\}$ *et* $G_m = \{a_{m1}, a_{m2}, ... a_{mm}\}$.

G_n *et* G_m *sont discriminés s'il ne peuvent pas être appariés :*

$$\neg\exists F_{n,m} / F_{n,m}(G_n) = G_m,$$

c'est-à-dire s'il existe un arc de G_n *qui ne peut être apparié avec aucun arc de* G_m :

$$\exists a_{ni} \in G_n, \neg\exists f / f(a_{ni}) \in G_m.$$

Sur la base de cette définition, nous pouvons maintenant donner l'algorithme de la phase de discrimination des descriptions des taxons : il est fondé sur la recherche d'appariements d'arcs conceptuels.

3.3.3.2. Algorithme

Dans la phase de discrimination de taxons, on distingue trois étapes. La première étape consiste à tester si le dernier taxon acquis est discriminé de ses frères déjà présents dans la base de connaissances à l'aide du ou des traits descriptifs qui discriminent ces derniers entre eux. Dans la négative, la seconde étape est la recherche dans la description du dernier taxon acquis d'un trait descriptif supplémentaire qui le discrimine du frère dont il n'a pas été discriminé lors de la première étape. Enfin, en cas d'échec de cette recherche, la troisième étape est un joint entre chacun des graphes décrivant les taxons frères et un nouveau trait descriptif ; elle est suivie, après un retour à la phase de complétion des descriptions des taxons, d'un test qui établit si le dernier taxon acquis est maintenant discriminé de ses frères.

Considérons l'état de la base au moment de l'acquisition d'un taxon T_n. Soit T son père dans la taxinomie et soit LT = $\{T_1, \ldots T_{n-1}\}$ la liste de ses n-1 taxons frères déjà acquis et discriminés ; soient A, $A_1, \ldots A_{n-1}$ et C, $C_1, \ldots C_{n-1}$ les représentations atomiques et celles contractées de ces taxons et soient $D_1, \ldots D_{n-1}$ les parties discriminantes de $C_1, \ldots C_{n-1}$. On a donc $\forall i=1..n-1,\ D_i \subset C_i \subset A_i$. Soit LD la liste $\{d_1, \ldots d_p\}_{p \leq n-1}$ des traits descriptifs génériques qui discriminent $T_1, \ldots T_{n-1}$ (ils sont instanciés dans chaque D_i, i=1..n-1). Soient alors A_n et C_n les descriptions atomique et contractée de T_n obtenues à l'issue de la phase d'acquisition de description initiale (cf. § 3.2.). Nous décrivons sur la table 3.2. ci-après l'algorithme mis en oeuvre dans la phase de discrimination de T_n de ses n-1 frères.

```
1. Test de discrimination
    $R:=\emptyset$
    $D_n:=\{f(d_1), \dots f(d_p)\}$ avec $f: LD \rightarrow C_n$ un appariement d'arcs
    Pour $i:=1..n-1$
        S'il existe un appariement de graphes F tel que $F(D_i)=D_n$
            alors $R:=C_i-D_i$
    Si $R=\emptyset$ alors FIN sinon ($R=\{a_{ik}\}_k$) :
2. Recherche d'un nouveau trait discriminant
    S'il existe un arc $a_{nk} \in C_n-D_n$ tel que. il n'existe pas d'arc $a_{ik}$ tel que. il
    existe un appariement f tel que. $f(a_{nk})=a_{ik}$
            alors la généralisation de $a_{nk}$ est un trait descriptif
            discriminant ; il est retenu s'il est déclaré pertinent par l'expert.
    Sinon :
3. Ajout d'un nouveau trait descriptif
    Soit d un nouveau trait descriptif demandé à l'expert
    Ajout des nouveaux types de d dans $T_c$ et $T_r$
    $A:=Joint(A,d)$
    Pour $j:=1..n$, $A_j:=Complétion(Joint(A_j,d))$
    Soient $d_i$ et $d_n$ les spécialisations de d dans $A_i$ et $A_n$
    S'il n'existe pas d'appariement d'arcs f tel que $f(d_i)=d_n$ alors FIN
    Sinon répéter 3.
FIN
```

table 3.2. Algorithme de discrimination d'un taxon

3.3.3.3. Complexité

Considérons dans un premier temps la complexité du test d'appariement de deux arcs conceptuels qui intervient tout au long de cet algorithme. La prise en compte de connaissances du domaine représentées par des hiérarchies de types telles que celles du support des graphes conceptuels implique un coût pour l'opération d'appariement de deux arcs : celui des comparaisons des types des concepts et des relations qui composent les deux arcs (cf. définition 3.1.). Plus précisément, la comparaison de deux types dans une même hiérarchie se ramène à la recherche d'un des types dans le sous-arbre issu de l'autre type et dans les ancêtres de ce type. Le coût de la comparaison de deux types dépend donc de la taille de la hiérarchie dans laquelle ils sont ordonnés : si la hiérarchie comporte n types, la comparaison de deux types requiert au plus n tests d'égalité; elle est donc en

$O(n)$. Le test d'appariement de deux arcs requiert donc au plus $2n_c + n_r$ tests d'égalité, où n_c et n_r sont les cardinalités respectives de la hiérarchie des types de concepts T_C et de la hiérarchie des types de relations T_r. Il est donc en $O(n_c + n_r)$.

Intéressons-nous maintenant à la complexité de l'algorithme général de la phase de discrimination décrit ci-dessus. Nous allons montrer dans quelle mesure cette complexité est contrôlée grâce à la décomposition des graphes en arcs conceptuels. Dans les étapes 2 et 3 de cet algorithme, seuls deux graphes sont appariés. En effet, étant donnés $n-1$ graphes discriminés deux à deux, si un $n^{ième}$ graphe n'est pas discriminé de tous ces $n-1$ graphes, c'est qu'il s'apparie avec exactement un de ces graphes. C'est l'étape 1, le test de discrimination, qui est coûteuse. Cependant, elle ne nécessite pas une recherche d'appariements sur les descriptions entières des taxons. Remarquons en effet que la place dans la hiérarchie aristotélicienne des types dont il s'agit d'expliciter des descriptions discriminantes est connue : ces types représentent des concepts frères et leurs définitions ont donc le même genre. En d'autre termes, sachant que les graphes différences des formes atomiques des définitions de ces types contiennent les mêmes arcs conceptuels représentant les propriétés héritées des ancêtres communs des concepts, il est suffisant de ne chercher à apparier que les graphes différences des formes contractées de leurs définitions. En cela, la complexité de la recherche d'appariements nécessaire à l'explicitation de descriptions de concepts discriminantes est bien moindre que celle des algorithmes de regroupement conceptuel ou encore de classification de concepts. La formation de concepts par généralisation de graphes [Mineau, 90][Bournaud, 96] ou la classification de concepts par insertion de types définis dans une hiérarchie aristotélicienne [Leclère, 95] nécessitent en effet l'appariement des graphes différences des formes atomiques (cf. Définition 1.12) des définitions des types.

Pour calculer la complexité du test de discrimination, considérons le nombre d'arcs conceptuels nécessaires pour discriminer $n-1$ taxons : il est au plus égal à $n-2$. Le test de discrimination du $n^{ième}$ taxon d'un de ses $n-1$ frères requiert donc au plus :

$$\frac{(n-2)(n-1)}{2} \text{ tests d'appariement d'arcs.}$$

Cela correspond au nombre de mises en correspondances possibles entre deux graphes décomposés en n_i et n_j arcs quand $n_i = n_j = n-2$ (cf. § 3.3.2.).
Le test de discrimination du $n^{ième}$ taxon de ses $n-1$ frères requiert donc au plus :

$$\frac{(n-1)^2(n-2)}{2} \text{ tests d'appariement d'arcs.}$$

Discriminer un taxon de ses $n-1$ frères a donc une complexité au pire en $O(n^3(n_c + n_r))$.
Soulignons bien que cette complexité n'est pas celle de l'algorithme général d'acquisition

d'une taxinomie. Cette complexité dépend du nombre total de taxons et de la profondeur de la taxinomie et est plus faible (cf. § 3.4.2.).

3.3.4. Discrimination de descriptions comportant des défauts et exceptions

Dans cette partie, nous définissons deux principes selon lesquels discriminer des descriptions de taxons comportant des défauts et des exceptions. Pour respecter ces principes, nous seront amenés à réviser la définition de la discrimination de graphes conceptuels. Nous mettrons en évidence que dans ces conditions, l'algorithme de discrimination de taxons décrit dans la partie précédente convient pour toute description de taxon.

3.3.4.1. Discrimination de descriptions comportant des exceptions

Considérons un taxon dont un trait descriptif du graphe différence de la définition du type qui le représente est une exception à l'un des traits descriptifs de son père dans la taxinomie. Nous définissons le principe selon lequel il sera discriminé de ses frères dans la taxinomie comme suit :

Principe 1 : *Le fait qu'un taxon constitue une exception à un trait descriptif de son père dans la taxinomie doit permettre de discriminer ce taxon de tous ses frères - sauf ceux qui constituent une exception à ce même trait descriptif.*

Soit T_i un tel taxon ; examinons les appariements possibles entre sa description et celles de ses frères. T_i possède un arc de la forme [C]→(R)→[εt' : r'] qui vient excepter l'arc [C]→(R)→[δt' : r'] de la description du taxon père et un arc de la forme [C]→(R)→[t'_i : r'] tel que t'_i et t' sont incomparables (cf. § 3.2.2.). Ces deux arcs sont susceptibles de s'apparier avec un même arc [C]→(R)→[C_j] de la description d'un taxon T_j frère de T_i. Cet arc de la description de T_j est soit l'arc [C]→(R)→[t' : r'] hérité de la description du taxon père avant qu'il ne soit généralisé dans la description de ce dernier au moment de l'introduction de T_i (cf. § 3.2.2.) ; soit l'arc [C]→(R)→[εt' : r'] ou l'arc [C]→(R)→[t'_j : r'] (avec t'_j et t' incomparables) dans le cas où T_j constitue également une exception au trait descriptif [C]→(R)→[δt' : r'] du taxon père ; soit enfin l'arc [C]→(R)→[δt' : r'] parce que la description de l'un de ses fils comporte une exception à ce trait descriptif.

<u>Cas 1</u> : [C]→(R)→[t' : r'] ne s'apparie ni avec [C]→(R)→[εt' : r'] ni avec [C]→(R)→[t'_i : r'] car t', εt' et t'_i sont incomparables. Le principe énoncé ci-dessus est

donc bien respecté : le fait de constituer une exception à un trait descriptif du taxon père suffit à discriminer T_i d'un frère pour lequel ce trait descriptif est strict.

Cas 2 : l'arc [C]→(R)→[εt' : r'] est présent dans la description de T_i et celle de T_j ; la discrimination des descriptions de ces deux taxons dépend donc de l'appariement des autres arcs qui les composent, à commencer par [C]→(R)→[t'_i : r'] et [C]→(R)→[t'_j : r']. Le principe énoncé ci-dessus est donc bien respecté : le fait que T_i et T_j fassent exception au même trait descriptif du taxon père ne discrimine pas leurs descriptions. Par contre, selon les nouveaux traits descriptifs avec lesquels ils font exception, ils sont susceptibles d'être discriminés.

Cas 3 : l'arc [C]→(R)→[δt' : r'] ne s'apparie pas avec [C]→(R)→[t'_i : r'] mais il s'apparie avec [C]→(R)→[εt' : r']. T_i et T_j ne sont donc pas discriminés, ce qui va à l'encontre du principe que nous avons énoncé : selon la définition 3.3. de la discrimination de graphes conceptuels, le fait que T_i constitue une exception à un trait descriptif du taxon père ne le discrimine pas d'un taxon frère T_j pour lequel ce trait descriptif est par défaut, un de ses descendants constituant une exception à ce trait.

Pour pallier ce problème, nous redéfinissons la discrimination de graphes en excluant de l'ensemble des arcs à examiner ceux qui comportent des types exceptions. Soulignons que, dans les graphes considérés, un tel arc est toujours accompagné d'un autre arc représentant le trait descriptif qui nécessite l'exception, de sorte que deux arcs peuvent être appariés avec un même arc de la description d'un taxon frère. En excluant les arcs qui comportent un type exception, il existe au plus un appariement entre deux graphes décrivant deux taxons frères. La définition de la discrimination de graphes conceptuels que nous adoptons est la suivante :

Définition 3.4. [Discrimination de graphes conceptuels construits sur $S_{\delta\varepsilon}$]:

Soient deux graphes conceptuels représentés par deux ensembles d'arcs de cardinalités respectives n et m, n ≤m :

$$G_n = \{a_{n1}, a_{n2}, \dots a_{nn}\} \text{ et } G_m = \{a_{m1}, a_{m2}, \dots a_{mm}\}.$$

Soient G_{np} et G_{mq} les sous-ensembles de G_n et G_m de cardinalités respectives p et q qui ne comportent aucun arc dont un noeud concept est d'un type exception.

G_n et G_m sont discriminés si G_{np} et G_{mq} ne peuvent pas être appariés :

$$\neg\exists\, F_{p,q} \,/\, F_{p,q}(G_p) = G_q,$$

c'est-à-dire s'il existe un arc de G_{np} qui ne peut être apparié avec aucun arc de G_{mq} :

$$\exists\, a_{ni} \in G_{np},\ \neg\exists f \,/\, f(a_{ni}) \in G_{mq}.$$

Remarquons que dans le cas particulier où ni G_n ni G_m ne comportent d'arc dont un noeud concept est de type exception, la définition 3.4. est équivalente à la définition 3.3. établie précédemment pour des graphes conceptuels simples ($G_{np} = G_n$ et $G_{mq} = G_m$).

Considérons à nouveau la discrimination de T_i de ses frères. Dans le cas 3 qui posait problème, le seul arc de la description de T_i susceptible de s'apparier avec l'arc [C]→(R)→[δt' : r'] de T_j est l'arc [C]→(R)→[t'_i : r']. Or t'_i et δt' sont incomparables donc ces arcs ne peuvent être appariés. Conformément au principe de discrimination énoncé, les descriptions de T_i et T_j sont maintenant bien discriminées dans ce cas. Vérifions que les cas 1 et 2 exposés ci-dessus sont encore conformes au principe de discrimination. Dans le cas 1, le seul arc de la description de T_i susceptible maintenant de s'apparier avec l'arc [C]→(R)→[t' : r'] de T_j est l'arc [C]→(R)→[t'_i : r'] ; or t'_i et t' sont incomparables donc ces arcs ne peuvent être appariés. Les descriptions de T_i et T_j sont donc bien discriminées. Dans le cas 2, la discrimination des descriptions de T_i et T_j qui constituent une exception à un même trait descriptif du taxon père dépend bien de l'appariement entre leurs arcs autres que ceux comportant des types exceptions.

3.3.4.2. Discrimination de descriptions comportant des défauts

Considérons un taxon dont un trait descriptif du graphe différence de la définition du type qui le représente comporte un type défaut. Rappelons qu'un tel trait n'est pas explicité dans la description initiale du taxon ; il remplace un trait descriptif strict par généralisation de celui-ci au moment de l'acquisition de la description d'un descendant du taxon (cf. § 3.2.). Nous définissons le principe selon lequel la description d'un tel taxon sera discriminé de celles de ses frères dans la taxinomie comme suit :

Principe 2 : *La description d'un taxon doit pouvoir être discriminée aussi bien à l'aide de traits descriptifs comportant des défauts que de traits n'en comportant pas.*

En d'autres termes, la description d'un taxon T_i initialement discriminée de celles de certains des taxons frères par un trait descriptif [C_i]→(R_i)→[t'_i : r'_i] doit encore être discriminée de ces dernières dans le cas où l'acquisition d'un taxon descendant de T_i requiert de généraliser ce trait descriptif discriminant en [C_i]→(R_i)→[δt'_i : r'_i]. Il s'agit donc de vérifier, dans le cas où un arc [C_i]→(R_i)→[t'_i : r'_i] permet de discriminer la description d'un taxon T_i de celle de l'un de ses frères T_j parce que t'_i et le type t'_j dans l'arc correspondant de la description de T_j sont incomparables, si l'arc [C_i]→(R_i)→[δt'_i : r'_i] permet également de discriminer la description de T_i de celle de T_j, c'est-à-dire si δt'_i et t'_j sont également incomparables.

Déterminons d'abord quelle peut être la description de T_j. Avant l'acquisition d'une description initiale d'un taxon descendant de T_i comportant une exception, la partie discriminante de la description de T_i contenait l'arc $[C_i]\rightarrow(R_i)\rightarrow[t'_i : r'_i]$. Cela signifie que la description du taxon père de T_i contient soit l'arc $[C_i]\rightarrow(R_i)\rightarrow[\delta t'_i : r'_i]$ soit l'arc $[C_i]\rightarrow(R_i)\rightarrow[t' : r'_i]$ qui représente un trait générique ; elle ne peut pas contenir l'arc $[C_i]\rightarrow(R_i)\rightarrow[t'_i : r'_i]$ car cela signifierait que tous ses descendants donc tous les frères de T_i auraient hérité de cet arc qui n'aurait donc pas pu être utilisé pour discriminer la description de T_i de celle d'un de ses frères T_j. Examinons donc les deux autres cas.

<u>Cas 1</u> : la description du père de T_i contient l'arc $[C_i]\rightarrow(R_i)\rightarrow[\delta t'_i : r'_i]$ et la description de T_i est discriminée de celle de son frère T_j à l'aide de son arc $[C_i]\rightarrow(R_i)\rightarrow[t'_i : r'_i]$. Cela signifie que la description de T_j constitue une exception au trait descriptif que représente l'arc $[C_i]\rightarrow(R_i)\rightarrow[\delta t'_i : r'_i]$: elle contient l'arc $[C_i]\rightarrow(R_i)\rightarrow[\varepsilon t'_i : r'_i]$ et un autre arc $[C_i]\rightarrow(R_i)\rightarrow[t'_j : r'_i]$ avec t'_j et t'_i incomparables. Dans ce cas, quand l'arc $[C_i]\rightarrow(R_i)\rightarrow[t'_i : r'_i]$ de la description de T_i est généralisé en $[C_i]\rightarrow(R_i)\rightarrow[\delta t'_i : r'_i]$, on est ramené au cas 3 décrit dans la partie précédente ; la description de T_i est donc encore discriminée de celle de T_j.

<u>Cas 2</u> : la description du père de T_i contient l'arc $[C_i]\rightarrow(R_i)\rightarrow[t' : r'_i]$ et la description de T_i est discriminée de celle de son frère T_j à l'aide de son arc $[C_i]\rightarrow(R_i)\rightarrow[t'_i : r'_i]$. Cela signifie que la description de T_j contient un arc $[C_i]\rightarrow(R_i)\rightarrow[t'_j : r'_i]$, avec t'_j et t'_i incomparables. Il s'agit de la discrimination de graphes ne comportant ni défaut ni exception. Dans ce cas, quand l'arc $[C_i]\rightarrow(R_i)\rightarrow[t'_i : r'_i]$ de la description de T_i est généralisé en l'arc $[C_i]\rightarrow(R_i)\rightarrow[\delta t'_i : r'_i]$, il est encore impossible de l'apparier avec l'arc $[C_i]\rightarrow(R_i)\rightarrow[t'_j : r'_i]$ de T_j . En effet, t'_j et t'_i étant incomparables, t'_j et $\delta t'_i$ le sont également ; la description de T_i est donc encore discriminée de celle de T_j.

3.3.4.3. Avantage d'une approche définitionnelle des défauts et exceptions

A l'issue de cette partie, eu égard à la définition 3.4. de la discrimination de graphes conceptuels, il apparaît que la discrimination de descriptions de taxons comportant des défauts ou des exceptions n'est pas différente de la discrimination de descriptions de taxons qui n'en comportent pas. L'algorithme décrit dans la partie précédente convient donc pour discriminer n'importe quelles descriptions de taxons.

L'avantage d'une approche définitionnelle des défauts et exceptions apparaît ici clairement. Leur représentation dans les descriptions mêmes des taxons permet qu'ils soient gérés en même temps que sont manipulées les descriptions de taxons et ce de la même manière que celles qui n'en comportent pas. En d'autres termes, l'avantage d'une telle approche est que la gestion d'une hiérarchie d'héritage par défaut n'est pas différente de celle d'une hiérarchie d'héritage stricte [Faron et Ganascia, 97b].

3.4. Construction d'une hiérarchie de types définis

3.4.1. Algorithme

Le principe de l'algorithme général de la construction d'une hiérarchie aristotélicienne de types représentant une taxinomie est celui décrit dans le § 3.1.3. Il consiste en l'acquisition incrémentale de définitions de types de concepts représentant chacun un taxon en itérant la succession des trois phases présentées sur la table 3.3. ci-dessous.

Pour tout taxon,

1. Explicitation d'une définition de type (table 3.1.) :
   le graphe différence de cette définition représente la description initiale d'un taxon;
2. Complétion de la description du taxon représenté :
   le graphe différence de cette définition de type est spécialisé par restriction de concepts;
3. Discrimination de la description du taxon représenté de celles des taxons frères (table 4.2.) :
   les graphes différences des définitions des types représentant le taxon considéré et ses frères sont appariés;
   Si nécessaire, spécialisation par joint puis restriction de concept.

table 3.3. Acquisition incrémentale de descriptions de taxons

3.4.2. Complexité

La complexité de l'algorithme général d'acquisition de descriptions de taxons est liée au coût de la phase de discrimination. Rappelons le coût du test de discrimination d'un $n^{ième}$ taxon de ses $n-1$ frères (cf. § 4.1.3.3.) ; il requiert au plus :

$$\frac{(n-1)^2(n-2)}{2}$$ tests d'appariement d'arcs deux à deux.

Le nombre de tests d'appariement d'arcs nécessaires à l'acquisition incrémentale de n taxons frères est donc au plus égal à :

$$\sum_{i=1}^{n}\frac{(i-1)^2(i-2)}{2}=\sum_{i=1}^{n-1}\frac{i^2(i-1)}{2}=\frac{1}{2}\left(\sum_{i=1}^{n-1}i^3-\sum_{i=1}^{n-1}i^2\right)\ (1).$$

Ce nombre est borné supérieurement par son approximation asymptotique :

$$\int_0^{n-1}x^3-\int_0^{n-1}x^2=\frac{(n-1)^4}{4}-\frac{(n-1)^3}{3}$$

Donc :

$$\sum_{i=1}^{n}\frac{(i-1)^2(i-2)}{2}<\frac{1}{2}\left(\frac{(n-1)^4}{4}-\frac{(n-1)^3}{3}\right)\ (2).$$

L'acquisition de n descriptions de taxons frères discriminantes a donc une complexité en :

$$O(n^4(n_c+n_r)).$$

Considérons maintenant la complexité globale de l'acquisition d'une taxinomie comportant N taxons et de profondeur P. Pour cela, nous approximons le nombre de taxons frères à discriminer sous chaque noeud interne de la taxinomie par le nombre moyen de taxons frères dans l'ensemble de la taxinomie. Soit n ce nombre ; on a la relation suivante entre N, P et n :

$$N=\sum_{i=1}^{P}n^i=\frac{1-n^{P+1}}{1-n}-1=\frac{n-n^{P+1}}{1-n}\ (3).$$

Des tests de discrimination de descriptions de taxons sont effectués sous chaque taxon qui n'est pas une feuille de la taxinomie, soit :

$$\sum_{i=1}^{P-1}n^i=\frac{n-n^P}{1-n}\ \text{fois.}$$

Le nombre total de tests de discrimination de descriptions de taxons est donc égal à :

$$\frac{n-n^P}{1-n}\sum_{i=1}^{n}\frac{(i-1)^2(i-2)}{2}.$$

Ce nombre est borné supérieurement par une approximation asymptotique : d'après (2),

$$\frac{n-n^P}{1-n}\sum_{i=1}^{n}\frac{(i-1)^2(i-2)}{2}<\frac{n-n^P}{2(1-n)}\left(\frac{(n-1)^4}{4}-\frac{(n-1)^3}{3}\right)\ (4).$$

D'après cette inéquation, l'acquisition d'une taxinomie de profondeur P dont le nombre moyen de taxons frères est n a donc une complexité au pire en :

$$O(n^{P+3}(n_c+n_r)).$$

En considérant l'équation (3) qui relie n, P et N, cette complexité peut encore être exprimée en fonction du nombre total de taxons de la taxinomie :

$$O(N^{1+\frac{3}{P}}(n_c+n_r)).$$

Dans le cas d'une taxinomie comportant un unique niveau taxinomique, c'est-à-dire quand $P=1$ et $N=n$, on retrouve la complexité en $O(N^4(n_c+n_r))$ de l'acquisition de N taxons frères. Pratiquement, cette borne n'est jamais atteinte. Généralement, on compte en effet au moins trois niveaux taxinomiques : famille, genre et espèce. En orchidologie par exemple, 7 niveaux taxinomiques sont définis (cf. § 1.4). L'acquisition de descriptions de taxons discriminantes a donc, en pratique, une complexité plus faible que $O(N^2(n_c+n_r))$, voire proche de $O(N(n_c+n_r))$.

Le support $S_{\delta\varepsilon}$ sur lequel sont définies les descriptions des taxons est enrichi au fur et à mesure de l'acquisition de la taxinomie, chaque fois que de nouveaux types de concepts sont introduits pour discriminer les graphes décrivant les taxons. Nous pouvons donc exprimer les cardinalités respectives de T_c et T_r en fonction de celle de la taxinomie acquise. La discrimination des descriptions de n taxons frères requiert au plus $2(n-1)$ arcs distincts donc au plus $2n(n-1)$ types de relations différents et $4n(n-1)$ types de concepts différents. Sur un niveau taxinomique i, il y a n^i ensembles de taxons frères à discriminer. Pour l'ensemble de la taxinomie acquise, il y a donc au plus :

$n_r = 2n(n-1)\sum_{i=1}^{P} n^i = 2(n^{P+2}+n^2)$ types de relations différents et

$n_c = 4n(n-1)\sum_{i=1}^{P} n^i = 4(n^{P+2}+n^2)$ types de concepts différents.

La complexité en $O(n^{P+3}(n_c+n_r))$ de l'acquisition d'une taxinomie où P est la profondeur de la taxinomie et n le nombre moyen de taxons frères s'exprime donc encore en $O(n^{2P+5})$ ou encore, en considérant l'équation (3) qui relie n, P et N en :

$$O(N^{2+\frac{5}{P}}).$$

Pour trois niveaux taxinomiques (famille, genre, espèce), cette complexité est plus faible que $O(N^4)$; à partir de six niveaux taxinomiques, elle est plus faible que $O(N^3)$ et tend vers $O(N^2)$. Ce résultat est la preuve de la faisabilité algorithmique de notre approche de l'acquisition de connaissances taxinomiques.

3.5. Identification d'objet

Dans une base de connaissances construite selon le processus d'acquisition que nous proposons, des traits descriptifs génériques ont été explicités qui discriminent entre taxons frères. Ils sont représentés par des arcs conceptuels qui font partie de la description du taxon père et qui sont spécialisés dans chacune des descriptions des taxons frères. Ces ensembles de traits descriptifs sont autant de clefs d'identification qui vont permettre d'identifier pas à pas le taxon le plus spécifique auquel appartient une instance décrite par l'utilisateur. Ces clefs forment un arbre dit de détermination dont la structure coïncide avec celle de la hiérarchie des types qui représentent les taxons. Chaque noeud de l'arbre correspond à un taxon et contient les arcs conceptuels représentant les traits descriptifs génériques de ce taxon (cf. § 1.4.1).

Le processus d'identification d'une instance est itératif et consiste, étant donné un arbre de détermination, à spécialiser à chaque pas d'itération les arcs conceptuels contenus sur un noeud de l'arbre en interrogeant l'utilisateur sur la description de l'instance à classer. Le graphe conceptuel ainsi construit permet de déterminer un (unique) fils du noeud courant qui devient le nouveau noeud courant : on recherche le fils du taxon correspondant au noeud courant dont la description s'apparie avec celle qui est construite de l'instance. Lorsqu'un chemin de la profondeur de l'arbre de détermination a été ainsi parcouru à partir de la racine, l'instance est identifiée comme appartenant au taxon correspondant à la feuille de l'arbre sur laquelle on se trouve.

Cependant, dans le cas d'une taxinomie comportant des exceptions, le processus décrit ci-dessus ne permet de reconnaître que les instances *typiques* d'un taxon. La description d'une instance fournie par l'utilisateur ne peut en effet être que stricte et ne s'apparie donc avec aucune description de taxon comportant des exceptions. En d'autres termes, les noeuds de l'arbre de détermination qui correspondent à des taxons dont les descriptions comportent des exceptions ne sont jamais explorés et, de ce fait, certaines instances ne peuvent pas être identifiées. Considérons par exemple l'extrait de la taxinomie des animaux contemporains présentée dans le chapitre 2 (figure 2.12.). Un animal contemporain décrit par les arcs conceptuels [ANIMAL]→(COULEUR)→[BLANC] et [ANIMAL]→(COMPO)→[TROMPE] ne peut être reconnu instance du taxon *Eléphant_Royal* : il doit pour cela d'abord être reconnu instance du taxon *Éléphant*, ce que ne permet pas sa description. Leurs arcs [ANIMAL]→(COULEUR)→[BLANC] et [ANIMAL]→(COULEUR)→[δGRIS] ne peuvent en effet pas être appariés. De même, un animal décrit par les arcs [ANIMAL]→(COULEUR)→[GRIS] et [ANIMAL]→(COMPO)→[TROMPE], qui est donc reconnu instance du taxon *Éléphant*, ne peut être reconnu instance du taxon *Eléphant_Royal_Sale* car sa description ne permet pas de le reconnaître instance du taxon *Éléphant_Royal*. Leurs arcs [ANIMAL]→(COULEUR)→[GRIS] et [ANIMAL]→(COULEUR)→[δBLANC] ne peuvent pas être appariés. Dès lors que les descriptions de taxons sont construites sur un support $S_{\delta\varepsilon}$, le principe d'identification décrit ci-dessus doit donc être modifié. Pour cela, nous adoptons la démarche proposée dans [Coupey et Fouqueré, 94b] qui consiste à distinguer des instances typiques et des instances exceptionnelles et à raffiner les descriptions des instances exceptionnelles pour pouvoir les identifier.

Étant donné un spécimen O, le processus d'identification décrit ci-dessus permet de déterminer les taxons dont cet objet est une instance typique. Pour reconnaître ceux dont il est une instance exceptionnelle, il faut raffiner sa description à chaque pas d'identification, c'est-à-dire à chaque niveau taxinomique, où la description d'un taxon candidat dont O n'est pas une instance typique comporte un défaut sur un des arcs conceptuels qui sert à le discriminer de ses frères. En introduisant la mention d'une exception dans la description de l'objet, celle-ci peut s'apparier avec celle du taxon ; l'objet à identifier est une instance exceptionnelle de ce taxon. Un fils de ce taxon qui fait exception à celui-ci sur certains traits descriptifs est susceptible d'avoir une description qui s'apparie avec celle de l'objet. Considérons à nouveau les deux animaux contemporains décrits ci-dessus. Le premier n'étant pas reconnu instance typique du taxon *Éléphant*, sa description est spécialisée en y introduisant (par joint) l'arc [ANIMAL]→(COULEUR)→[εGRIS] qui s'apparie avec l'arc [ANIMAL]→(COULEUR)→[δGRIS] de la description du taxon *Éléphant*. L'animal est alors reconnu instance exceptionnelle de ce taxon. Au pas d'itération suivant, il sera reconnu instance typique du taxon *Eléphant_Royal*. De même, le second animal décrit, qui lui est reconnu instance typique du taxon *Éléphant,* n'est pas ensuite reconnu instance typique du taxon *Eléphant_Royal*. Sa description est alors raffinée en y introduisant l'arc [ANIMAL]→(COULEUR)→[εBLANC] qui s'apparie avec l'arc [ANIMAL]→(COULEUR)→[δGRIS] de la description du taxon *Eléphant_Royal*. L'animal est alors reconnu instance exceptionnelle de ce taxon. Au pas d'itération suivant, il sera reconnu instance typique du taxon *Eléphant_Royal_Sale*.

L'algorithme d'exploration du graphe de détermination est décrit sur la table 3.4. ci-dessous. La taxinomie et l'arbre de détermination sont confondus ; soit P leur profondeur. Pour chaque noeud, c'est-à-dire chaque taxon, on distingue la partie D de sa description qui contient les arcs conceptuels qui permettent de le discriminer de ses frères et la partie G qui contient les arcs conceptuels qui représentent des traits descriptifs génériques et qui, instanciés, permettent de reconnaître un taxon fils en le discriminant de ses frères. On appelle D^* le sous-ensemble des arcs conceptuels de D qui ne contiennent aucun type défaut.

```
0. Initialisation:
   T:=racine de la taxinomie; Ex:={}; n:=1.
   Description initiale de O par spécialisation des traits descriptifs de G:
      $D_O:=\{a_k\}_k$.
1. Détermination du taxon dont O est instance typique:
   Tant que n<P faire
      Pour tout fils $T_i$ de T faire
         Si $D_O$ et $D^*_i$ s'apparient alors $Ex:=Ex \cup \{(T_i,n+1)\}$
      S'il existe un $T_i$ tel que $D_O$ et $D_i$ s'apparient alors
         $T:=T_i$; $D_O:=D_O \cup \{a_{ik}\}_k$; n:=n+1
      Sinon n:=P.
   FinTantQue
2. Détermination des taxons dont O est instance exceptionnelle:
   Pour tout $(T,n) \in Ex$ faire
      Soit $\{a, a \in G\}$ tel que, spécialisé en $\{a_k\}_k$, $D_O \cup \{a_k\}_k$ s'apparie avec D;
      $D_O:=D_O \cup \{a_k\}_k$
      Tant que n<P faire
         Pour tout fils $T_i$ de T faire
            Si $D_O$ et $D^*_i$ s'apparient et pas $D_O$ et $D_i$ alors
               $Ex:=Ex \cup \{(T_i,n+1)\}$
            S'il existe un $T_i$ tel que $D_O$ et $D_i$ s'apparient alors
               $T:=T_i$; $D_O:=D_O \cup \{a_{ik}\}_k$; n:=n+1
            Sinon n:=P.
         FinPour
      FinTantQue
   FinPour
```

table 3.4. Algorithme d'identification d'un objet

La démarche que nous adoptons revient à considérer non plus un arbre de détermination mais un graphe comportant deux types d'arcs : ceux qui correspondent à la structure de la taxinomie et ceux qui mettent en relation un noeud correspondant à un taxon T avec un noeud qui correspond à un descendant T' de T et qui fait exception à un fils de T. Ces arcs qui permettent un test d'instance exceptionnelle sont explorés après ceux du type précédent qui permettent un test d'instance typique.

La figure 3.2. ci-dessous présente le graphe de détermination associé à la taxinomie des animaux contemporains.

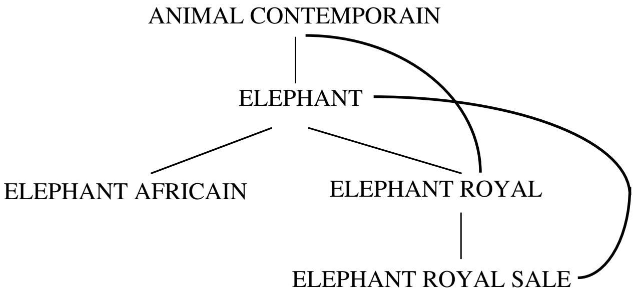


figure 3.2. Graphe de détermination

3.6. Conclusion

Dans ce chapitre, nous avons présenté les raisonnements sur les graphes conceptuels que nous mettons en oeuvre pour acquérir des connaissances taxinomiques et identifier ensuite les objets du domaine à l'aide de la base ainsi construite. Dans le chapitre suivant, nous montrons qu'une telle base de connaissances taxinomiques formalisées dans le modèle des graphes conceptuels peut constituer la structure d'un réseau hypermédia qui offre, de ce fait, des possibilités d'accès à l'information basées sur la manipulation formelle des connaissances sous-jacentes. La méthode d'acquisition des connaissances que nous avons présentée nous permet de définir une méthode de conception d'un tel hypermédia, c'est-à-dire d'organisation du réseau hypermédia.

Chapitre 4

Conception d'hypermédias à base de connaissances taxinomiques

Introduction

Dans ce chapitre, nous nous intéressons à l'intégration de connaissances formelles et informelles au sein d'un même système ou plus précisément à la collaboration entre base de connaissances et hypermédia. Un tel système qui intègre ces deux formes de mémorisation et d'organisation de l'information est de fait appelé hypermédia à base de connaissances. Nous présentons d'abord les apports mutuels d'une base de connaissances et d'un réseau d'information hypermédia (§ 4.1.). Par la suite, nous nous concentrons sur les hypermédias à base de connaissances du domaine. Nous mettons en évidence la problématique qui leur est associé, à savoir la conception du réseau hypermédia et l'accès à l'information sur un ce réseau organisé à l'aide de connaissances formalisées. Nous proposons une revue des différents types d'hypermédia à base de connaissances (§ 4.2.). Nous présentons plus en détail les travaux de conception d’hypermédia à base de connaissances du domaine dans lequel s’inscrit notre travail (§ 4.3.). Nous proposons une approche originale de la conception d’hypermédias à base de connaissances du domaine représentées dans le formalisme des graphes conceptuels et explicitées selon une méthode d’acquisition de connaissances (§ 4.4.).

4.1. Hypermédia à base de connaissances

4.1.1. Un hypermédia pour une base de connaissances

4.1.1.1. Le contexte de conception d'une base de connaissances

Utiliser une base de connaissances requiert, de la part de l'utilisateur, l'exactitude de son interprétation des expressions formelles qu'il manipule. Les sources d'information à partir desquelles sont explicitées des connaissances constituent un moyen d'en assurer l'intelligibilité. C'est ce que Rechenmann appelle le contexte de conception [Rechenmann, 93]. Or une base de connaissances est généralement destinée à des utilisateurs autres que son concepteur, de sorte que le contexte de conception d'une base de connaissances est inconnu des utilisateurs de celle-ci. La taille des grandes bases de connaissances est un autre facteur d'ignorance de leur contexte de conception [Rechenmann, 93][Lemaire et Rechenmann, 96].

Préserver le contexte de conception d'une base de connaissances consiste à mémoriser les sources d'information à partir desquelles ont été explicitées les connaissances et à maintenir des liens entre l'expression formelle des connaissances et les sources d'information. Un hypermédia est particulièrement bien adapté aux spécificités de cette tâche de mémorisation. Plus précisément, le maintien d'une base de connaissances dans son contexte de conception est réalisable en intégrant cette base de connaissances à un hypermédia, c'est-à-dire en indexant des documents multimédias et les relations qu'ils entretiennent à leurs expressions formelles dans la base de connaissances. Les connaissances formalisées sont ainsi accessibles grâce à la structure hypertexte du système [Grivaux et Rechenmann, 92].

4.1.1.2. Une base de connaissances consensuelle

Sans documents textuels et graphiques additionnels, des incompréhensions sont inévitables de la part des utilisateurs d'une base de connaissances non experts du domaine. Elles sont même susceptibles d'apparaître entre experts du domaine, du fait d'ambiguïtés entre leurs langages respectifs. En particulier si ces experts participent ensemble à l'explicitation d'une base de connaissances, il est important qu'ils accordent leurs langages de description du domaine. La base d'un tel accord peut être, de nouveau, un ensemble de documents textuels et graphiques organisé sous une forme hypertextuelle qui rassemble les connaissances du domaine que l'on cherche à formaliser [Rechenmann, 93][Euzenat, 95][Euzenat, 96].

4.1.2. Une base de connaissances pour un hypermédia

4.1.2.1. Intelligibilité du réseau hypermédia

De part sa structure non linéaire, un hypermédia impose à l'utilisateur une charge cognitive importante. Celui-ci doit en effet traiter à la fois le contenu informatif des noeuds du réseau d'information et les relations entre ces noeuds. En d'autres termes, à sa tâche de compréhension des documents hypermédias s'ajoute celle de recherche des documents pertinents dans le réseau : l'utilisateur alterne lecture et navigation et le passage d'une tâche à l'autre est sans transition.

La charge cognitive de l'utilisateur d'un hypermédia est d'autant plus importante que l'efficacité de sa recherche d'une information pertinente dépend de sa faculté d'orientation dans le réseau d'information. Celle-ci est le résultat de la compréhension qu'il a de la structure du réseau d'information. Or plus la taille et la complexité du réseau d'information sont importantes, plus il est difficile d'en appréhender la structure.

La navigation est donc un mode d'accès à l'information qui présente pour l'utilisateur un risque de *désorientation* [Haake, 94]. Dans la conception d'un hypermédia, il est de ce fait nécessaire de rendre le plus *intelligible* possible la structure de l'ensemble des liens le long desquels peut cheminer l'utilisateur avant d'atteindre un noeud d'information qui l'intéresse. Pour ce faire, une solution est d'introduire dans l'hypermédia des connaissances sur le domaine de l'application. Indexer les éléments d'un réseau hypermédia à une base de connaissances du domaine est un moyen d'une part d'organiser le réseau hypermédia pour le rendre plus intelligible à l'utilisateur et d'autre part de le rendre manipulable par le système pour aider l'utilisateur à accéder à l'information.

4.1.2.2. Accessibilité de l'information

Une base de connaissances améliore l'accès à l'information d'un hypermédia non seulement en procurant au réseau d'information une sémantique qui le rende intelligible pour l'utilisateur mais aussi en le dispensant d'une trop grande complexité. En effet, une alternative à la définition d'un vaste et complexe ensemble de liens entre des noeuds d'information hypermédias consiste à ne construire qu'un réseau relativement simple sur l'ensemble des noeuds d'information et de *calculer* des liens entre ces noeuds *dynamiquement*, selon les besoins de l'utilisateur. La navigation n'est plus alors un cheminement le long des liens prédéfinis mais un tracé de route dynamique, les liens nécessaires à la navigation étant créés, effectivement ou virtuellement, au fur et à mesure. Cela est d'autant plus utile que tous les liens ne peuvent pas être prédéfinis sur un hypermédia de grande taille. La navigation sur un réseau hypermédia dont les noeuds d'information sont indexés à une base de connaissances est de ce fait plus facile et les cheminements sur le réseau plus intelligibles que sur un réseau hypermédia totalement prédéfini, par trop complexe.

D'autre part, l'amélioration de l'accès à l'information dans un hypermédia enrichi d'une base de connaissances ne se limite pas à de nouvelles possibilités de navigation. Une base de connaissances offre également des moyens d'accès à l'information complémentaires. Typiquement, l'utilisateur d'une base de connaissances accède à l'information représentée en formulant des *requêtes*. Des procédures de recherche d'information peuvent donc être intégrées à un hypermédia dont les noeuds d'information sont indexés à une base de connaissances. Mais le traitement d'une requête n'est pas différent de celui d'un “pas” de navigation dynamique. On peut considérer que l'expression d'une requête est préalable au calcul d'un lien du réseau hypermédia. En quelque sorte, un lien calculé est étiqueté par la requête dont il est la réponse, ou plutôt le moyen d'accès à la réponse. En d'autres termes, dans les procédures d'accès à l'information sur un hypermédia dont les noeuds d'information sont indexés à leur expression formelle dans une base de connaissances, il n'y a pas de frontière entre navigation et recherche d'information. L'exploitation des capacités de la base de connaissances rend l'accès à l'information, en mode de navigation ou de recherche d'information, moins dépendant d'un cheminement sur un réseau prédéfini de liens entre des noeuds dont l'un d'entre eux est susceptible de contenir l'information pertinente. L'information pertinente est atteinte non pas en cheminant le long des liens qui existent entre le noeud d'information courant et celui qui contient l'information recherchée mais en calculant dynamiquement un nouveau lien entre ces deux noeuds. L'accès à l'information n'est plus envisagé comme la recherche d'un noeud susceptible de contenir une information pertinente en cheminant sur le réseau hypermédia mais comme la recherche d'une information figurant sur un noeud ou pouvant être déduite de celles contenues dans le réseau hypermédia, à l'aide de structures d'inférence sur des connaissances formalisées.

4.2. Hypermédia à base de connaissances du domaine

La problématique générale associée aux hypermédias à base de connaissances se différencie selon la nature des connaissances introduites, c'est-à-dire selon ce que cherche à modéliser le concepteur et à quelles fins. On peut en effet distinguer trois classes d'hypermédias à base de connaissances : des hypermédias à base de connaissances sur l'organisation des documents multimédias, des hypermédias à base de connaissances sur le domaine et des hypermédias à base de connaissances sur l'utilisateur.

Les travaux relatifs aux hypermédias à base de connaissances sur l'utilisateur ont pour but d'offrir à l'utilisateur une aide automatique à la navigation, en calculant automatiquement les liens les plus prometteurs dans son contexte et/ou en restreignant automatiquement l'espace de navigation en fonction de son profil. Pour cela un modèle de l'utilisateur en situation est donné au système, c'est-à-dire un modèle de l'utilisateur (profil et but) et de son contexte de navigation. Il est soit construit au préalable soit appris à partir de l'observation de l'utilisateur en situation de navigation. Le lecteur intéressé peut par exemple se reporter à [Bareiss et Osgood, 93] et [Shibata et Katsumono, 93] pour la première approche et à [Boy, 91] et [Chang, 93] pour la seconde.

Les travaux relatifs aux hypermédias à base de connaissances sur l'organisation des documents portent sur la *conception*, c'est-à-dire l'organisation du réseau hypermédia. Les travaux de conception d'hypermédias à base de connaissances sur l'organisation des documents s'intéressent à fonder la structure du réseau hypermédia sur l'organisation sémantique des documents, la structure logique des unités logiques que constituent les blocs d'information [Balpe et al, 96]. Ces travaux s'inspirent des principes de l'édition de documents électroniques structurés [Vercoustre, 90][Quint et Vatton, 92].

Dans cette partie, nous présentons de manière plus détaillée les travaux relatifs aux hypermédias à base de connaissances du domaine, auxquels nous nous intéressons plus particulièrement. Ils s'adressent à la fois à la conception de l'hypermédia et à l'amélioration des possibilités d'accès à l'information sur le réseau construit. Nous décrivons les différentes approches de l'introduction de connaissances, acquises à partir des documents de l'hypermédia ou d'un expert du domaine. Nous mettons ensuite en évidence les différents modes de représentation des connaissances explicitées. Puis nous montrons quelles procédures d'accès à l'information sont proposées sur de tels hypermédias à base de connaissances.

4.2.1. Explicitation de connaissances du domaine

4.2.1.1. A partir des documents hypermédias

Les travaux de conception automatique de réseaux hypermédias à base de connaissances sur le contenu sémantique des documents reposent sur l'hypothèse que le texte des documents est porteur d'information suffisamment riche et analysable pour en extraire des liens hypertextes. Le principe général est d'expliciter les connaissances que peut fournir le texte des documents pour en extraire automatiquement un réseau de parcours possible sur ces documents. La conception automatique d'hypermédias à base sémantique s'apparente à la recherche d'information dans les bases documentaires et à l'acquisition de connaissances à partir de textes. Elle fait appel aux techniques d'analyse de textes syntaxiques ou statistiques [Balpe et al, 96].

L'analyse syntaxique et l'analyse statistique de documents sont deux approches complémentaires pour l'explicitation de connaissances à partir du contenu de documents. L'analyse syntaxique des documents permet d'établir un lexique des racines des termes présents dans les textes des documents. Un réseau hypertexte peut alors être constitué en reliant automatiquement les documents par une recherche "plein texte" des occurrences de racines communes. Ce réseau peut être encore raffiné, en faisant appel à des techniques d'analyse statistique. Essentiellement, il s'agit de vérifier la pertinence des termes utilisés comme ancres des liens du réseau. Les expériences relatées dans [Nanard et al, 93], [Salton et Allan, 93] et [Zizi et Beaudouin-Lafont, 95] illustrent l'utilisation de telles techniques pour la conception automatique de réseaux hypermédias.

A l'issue de cette analyse, le lexique constitué peut encore être enrichi par un réseau de liens sémantiques entre les termes de sorte que des liens hypertextes soient créés non seulement entre les documents caractérisés par l'occurrence d'un même terme mais aussi entre ceux dans lesquels apparaissent deux termes en relation dans le réseau du lexique. Cette méthode de conception est par exemple mise en oeuvre dans les Hyperlivres des éditions Ilias [Balpe et al, 96] et dans le système TAIGA [Marteau et Krumeich, 95].

4.2.1.2. Auprès du concepteur de l'hypermédia

Certaines limites des méthodes de conception automatique d'hypertextes que nous venons de présenter ont été mises en évidence [Nanard et Nanard, 91]. En particulier, la sémantique des liens est limitée à des relations de *référence*. L'approche dominante pour la conception d'applications hypermédias à base de connaissances du domaine est de fait celle qui consiste à expliciter la sémantique précise des liens hypermédias auprès de l'auteur de l'application, au moment de leur création - manuelle. Les motivations de l'auteur pour la création d'un noeud ou d'un lien contiennent en effet la sémantique du noeud ou du lien. Celle-ci est disponible au moment de la création du noeud ou du lien mais risque d'être perdue (oubliée) ultérieurement.

Dans cette approche, les connaissances du domaine de l'application hypermédia sont explicitées auprès de son concepteur en termes de relations sémantiques dont il s'agit de préciser la nature. Pour ce faire, le concepteur de l'application attribue des *types* aux liens, aux noeuds et aux ancres qu'il introduit dans le réseau hypermédia. En d'autres termes, des connaissances du domaine sont intégrées au réseau hypermédia sous la forme de *types* de liens, de noeuds et d'ancres. Le typage des liens hypermédias est une fonctionnalité caractéristique des systèmes de conception d'applications hypermédias à base de connaissances du domaine. C'est par exemple le cas des systèmes CONCORDE [Hofmann et al, 90], MacWeb [Nanard et Nanard, 91], SEPIA [Streitz et al, 92], Aquanet [Marschall et al, 92], ou HyperPro [Normark et Osterbye, 95].

Parmi ces systèmes, CONCORDE est le plus simple et ne permet pas de typer les noeuds hypermédias ; MacWeb, SEPIA, Aquanet et HyperPro permettent de typer à la fois les noeuds et les liens hypermédias ; seul MacWeb permet de typer les ancres [Nanard et Nanard, 93]. De manière générale, les types de noeuds et d'ancres représentent des concepts du domaine dont les noeuds et les ancres sont les instances ; les types de liens représentent des relations sémantiques dont les liens sont les instances. Il s'agit des relations `part-of`, `has-property`, etc. ou de relations dépendantes du domaine. Dans le système CONCORDE, les types sont simples. Dans SEPIA, une hiérarchie d'héritage simple est définie sur les types. Dans MacWeb et Aquanet, les types sont organisés en hiérarchies d'héritage multiple.

4.2.2. Représentation de connaissances du domaine

4.2.2.1. Réseaux sémantiques

Un réseau hypermédia supporte des connaissances du domaine et constitue un *modèle du domaine*. En ce sens, il peut être comparé à un réseau sémantique. Un réseau sémantique caractérise en effet les concepts d'un domaine en décrivant des relations entre eux. De façon similaire, les liens d'un réseau hypermédia représentent des relations sémantiques entre unités d'information. Un réseau sémantique caractérise des concepts, un hypermédia met ces concepts sous une forme directement perceptible par l'utilisateur. En d'autres termes, la structure d'un réseau hypermédia qui supporte des connaissances du domaine est calquée sur celle d'un réseau sémantique qui modéliserait le domaine de l'application hypermédia. Un réseau hypermédia peut donc être vu comme une “interface” qui rend un réseau sémantique accessible à l'utilisateur.

Considérons un réseau hypermédia construit automatiquement, par exemple avec le système TAIGA (cf. § 4.2.1.1.). Sa structure est celle d'un réseau sémantique simple, dont les liens sont tous des relations de référence. Considérons maintenant un réseau hypermédia construit en explicitant des connaissances d'un expert du domaine, par exemple avec le système CONCORDE, le plus rudimentaire des systèmes de conception d'applications hypermédias cités ci-dessus (cf. § 4.2.1.2.). On constate que ce réseau a un potentiel d'expressivité supérieur au précédent. En effet, il représente un réseau sémantique plus riche, en ce qu'il intègre plusieurs types de relations.

Dans un réseau sémantique, on distingue les noeuds *concepts* et les noeuds *instances* de ces concepts. Le typage des noeuds et ancres d'un réseau hypermédia revient à construire un réseau sémantique dont les noeuds instances sont les noeuds et ancres hypermédias et dont les noeuds concepts sont les types des noeuds et ancres hypermédias. Les types des liens hypermédias représentent les relations sémantiques entre les noeuds instances du réseau sémantique. Les réseaux hypermédias construits avec les systèmes Aquanet, SEPIA et MacWeb, dont les types des noeuds d'information entretiennent des relations d'héritage représentent des réseaux sémantiques dont les noeuds concepts entretiennent des relations d'héritage `is-a`. Le système MacWeb permet en outre de représenter d‘autres relations sémantiques entre les noeuds concepts qui dépendantes du domaine.

4.2.2.2. Frames

La comparaison d'un réseau hypermédia à une hiérarchie de frames est fondée sur l'association d'une liste d'attributs et d'un ensemble de méthodes aux noeuds et liens du réseau. C'est le cas des réseaux hypermédias construits avec les systèmes Aquanet, SEPIA et MacWeb. Les noeuds et liens d'un réseau hypermédia sont assimilés à des frames, les types de noeuds et de liens sont des classes de frames et la relation d'héritage sur les types est identique à une hiérarchie de classes de frames. Une classe hérite des attributs et des méthodes de la (des) classe(s) supérieure(s). La définition d'un type de noeuds ou de liens, c'est-à-dire d'une classe de frames, consiste à définir la liste des attributs des frames de cette classe et le type des objets qui peuvent instancier ces attributs. Streitz et al. mettent en évidence les atouts d'une représentation à base de frames pour la modélisation de structures hypermédias typées [Streitz et al., 92]. Elle permet d'une part une spécification déclarative de contraintes, telles que la validité des sources et destinations des liens, qui peuvent être prises en compte au cours de l'exécution des procédures d'organisation du réseau. Elle est d'autre part facilement extensible et maintenable.

4.2.3. Exploitation de connaissances du domaine

D'une manière générale, des connaissances déclaratives du domaine ont pour fonction d'améliorer l'accès à l'information en rendant le réseau hypermédia plus intelligible et donc mieux manipulable. L'introduction de connaissances dans les types des noeuds, liens et ancres du réseau et l'organisation de ces types en hiérarchies d'héritage assure au réseau hypermédia à la fois une sémantique riche et une forte structuration. Un tel réseau hypermédia, qui offre à l'utilisateur un modèle expressif du domaine, est à même de pallier les risques de désorientation et de surcharge cognitive de l'utilisateur au cours d'une session de navigation. Mais les connaissances introduites ne sont pas seulement destinées à l'utilisateur ; elles sont également explicitées pour être manipulées par le système. C'est ainsi que Lucarella définit la recherche d'information sur un réseau hypermédia comme une procédure d'inférence qui peut être accomplie par l'utilisateur qui explore le réseau (browsing) mais aussi par le système qui exploite le réseau hypermédia comme une base de connaissances (searching) [Lucarella, 90]. L'exploitation automatique des connaissances explicitées dépend de leur degré de formalisation.

4.2.3.1. Degrés de formalisation des connaissances

Des représentations déclaratives telles que les réseaux sémantiques ou les hiérarchies de frames se prêtent bien à l'intégration de procédures de recherche par résolution logique ou à base de règles et de techniques de décision heuristiques propres à l'IA qui améliorent les procédures de recherche. Parmi les systèmes de conception d'applications hypermédias précédemment cités, MacWeb permet de construire un réseau hypermédia non seulement interprétable en termes de réseau sémantique mais aussi effectivement implanté comme tel et doté des mécanismes de raisonnement automatiques propres à une telle structure de connaissances [Nanard et Nanard, 91].

Au contraire, les systèmes CONCORDE, Aquanet, SEPIA ou HyperPro ne sont pas dotés de capacités de raisonnement automatique. La finalité de ces systèmes est de présenter à l'utilisateur un modèle hypermédia du domaine. Cependant, dans Aquanet et HyperPro, le typage des noeuds et des liens supporte une forme de raisonnement automatique élémentaire pour gérer automatiquement une représentation graphique du réseau hypermédia et aider ainsi l'utilisateur à mieux appréhender celui-ci et à mieux s'y orienter. Dans Aquanet, les attributs des types (frames) ont trait à l'apparence graphique des noeuds et liens hypermédias. Dans HyperPro, des règles de présentation sont associées aux types des noeuds et liens.

4.2.3.2. Exploitation des connaissances par l'utilisateur et par le système

A mesure que la quantité d'information disponible dans un hypermédia augmente, il devient nécessaire de disposer de moyens auxiliaires de prise de décision, pour aider l'utilisateur à s'orienter et à naviguer ou pour rechercher automatiquement une information sur le réseau hypermédia [Woodhead, 91].

La mise en oeuvre de procédures de recherche automatique sur des connaissances du domaine intégrées à un réseau hypermédia concerne le domaine de la *recherche d'information* dans les hypermédia. Dans ce cadre, un hypermédia est envisagé comme une base de connaissances, utilisée pour construire des systèmes de recherche intelligents [Lucarella, 90]. De fait, certains travaux en recherche d'information font appel à des formalismes de représentation de connaissances tels que les réseaux sémantiques ou les graphes conceptuels [Myaeng, 92] [Kheirbek et Chiaramella, 95] [Martin, 96].

L'*aide à la navigation* sur des réseaux hypermédias intéresse marginalement le domaine de la recherche d'information. Dans le système I^3R [Croft et Turtle, 89] et le manuel médical électronique décrit dans [Frisse et Cousins, 89], il est fait appel à des stratégies de recherche d'information pour aider l'utilisateur au cours d'une session de navigation dynamique. L'ensemble des noeuds à explorer par l'utilisateur pour juger de leur pertinence eu égard à sa requête implicite est restreint grâce à un mécanisme de présélection automatique d'une liste de noeuds candidats. Cependant, les procédures d'aide à la navigation dont sont munis ces systèmes sont relativement limitées. Pour être efficace, un système d'aide à la navigation a en effet besoin de connaissances sur l'utilisateur en situation de navigation.

Les systèmes de conception d'applications hypermédias qui fondent l'organisation du réseau hypermédia sur des connaissances du domaine ne s'attachent que peu ou pas à l'opérationnalisation de ces connaissances pour la recherche d'information. Dans ces systèmes, on s'intéresse davantage à la structuration et à la présentation d'un réseau hypermédia à l'aide de connaissances, l'objectif étant d'offrir un réseau intelligible, c'est-à-dire structuré et doté de formats de présentation graphiques qui permettent à l'utilisateur de s'orienter. De fait, parmi les systèmes présentés, seul dans MacWeb les connaissances explicitées sont suffisamment formalisées pour être automatiquement manipulées.

4.3. Conception d'hypermédia à base de connaissances du domaine

Nous nous concentrons dans cette partie sur la *conception* d'applications hypermédias à base de connaissances du domaine. Nous précisons d'abord la problématique générale qui lui est associée, à savoir l'explicitation d'un modèle du domaine au cours de la conception du réseau hypermédia (§ 4.3.1.). Nous présentons ensuite différents méthodes et systèmes de conception d'applications hypermédias (§ 4.3.2.). Nous exposons alors les principes de notre approche de la conception d'applications hypermédias à base de connaissances du domaine (§ 4.3.3.).

4.3.1. Problématique

4.3.1.1. Du point de vue de l'utilisateur à celui du concepteur

Nous avons vu que, de manière générale, on intègre des connaissances dans un hypermédia pour rendre les procédures d'accès à l'information les plus performantes possibles (cf. § 4.2.3.). En particulier, des connaissances du domaine intégrées dans un hypermédia permettent à l'utilisateur d'appréhender aisément la structure générale du réseau hypermédia conçu à l'aide de ces connaissances et ainsi de mieux comprendre les choix de navigation associés au réseau. Essentiellement, il s'agit de pallier les problèmes auxquels l'utilisateur est susceptible d'être confronté : les risques de désorientation et de mauvaise compréhension de la structure générale du réseau, c'est-à-dire de la sémantique des liens entre les noeuds du réseau. Également, des connaissances du domaine sont interprétables par le système, automatiquement manipulées dans des procédures de recherche d'information.

Pour assurer la cohérence d'un réseau hypermédia et une navigation efficace sur ce réseau, le concepteur d'une application hypermédia s'attache à organiser au mieux le réseau hypermédia. Il s'agit plus précisément de concevoir des structures organisatrices et des formats de présentation des documents de l'application hypermédia à l'aide de connaissances du domaine. Or, en l'absence de méthodologie générale de conception, ces tâches constituent pour le concepteur une lourde charge cognitive. Il apparaît donc que pallier le problème de désorientation et d'incompréhension de la structure d'une application hypermédia par l'utilisateur requiert de pallier celui de surcharge cognitive du concepteur.

4.3.1.2. Le point de vue du concepteur

Pallier la surcharge cognitive du concepteur d'une application hypermédia à base de connaissances est un but intermédiaire à celui, final, d'améliorer l'accessibilité de l'information dans une application hypermédia en y intégrant des connaissances du domaine. En d'autres termes, pour qu'une application hypermédia soit le plus riche possible en structures organisatrices et formats de présentation, il faut que le système de conception d'applications hypermédias utilisé soit conçu de telle sorte qu'il limite la charge cognitive du concepteur.

La tâche d'explicitation de connaissances du domaine peut apparaître comme une charge cognitive pour le concepteur de l'application hypermédia qui vient s'ajouter à celle d'organisation du réseau hypermédia. Cependant, cette surcharge est largement compensée par l'utilité des connaissances explicitées pour la conception même de l'application. Une base de connaissances du domaine constitue en effet un schéma conceptuel qui fait office de *guide* pour la conception de l'application hypermédia. Plus précisément, la structure de la base de connaissances est le fondement de l'organisation du réseau hypermédia. Les travaux sur la conception d'hypermédias à base de connaissances du domaine s'attachent donc à optimiser la tâche (centrale) d'explicitation d'une base de connaissances structurée. Nous présentons dans la partie suivante les méthodes et systèmes de conception d'applications hypermédias à base de connaissances du domaine.

4.3.2. Méthodes et systèmes de conception d'applications hypermédias à base de connaissances du domaine

Dans la conception d'une application hypermédia, l'explicitation d'un modèle du domaine est essentiel. De celui-ci dépend en effet la qualité de l'organisation du réseau hypermédia. Le concepteur d'une application hypermédia est d'autant plus performant dans sa tâche d'organisation du réseau hypermédia qu'il est guidé par la base de connaissances du domaine qu'il a explicitées. Les travaux de recherche sur la conception d'applications hypermédias se concentrent donc sur l'explicitation d'une base de connaissances du domaine de qualité. Plus précisément, ils s'attachent d'une part à l'explicitation d'un modèle du domaine de l'application hypermédia suffisamment riche et structuré pour faciliter effectivement l'organisation du réseau hypermédia (§ 4.3.2.1.). Ils s'appliquent d'autre part à faciliter la tâche même d'explicitation de connaissances , à l'aide du réseau hypermédia en cours de construction (§ 4.3.2.2.).

4.3.2.1. Explicitation d'un modèle du domaine

Les différentes approches de la conception d'applications hypermédias à base de connaissance du domaine ont en commun de supporter le modèle du domaine qui doit servir de guide pour l'organisation du réseau hypermédia par un mécanisme de typage des noeuds et liens du réseau hypermédia et de structuration de ces types (cf. § 4.2.2.). Les types de noeuds représentent les concepts généraux du domaine de l'application hypermédia et permettent de distinguer entre différents types d'information. Les types de liens entre les noeuds représentent explicitement les relations sémantiques qu'entretiennent les noeuds du réseau hypermédia et permettent de distinguer entre différents types de relations.

Une organisation hiérarchique des types de noeuds et de liens, telle que dans les systèmes Aquanet [Marshall et al, 91 ; Marshall et Rogers, 92], SEPIA [Streitz et al, 92], permet de reconnaître dans le réseau hypermédia des structures régulières qui garantissent une organisation intelligible des informations. Une organisation en réseau sémantique des types de noeuds et de liens, telle que dans le système MacWeb [Nanard et Nanard, 91], est plus riche qu'une structure hiérarchique et accroît encore l'intelligibilité de l'organisation du réseau hypermédia.

Le modèle du domaine qui permet de guider la conception d'un réseau hypermédia est élaboré au fur et à mesure de la conception même du réseau. Pour sa construction, le concepteur est guidé dans sa tâche d'explicitation de connaissances par des schémas génériques que nous détaillons pour chacun des systèmes précédemment cités.

a) Aquanet

Dans le système Aquanet, l'explicitation d'un modèle du domaine est contrôlée à l'aide d'un *éditeur de schémas*. Un schéma de modèle permet au concepteur de définir l'ensemble des types de noeuds et de liens dont il peut déclarer instances les noeuds et liens hypermédias qu'il construit. Plus précisément, un modèle du domaine conçu avec Aquanet est supporté par une représentation orientée objet à base de frames. Les types de noeuds et de liens du réseau hypermédia sont représentés par des classes de frames.

L'ensemble des classes est organisé hiérarchiquement. Chaque classe est définie par un ensemble d'attributs et hérite des attributs des classes plus générales.

Un schéma de modèle consiste donc, dans Aquanet, en une hiérarchie de classes dont les attributs sont définis et les valeurs qu'ils peuvent prendre sont contraintes. Ce schéma explicite la nature et la structure du réseau hypermédia souhaitées par le concepteur. Les noeuds et liens du réseau sont en effet définis comme des instances de classes appartenant au schéma ; ils héritent des attributs de celles-ci.

b) SEPIA

La représentation et l'explicitation d'un modèle du domaine dans les systèmes SEPIA et Aquanet sont similaires. Ainsi, tout comme dans Aquanet, un modèle du domaine est supporté dans SEPIA par une représentation orientée objet à base de frames. Plus précisément, SEPIA intègre un "moteur hypermédia" indépendant, Hyperbase [Schütt et Streitz, 90], dédié aux définitions de types et de schémas de modèles. Il est comparable à l'éditeur de schémas couplé avec l'éditeur de types du système Aquanet.

Le système Hyperbase permet au concepteur d'une application hypermédia de définir un schéma de modèle à l'aide de types de liens et de noeuds, qui sont représentés par des sous-classes des classes les plus générales HB_Node et HB_Link. Un schéma de modèle consiste donc dans SEPIA en deux hiérarchies de classes : une hiérarchie de classes représentant des types de noeuds et une hiérarchie de classes représentant des types de liens hypermédias. Les noeuds et les liens du réseau hypermédia sont définis comme instances des classes de ces deux hiérarchies.

c) MacWeb

Dans le système MacWeb, la représentation d'un modèle du domaine est supportée par une structure de types orientée objet plus riche que celles des systèmes SEPIA et Aquanet. Les relations entre types ne sont en effet pas limitées à une relation d'héritage. Le modèle qu'explicite le concepteur d'une application MacWeb est donc susceptible d'avoir une sémantique et une structure plus riches que les modèles conçus avec SEPIA ou Aquanet.

Comme dans SEPIA ou Aquanet, les types sont définis dans MacWeb comme des classes et les noeuds et liens hypermédias comme des instances de ces classes. L'originalité de MacWeb réside en ce que les types de noeuds et de liens, tout comme les noeuds hypermédias, sont des objets entre lesquels l'héritage n'est qu'une relation parmi d'autres - exprimées sous la forme de types de liens. En d'autres termes, les types de noeuds et de liens, comme les noeuds et les liens, sont organisés en réseau et non simplement en hiérarchie d'héritage.

Un schéma de modèle consiste donc, dans MacWeb, en une hiérarchie ou un réseau de classes définies à partir du schéma générique constitué des trois classes les plus générales ChunkType, LinkType et Inherit [Nanard et Nanard, 91]. Ce schéma de modèle est défini par le concepteur de l'application hypermédia et sa cohérence est contrôlée par le système. Comme dans Aquanet et SEPIA, les noeuds et les liens du réseau hypermédia sont définis comme instances des classes appartenant au schéma de modèle élaboré.

4.3.2.2. Explicitation d'un modèle hypermédia du domaine

Nous avons mis en évidence dans la partie précédente le processus d'explicitation d'un modèle du domaine à partir d'un schéma générique spécialisé au fur et à mesure de la conception d'une application hypermédia. Dans cette partie, nous nous intéressons aux implications de la simultanéité des processus d'explicitation du modèle du domaine et d'organisation du réseau hypermédia. Nous dégageons de l'étude des systèmes Aquanet, SEPIA et MacWeb le rôle de la tâche d'organisation du réseau hypermédia dans celle d'explicitation du modèle du domaine.

Dans les différentes approches de la conception d'applications hypermédias à base de connaissances du domaine, le concepteur de l'application fait appel au modèle du domaine qu'il explicite pour organiser le réseau hypermédia et, symétriquement, il fait appel au réseau hypermédia qu'il organise pour expliciter un modèle du domaine. Plus précisément, il est guidé dans sa tâche de spécialisation d'un schéma générique de modèle du domaine par un modèle de la tâche d'organisation d'un réseau hypermédia. Ce modèle peut ainsi être vu comme une méthode d'explicitation d'un modèle du domaine. Nous le détaillons pour chacun des systèmes précédemment décrits.

a) Aquanet

Le système Aquanet est présenté par ses auteurs comme un outil non seulement de présentation d'information mais aussi d'explicitation et de représentation de la structure sous-jacente de cette information [Marshall et al, 91]. Plus précisément, Aquanet est dédié à la *modélisation hypermédia* d'un domaine, c'est-à-dire à la modélisation d'un domaine à l'aide des capacités de représentation graphique des hypermédias. Les auteurs défendent en effet l'hypothèse que si dans un premier temps le concepteur d'une application hypermédia fait appel à certaines connaissances du domaine et certaines structures de ces connaissances pour amorcer l'organisation du réseau hypermédia, il développe dans un deuxième temps, à partir du réseau déjà élaboré, de nouveaux schémas de structuration de ses connaissances. Ceux-ci sont utilisés au fur et à mesure qu'ils sont explicités pour améliorer l'organisation du réseau hypermédia.

La conception d'une application hypermédia avec Aquanet procède donc d'un cycle qui alterne explicitation de connaissances du domaine à l'aide du support de représentation graphique que constitue le réseau hypermédia et organisation du réseau hypermédia à l'aide des connaissances explicitées, des schémas de modèle qui émergent de ces connaissances. Dans le système Aquanet, la tâche d'organisation du réseau hypermédia est donc centrale dans celle d'explicitation d'un modèle (hypermédia) du domaine.

b) SEPIA

Dans la présentation du système SEPIA, ses auteurs mettent l'accent sur la notion d'*environnement de conception* d'applications hypermédias [Streitz et al, 92]. Plus précisément, SEPIA est fondé sur un modèle du processus cognitif de conception. Les aspects dynamiques de la conception d'une application hypermédia sont modélisés par la division de l'“espace de conception” en “sous-espaces d'activité” en interaction. En particulier, le sous-espace de rhétorique qui contient le réseau hypermédia destiné aux utilisateurs résulte de la combinaison du sous-espace dans lequel est développé un modèle du domaine et du sous-espace dans lequel est développé un schéma d'argumentation.

Ainsi, dans SEPIA, l'explicitation d'un modèle du domaine est nécessaire à l'organisation d'un réseau hypermédia mais elle est aussi guidée par la réalisation d'un tel réseau et plus précisément par les interactions entre l'espace de modélisation du domaine et les autres espaces d'activité. En effet, le modèle du domaine et le schéma d'argumentation de l'application hypermédia sont représentés par des réseaux hypermédias, qui consistent chacun en un noeud composé, sous-ensemble des noeuds et liens du réseau hypermédia final. Le modèle hypermédia du domaine est ainsi élaboré en interaction avec l'espace d'argumentation et l'espace de rhétorique. La tâche d'explicitation d'un modèle hypermédia du domaine dans SEPIA est donc indissociable et guidée par celles d'argumentation et de rhétorique.

c) MacWeb

Dans le système MacWeb, à la différence de SEPIA ou Aquanet, le modèle du domaine explicité pour une application hypermédia est formalisé, sous une forme exploitable par des procédures de recherche d'information à base de connaissances générales relatives au domaine [Nanard et al, 93]. Ces connaissances du domaine sont également appréhendables par l'utilisateur à travers le réseau hypermédia. En ce sens, MacWeb est comparable à Aquanet ou SEPIA : les connaissances explicitées constituent bien un modèle hypermédia du domaine [Nanard et Nanard, 91].

Plus précisément, le concepteur d'une application MacWeb explicite un modèle du domaine dans le même environnement que celui dans lequel il organise un réseau hypermédia. La base de connaissances et le réseau hypermédia étant représentées et manipulées sous la forme d'un réseau sémantique unique, le concepteur de l'application hypermédia est guidé dans sa tâche d'explicitation d'un modèle du domaine par son environnement de travail orienté vers l'organisation du réseau hypermédia. Ainsi, de manière comparable à Aquanet, l'organisation d'un réseau hypermédia avec MacWeb est fondée sur un modèle du domaine dont l'explicitation est elle-même guidée par la tâche d'organisation du réseau hypermédia. En outre, les intentions et les effets de rhétorique du concepteur lorsqu'il crée un noeud ou un lien hypermédia, sans être explicités comme dans SEPIA, interviennent cependant dans le processus d'explicitation du modèle.

4.3.2.3. Conclusion

De l'étude des principaux systèmes de conception d'applications hypermédias à base de connaissances, il ressort, au-delà de leurs différences, une préoccupation commune chez leurs auteurs d'alléger la charge cognitive du concepteur dans sa tâche d'organisation du réseau hypermédia. Pour ce faire, une tâche d'explicitation d'un modèle du domaine est ajoutée à celle d'organisation du réseau hypermédia, le modèle du domaine explicité devant servir de guide dans l'organisation du réseau hypermédia. De fait, les différentes approches que nous avons présentées se concentrent sur la tâche d'explicitation de connaissances du domaine auprès du concepteur de l'application hypermédia (expert du domaine). Le mode de représentation des connaissances du domaine commun aux différentes approches consiste en un typage des noeuds et liens hypermédias et une organisation hiérarchique ou en réseau sémantique de ces types. Le schéma du modèle du domaine destiné à fonder le réseau hypermédia est construit à partir d'un schéma générique de modèle sous la forme de types généraux de noeuds et liens hypermédias. Enfin, le modèle du domaine est élaboré suivant un modèle de la tâche d'organisation du réseau hypermédia, l'hypothèse sous-jacente à cette démarche étant que, lorsqu'il explicite un modèle du domaine, le concepteur de l'application hypermédia pense dans les termes du réseau hypermédia.

4.3.3. Une approche originale de la conception d’applications hypermédias à base de connaissances du domaine

Dans cette partie, nous proposons une approche de la conception d’applications hypermédias comparable à celles qu’implémentent les systèmes décrits dans la partie précédente. Nous nous intéressons en effet à l’organisation de réseaux hypermédias par indexation de leurs éléments à une base de connaissances qui modélise le domaine et dont la structure constitue un fil directeur pour le concepteur de l’application. La tâche d’explicitation des connaissances du domaine est donc centrale dans notre approche, notre objectif étant de fournir au concepteur expert du domaine un formalisme de représentation et une méthode pour expliciter une base de connaissances à même de fonder l’organisation d’un réseau hypermédia.

L'originalité de notre approche réside d'une part dans le choix du formalisme des graphes conceptuels pour représenter un modèle du domaine. Son adéquation pour la représentation de connaissances taxinomiques et la mise en oeuvre de procédures d’identification a été mise en évidence au chapitre 1. Nous montrons ici que, de plus, ce formalisme se prête bien à la structuration d’un réseau hypermédia (§ 4.3.3.1.).

D’autre part, notre approche s’éloigne de la démarche commune aux travaux décrits qui consiste à guider l’explicitation de connaissances par la tâche d’organisation du réseau hypermédia. A l'inverse, nous proposons une méthode de conception d'un hypermédia qui consiste à guider le processus d’organisation du réseau hypermédia par celui d’explicitation des connaissances. Plus précisément, nous faisons appel à une méthode d’acquisition de connaissances pour aider le concepteur dans sa tâche d’explicitation d’un modèle du domaine. Le réseau hypermédia est organisé au fur et à mesure de l’explicitation de ce modèle, en indexant des noeuds et des liens hypermédias aux éléments de la base de connaissances (§ 4.3.3.2.).

4.3.3.1. Explicitation d’un modèle du domaine représenté par des graphes conceptuels

Les domaines pour lesquels nous nous intéressons à la conception d’applications hypermédias sont ceux des sciences naturelles caractérisés par une organisation taxinomique et une forte structuration des objets. L’adéquation du formalisme des graphes conceptuels pour la modélisation de tels domaines a été mise en évidence dans le chapitre 1. Nous proposons donc de faire appel à un modèle du domaine représenté par des graphes conceptuels pour guider le concepteur d'une application hypermédia dans l'organisation du réseau hypermédia. Nous montrons dans cette partie que le formalisme des graphes conceptuels se prête à l'organisation d'un réseau hypermédia sémantiquement riche et structuré. Nous proposons un mécanisme de typage des noeuds et des liens d'un réseau hypermédia à l'aide des éléments d'une base de connaissances du domaine représentée dans le formalisme des graphes conceptuels. Nous décrivons ensuite un modèle générique représenté à l'aide de graphes et à partir duquel expliciter les connaissances du domaine.

a) Typage de noeuds et de liens hypermédias dans le formalisme des graphes conceptuels

Afin de construire un schéma de modèle du domaine pour organiser un réseau hypermédia, nous proposons un typage des noeuds et des liens hypermédias à l'aide des “briques” de base du formalisme des graphes conceptuels : les types de concepts et de relations conceptuelles. Dans notre approche, les noeuds et les ancres hypermédias sont vus comme des *référents* de types de concepts et les liens hypermédias comme des types de relations conceptuelles ou des relations de généralisation. De la sorte, les relations de généralisation définies sur les types de concepts et les types de relations conceptuelles et les relations conceptuelles qu'entretiennent les concepts qui interviennent dans des graphes conceptuels décrivant les objets du domaine constituent un modèle pour organiser un réseau hypermédia.

Les domaines des sciences naturelles pour lesquels nous nous intéressons à l'organisation d'un réseau hypermédia se prêtent bien à une modélisation par une base de connaissances taxinomiques représentées dans le formalisme des graphes conceptuels (cf. § 1.4.) :

- les types de concepts définis par des graphes conceptuels représentent les taxons du domaine ;
- les relations de généralisation définies sur ces types de concepts représentent la structure taxinomique du domaine ;
- les types de concepts et les relations conceptuelles qui interviennent dans les graphes de définition des taxons représentent les primitives du langage de description du domaine ; ils sont ordonnés par des relations de généralisation.

Tout type de concept ou de relation conceptuelle est donc en relation avec ses descendants et ancêtres dans la hiérarchie de généralisation des types de concepts ou des relations conceptuelles. Les types de concepts qui constituent les primitives du langage de description des taxons sont en outre en relations conceptuelles avec les autres types de concepts des graphes de définition dans lesquels ils interviennent.

Ainsi, l'ensemble des types des noeuds hypermédias n'est pas simplement organisé hiérarchiquement - par la relation de généralisation ; il est doté d'une structure sémantiquement plus riche, sous la forme de relations conceptuelles. En cela, notre approche est tout à fait comparable à celle adoptée dans le système MacWeb. Soulignons à ce propos que le formalisme des graphes conceptuels est issu des réseaux sémantiques auxquels fait appel la représentation d'un modèle du domaine dans MacWeb. Cependant, les graphes conceptuels sont un formalisme plus puissant encore que les réseaux sémantiques. Il permettent en effet de définir des types de concepts structurés, sous la forme de graphes, et de distinguer entre l'organisation hiérarchique de ces types définis et leurs définitions sous forme de graphes. Le formalisme des graphes conceptuels permet donc un typage et donc une organisation du réseau hypermédia sémantiquement plus riches et plus structurés qu'avec des réseaux sémantiques.

b) Explicitation d'un modèle du domaine à partir d'une base canonique

De manière comparable aux approches existantes, nous proposons d'organiser le réseau hypermédia au fur et à mesure que le modèle du domaine est explicité. Plus précisément, les noeuds, les ancres et les liens hypermédias sont créés et typés au fur et à mesure que de nouveaux types de concepts atomiques ou définis et de nouveaux types de relations conceptuelles sont introduits pour enrichir le modèle du domaine. L'explicitation incrémentale d'un modèle du domaine est donc centrale dans l'organisation du réseau hypermédia et, en ce sens, notre approche est comparable à celles décrites dans la partie précédente. Nous définissons un modèle conceptuel générique de domaine, que nous spécialisons incrémentalement pour expliciter un modèle d'un domaine particulier.

Dans le formalisme des graphes conceptuels, la notion de modèle générique est définie par ce qu'on appelle une base canonique (cf. § 1.3.2.2.). Une base canonique est un ensemble “noyau” de graphes déclarés canoniques, c'est-à-dire sémantiquement corrects. Ces graphes sont utilisés comme générateurs d'autres graphes conceptuels (canoniques) à l'aide des règles de formation canonique que sont les opérations de spécialisation définies dans le formalisme [Mugnier et Chein, 93]. Une base canonique permet d'imposer des contraintes au modèle du domaine explicité, sous la forme de graphes conceptuels.

La démarche que nous proposons est de construire une base canonique en faisant appel au concepteur de l'hypermédia, expert du domaine. Celui-ci est en effet à même d'imposer des contraintes liées au domaine. En particulier, il est requis pour construire un graphe canonique `G` qui représente la description minimale des spécimens de la taxinomie. Ce graphe sera le générateur principal de la base de connaissances puisque le type de concept défini par ce graphe constitue la racine de la taxinomie à expliciter. Les autres taxons seront représentés par des types définis par des graphes construits en spécialisant ce graphe canonique.

D'autre part, nous avons identifié certaines contraintes liées non pas directement au domaine mais au type des domaines pour lesquels nous nous intéressons à l'explicitation d'un modèle. En d'autres termes, eu égard à l'organisation taxinomique de ces domaines et la structure des descriptions des objets du domaine, nous avons mis en évidence un schéma général de base canonique. Toute base canonique contient le type de concept le plus général T_c (le type défini par le graphe canonique `G` sera inséré sous ce type) et le type de relation le plus général T_r. Les types de concepts `composant` et `caractéristique` appartiennent également à toute base canonique, définis comme fils de T_c et frères du type défini par `G,` ainsi que les relation conceptuelles binaires `caractérisation` et `composition`, fils de T_r. Ces types rendent compte de la logique de décomposition adoptée par les systématiciens qui décrivent un spécimen par ses parties et sous-parties et les caractéristiques de celles-ci. Les types de concepts qui interviennent dans la définition du graphe `G` sont donc inférieurs aux types `composant` ou `caractéristique`, et les types de relations inférieurs aux relations `composition` ou `caractérisation`.

Ce schéma général d'une base canonique permet donc d'imposer certaines contraintes qui rendent compte des caractéristiques des domaines à modéliser et de la pratique des systématiciens. A partir de ce schéma de base canonique, le concepteur d'une application hypermédia particulière définit une base canonique particulière, pour un domaine particulier en introduisant un graphe `G` qui décrit la racine de la taxinomie ainsi que les primitives de langage qu'il utilise dans `G` (des types de concepts et de relations atomiques). Nous pensons que ce que perd notre approche en généricité par l'imposition d'un schéma pour la construction d'une base canonique, elle le gagne en terme d'assistance au concepteur de l'application hypermédia dans l'explicitation d'une base canonique puis dans l'explicitation d'un modèle du domaine, guidée par une méthode d'acquisition de connaissances basée sur la spécificité de la base canonique explicitée.

4.3.3.2. Explicitation d'un modèle du domaine selon une méthode d'acquisition de connaissances

Dans les approches de la conception d'applications hypermédias à base de connaissances que nous avons présentées dans la partie précédente, l'explicitation d'un modèle du domaine est guidée par la tâche d'organisation du réseau hypermédia pour laquelle le modèle du domaine fait office de guide. En particulier dans le système Aquanet, ce point de vue est clairement exprimé : penser en termes d'organisation d'un réseau hypermédia constitue une méthode d'explicitation de connaissances [Marschall et al., 91]. Nous proposons au contraire d'envisager l'explicitation d'un modèle du domaine en termes de construction d'une base de connaissances. En d'autres termes, notre approche s'éloigne de celles que nous avons présentées, dans lesquelles la conception d'une application hypermédia procède d'un cycle alternant explicitation d'un modèle du domaine guidée par la tâche d'organisation du réseau hypermédia et organisation du réseau hypermédia guidée par le modèle du domaine explicité.

En effet, dans notre cas, eu égard à l'organisation taxinomique des domaines auxquels nous nous intéressons et la logique de composition qui structure les descriptions de taxons, la tâche d'organisation du réseau hypermédia n'est pas utile à l'explicitation d'un modèle du domaine. De ce fait, dans notre approche, le concepteur d'une application hypermédia pense davantage en termes de base de connaissances que d'hypermédia pour expliciter un modèle du domaine. L'organisation du réseau hypermédia est corollaire à l'explicitation d'un modèle du domaine ; elle consiste en la création de noeuds, d'ancres et de liens hypermédias indexés aux éléments de la base de connaissances.

Plus précisément, nous proposons d'organiser le réseau hypermédia au fur et à mesure de l'explicitation d'un modèle du domaine, ce modèle étant explicité selon la méthode d'acquisition de connaissances décrite dans le chapitre 3. Cette méthode d'acquisition de connaissances est dédiée à la spécialisation du schéma de modèle que nous avons défini sous la forme d'une base canonique qui intègre les caractéristiques du domaine à modéliser (cf. § 4.3.3.1.). Nous l'avons choisie en fonction des caractéristiques du domaine et du formalisme des graphes conceptuels dans lequel nous représentons les connaissances du domaine ; elle est de fait fondée sur les propriétés du formalisme des graphes conceptuels.

Ainsi, l'organisation d'un réseau hypermédia est réalisée non seulement suivant la structure de la base de connaissances explicitée qui modélise le domaine mais aussi suivant la méthode d'explicitation mise en oeuvre pour construire ce modèle. Les noeuds et les ancres du réseau hypermédia sont créés incrémentalement, au fur et à mesure de l'explicitation du modèle, comme référents des types de concepts nouvellement introduits ; les liens entre les noeuds ou les ancres sont les relations de généralisation entre les types dont ces noeuds ou ancres sont référents et les relations conceptuelles binaires qu'entretiennent, dans les graphes de définition, les concepts dont les types ont pour référents ces noeuds ou ancres hypermédias. L'organisation d'un réseau hypermédia est ainsi réalisée en indexant ses éléments à la base de connaissances explicitée. Elle est partiellement automatisée grâce à la mise en oeuvre d'une méthode d'acquisition de connaissances qui fait aussi bien office de méthode d'organisation d'un réseau hypermédia. Plus généralement, l'organisation d'un réseau hypermédia par indexation automatique de ses éléments à ceux d'un modèle du domaine explicité selon une méthode d'acquisition de connaissances constitue une méthode de conception d'application hypermédia.

4.3.3.3. Conclusion

Dans ce chapitre, nous avons présenté la problématique générale de l'intégration de connaissances aux réseaux hypermédias pour la conception d'applications hypermédias intelligibles par leurs utilisateurs, c'est-à-dire dotées de possibilités d'accès conceptuel à l'information mémorisée, par navigation ou recherche d'information, qui pallient le risque de désorientation des utilisateurs. Nous avons plus particulièrement développé la présentation des approches qui s'intéressent à l'intégration de connaissances du domaine, en mettant en évidence l'intérêt d'un modèle du domaine pour l'organisation d'un réseau hypermédia sémantiquement riche. La problématique associée à la conception d'applications hypermédias à base de connaissances du domaine est donc la représentation et l'intégration d'un modèle du domaine à un réseau hypermédia et la réduction de la charge cognitive du concepteur que représente cette tâche d'explicitation de connaissances. C'est dans ce cadre que s'inscrit notre approche de la conception d'applications hypermédias.

Notre approche est comparable aux autres approches de la conception d'applications hypermédias à base de connaissances du domaine en ce que le concepteur est guidé dans sa tâche d'organisation d'un réseau hypermédia par la structure de la base de connaissances qui modélise le domaine et plus précisément par le typage des noeuds, ancres et liens hypermédias. De ce point de vue, l'originalité de notre approche réside dans le choix des graphes conceptuels pour représenter le modèle du domaine. Ce formalisme assure en effet un typage et donc une organisation du réseau hypermédia sémantiquement plus riches et plus structurés que ceux basés sur une simple hiérarchie de types ou un réseau sémantique. Le second aspect de notre approche est la définition d'un schéma de modèle à partir duquel le concepteur définit une base canonique destinée à générer un modèle du domaine. A la différence des schémas proposés dans les autres approches, ce schéma est dépendant des caractéristiques communes aux domaines auxquels nous nous intéressons, à savoir les domaines des sciences naturelles caractérisés par une organisation taxinomique et des descriptions de taxons structurées. Nous pensons que ce que perd notre approche en généricité, le concepteur d'une application hypermédia le gagne en terme d'aide à l'explicitation d'un modèle du domaine. En particulier, nous mettons en oeuvre une méthode d'acquisition de connaissances à partir d'une base canonique construite sur ce schéma et dans le formalisme des graphes conceptuels. C'est le troisième aspect de notre approche : l'organisation d'un réseau hypermédia est confondue avec l'explicitation d'un modèle du domaine, guidée par une méthode d'acquisition de connaissances. A l'inverse des autres approches où le concepteur pense en termes d'hypermédias et explicite un modèle hypermédia du domaine, notre approche de la conception d'une application hypermédia est de penser en termes de base de connaissances. Nous détaillons notre démarche dans la partie suivante.

4.4. Construction d'un hypermédia à base de connaissances : l'approche SATELIT

Dans cette partie, nous commençons par présenter les différents types de noeuds, ancres et liens hypermédias que nous avons définis préalablement à la construction d'un hypermédia à base de connaissances pour un domaine particulier (§ 4.4.1.). Ils sont indépendants du domaine mais ils dépendent de la nature des connaissances taxinomiques à acquérir et du formalisme des graphes conceptuel dans lequel sont représentées les connaissances taxinomiques sous-jacentes au réseau hypermédia. Nous proposons ensuite une méthode de construction d'un hypermédia à base de connaissances, fondée sur la méthode d'acquisition d'une base de connaissances taxinomiques décrite dans la partie précédente (§ 4.4.2.). Nous présentons alors le réseau hypermédia ainsi construit automatiquement et montrons quels liens peuvent encore être ajoutés au réseau, manuellement mais cependant guidé par les types de noeuds, ancres et liens hypermédias dépendants du domaine explicités au cours de la construction du réseau initial (§ 4.4.3.). Nous examinons alors les modes d'accès à l'information disponibles dans un hypermédia à base de connaissances ainsi construit. Ils consistent en primitives de navigation basées sur la structure de la base de connaissances. En outre, un calcul dynamique de taxon à partir de sa description fournie par l'utilisateur peut être mis en place, fondé sur la manipulation formelle des éléments de la base de connaissances.

4.4.1. Types primitifs de noeuds et liens hypermédias

4.4.1.1. Types primitifs liés à la structure taxinomique

Dans un hypermédia à base de connaissances construit selon SATELIT, les descriptions formelles de taxons sous la forme de graphes conceptuels sont supportées par des noeuds hypermédias de même type : le type F_TAXON que nous avons initialement défini. Aux descriptions formelles des taxons sous forme de graphes conceptuels, nous souhaitons indexer différents documents multimédias, en particulier des descriptions textuelles et graphiques (cf. § 4.3.3.). Pour cela, nous avons introduit deux types de noeuds hypermédias : le type T_TAXON et le type G_TAXON. Deux noeuds de types T_TAXON et G_TAXON relatifs à un même taxon peuvent être vus comme les référents dans l'espace des descriptions textuelles et l'espace des descriptions graphiques du type de concept représentant ce taxon et défini sur un noeud de type F_TAXON. Les types F_TAXON, T_TAXON, et G_TAXON se généralisent en le type TAXON. Pour accéder à partir d'un noeud de l'un de ces trois types à un noeud d'un autre de ces trois types relatif au même taxon, nous avons introduit le type de lien IN_TAXON défini entre deux noeuds de type TAXON. On indexerait de la même manière d'autres types de documents tels que des animations ou des documents sonores en introduisant de nouveaux types de noeuds hypermédias inférieurs au type TAXON : les types A_TAXON et S_TAXON.

Nous avons également défini différents liens entre deux noeuds hypermédias de type TAXON, qui reflètent la structure de la taxinomie, c'est-à-dire les relations de généralisation entre les graphes conceptuels qu'ils supportent. Un lien de type DESCENDANT_ANCETRE est établi entre un noeud de type TAXON supportant une description d'un taxon et les noeuds de même type supportant des descriptions plus générales, c'est-à-dire celles des ancêtres de ce taxon. Un lien de type PERE_FILS permet d'accéder à partir d'un noeud de type TAXON au noeud de même type supportant une description de l'un des taxons fils choisi par l'utilisateur. Un lien de type FRERE_FRERE permet d'accéder à partir d'un noeud de type TAXON à un noeud de même type supportant une description de l'un des taxons frères choisi par l'utilisateur. Ces types de liens se spécialisent respectivement en F_DESCENDANT_ANCETRE, T_DESCENDANT_ANCETRE et G_DESCENDANT_ANCETRE ; en F_PERE_FILS , T_PERE_FILS et G_PERE_FILS ; et en F_FRERE_FRERE, T_FRERE_FRERE et G_FRERE_FRERE.

Enfin, nous définissons le type de noeuds TAXINOMIE. Un noeud de ce type supportera la représentation graphique d'une taxinomie. Pour une application hypermédia donnée, un seul noeud sera de ce type. Nous définissons également le type d'ancres N_TAXON et le type de liens N_TAXON_TAXON défini entre une ancre de type N_TAXON et un noeud de type TAXON. Une ancre de type N_TAXON appartient à un noeud de type TAXINOMIE et, à partir d'une ancre de type N_TAXON, un lien de type N_TAXON_TAXON permet d'accéder à l'un des trois noeuds de type TAXON choisi par l'utilisateur et sur lequel figure une description formelle, textuelle ou graphique du taxon considéré.

4.4.1.2. Types primitifs liés au support des graphes conceptuels

La démarche que nous avons adoptée dans la partie précédente pour indexer des descriptions textuelles et graphiques à un type de concept représentant un taxon peut être généralisée pour indexer différents documents multimédias à un type de concept atomique, c'est-à-dire non défini. Plus précisément, afin que les éléments de la terminologie d'un domaine à modéliser soient facilement appréhendables, nous indexerons aux types de concepts qui représentent formellement ces primitives de langage des descriptions textuelles, des dessins et photos. Pour cela, nous introduisons les types de noeuds T_PRIMITIVE et G_PRIMITIVE et le type PRIMITIVE qui généralise ces deux types ; nous introduisons également le type de lien IN_PRIMITIVE défini entre deux noeuds de type PRIMITIVE qui permet, à partir de la description textuelle d'un type de concept atomique d'accéder à celle graphique et inversement. Enfin, le type de noeud TYPE_CONCEPT généralise les types PRIMITIVE et TAXON et le type de lien IN_TYPE_CONCEPT défini entre deux noeuds de type TYPE_CONCEPT généralise les types IN_PRIMITIVE et IN_TAXON.

Pour offrir une vue générale de la hiérarchie des types atomiques qui sera construite, similaire à celle de la taxinomie, nous introduisons le type de noeuds TERMINOLOGIE. Un noeud de ce type supportera la représentation graphique de la hiérarchie des types explicitée. Pour une application hypermédia donnée, un seul noeud sera de ce type. Nous définissons également le type d'ancres N_PRIMITIVE et le type de liens N_PRIMITIVE_PRIMITIVE défini entre une ancre de type N_PRIMITIVE et un noeud de type PRIMITIVE. Une ancre de type N_PRIMITIVE appartient à un noeud de type TERMINOLOGIE et à partir d'une telle ancre un lien de type IN_PRIMITIVE_PRIMITIVE permet d'accéder à l'un des deux noeuds de type PRIMITIVE choisi par l'utilisateur et sur lequel figure une description textuelle ou graphique du type de concept atomique considéré.

4.4.1.3. Types primitifs liés aux graphes conceptuels

Les relations qui existent entre types de concepts atomiques ne sont pas seulement des relations de généralisation. En effet, dans les graphes qui définissent les types représentant des taxons, les concepts entretiennent des relations de composition et de caractérisation. Nous introduisons donc deux types de noeud inférieurs au type PRIMITIVE : les types COMPOSANT et CARACTERISTIQUE ; nous introduisons encore les types T_COMPOSANT et T_CARACTERISTIQUE inférieurs au type T_PRIMITIVE et respectivement au type COMPOSANT et au type CARACTERISTIQUE, et les types G_COMPOSANT et G_CARACTERISTIQUE inférieurs au type G_PRIMITIVE et respectivement au type COMPOSANT et au type CARACTERISTIQUE. Le type de liens COMPOSITION est défini entre deux noeuds de types COMPOSANT ; le type CARACTERISATION est défini entre un noeud de type COMPOSANT et un noeud de type CARACTERISTIQUE.

Enfin, pour rendre accessibles, à partir d'un noeud de type F_TAXON qui supporte un graphe conceptuel décrivant un taxon, les informations relatives au type d'un sommet concept de ce graphe, nous introduisons le type d'ancre N_CONCEPT et le type de lien N_CONCEPT_PRIMITIVE défini entre une ancre de type N_CONCEPT et l'un des deux noeuds de type PRIMITIVE relatifs au type du concept considéré.

La figure 4.2. ci-dessous récapitule l'ensemble des types de noeuds, ancres et liens primitifs que nous avons définis indépendamment des domaines pour lesquels des hypermédias à base de connaissances pourront être construits selon SATELIT.

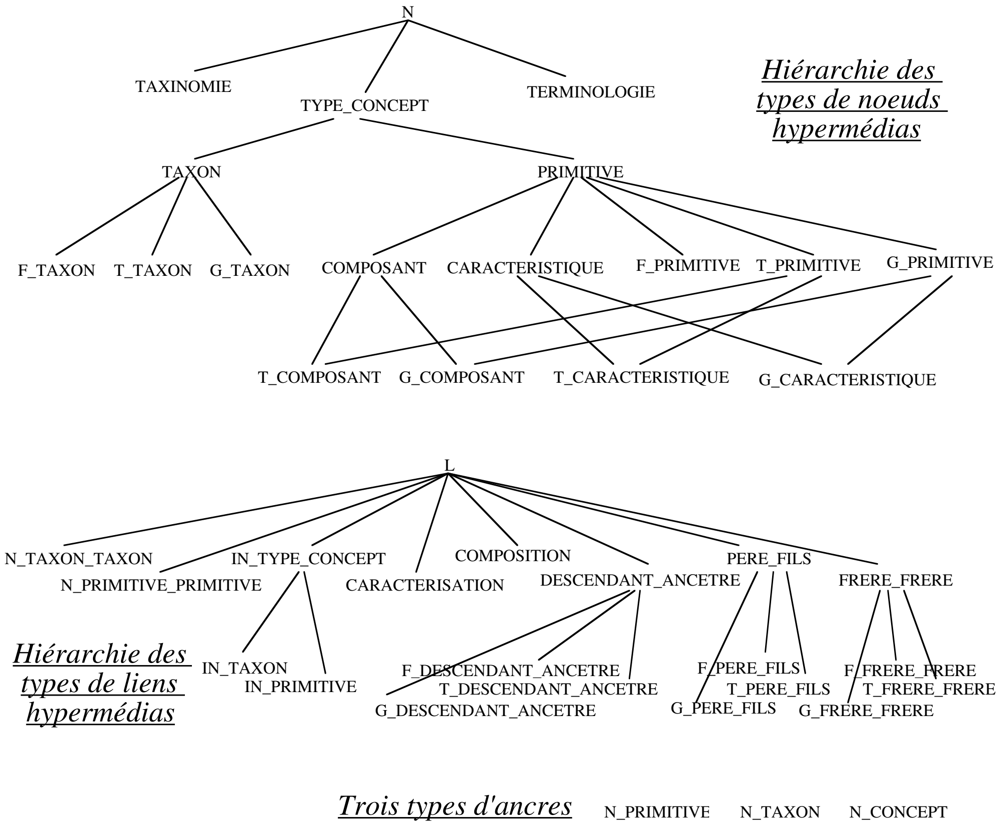


figure 4.2. Types primitifs de SATELIT

D'autres types seront ultérieurement introduits automatiquement au cours de la construction d'un hypermédia à base de connaissances pour un domaine particulier. A la différence de ceux que nous avons présentés dans cette partie, ils dépendront du domaine modélisé.

4.4.2. Acquisition d'une base de connaissances hypermédia

Dans cette partie, nous décrivons la méthode de construction d'un hypermédia à base de connaissances mise en oeuvre dans SATELIT. Elle utilise les types primitifs définis dans la partie précédente et suit le processus d'acquisition de connaissances taxinomiques décrit dans le chapitre 3. Dans la perspective de construire un hypermédia, l'acquisition incrémentale de descriptions de taxons est vue ici comme l'introduction successive dans un réseau hypermédia de noeuds d'information supportant chacun la description formelle d'un taxon sous la forme d'un graphe conceptuel. Les noeuds, ancres et liens hypermédias sont créés de façon semi-automatique, incrémentalement, en même temps que les connaissances sont explicitées, lorsqu'un nouveau taxon ou un nouveau type de concepts est acquis. Après avoir décrit la phase d'initialisation du processus de construction d'un hypermédia, nous reprenons l'algorithme d'acquisition en indiquant pour chacune de ses trois phases quels nouveaux noeuds, ancres et liens sont introduits dans le réseau hypermédia et quels nouveaux types sont créés qui dépendent du domaine et viennent enrichir la structure du réseau construit.

4.4.2.1. Phase d'initialisation

Préalablement à la mise en oeuvre de l'algorithme d'acquisition de connaissances taxinomiques, la construction d'un hypermédia avec SATELIT débute par la création de deux premiers noeuds du réseau hypermédia : celui destiné à supporter une représentation de l'ensemble de la taxinomie, de type TAXINOMIE, et un autre destiné à supporter celle de la hiérarchie des types de concept, de type TERMINOLOGIE. Sur ce dernier figurent initialement les noms des types prédéfinis de noeuds du réseau hypermédia inférieurs au type PRIMITIVE. Un troisième noeud est également créé destiné à interfacer une procédure d'identification d'objets (cf. § 4.3.1.) ; ce noeud est accessible à partir de tous les autres noeuds du réseau hypermédia grâce à des liens de type IDENTIFICATION construits automatiquement.

4.4.2.2. Acquisition d'une taxinomie hypermédia

La première phase de l'algorithme d'acquisition de connaissances taxinomiques mis en oeuvre dans SATELIT est l'explicitation d'une description formelle initiale d'un taxon sous forme de graphe conceptuel (cf. § 3.2.). Elle débute tout naturellement par la création d'un nouveau noeud hypermédia de type F_TAXON destiné à supporter ce graphe et par la mise à jour du noeud de type TAXINOMIE. Une nouvelle ancre de type N_TAXON est ajoutée sur ce noeud et un lien de type N_TAXON_TAXON est créé à partir de cette ancre vers le noeud nouvellement créé. L'introduction dans le réseau hypermédia de ce noeud de type F_TAXON s'accompagne également de la création d'un noeud de type T_TAXON et d'un noeud de type G_TAXON, destinés à recevoir les descriptions textuelles et graphiques du taxon. Des liens de type IN_TAXON sont créés entre ces trois noeuds.

Dans le même temps, trois nouveaux types de noeuds sont créés, inférieurs aux types F_TAXON, T_TAXON et G_TAXON. Ils ont pour noms le nom du taxon acquis, préfixé respectivement par F, par T et par G. Ce sont les types les plus spécifiques des trois noeuds hypermédias nouvellement créés qui supportent les descriptions formelle, textuelle et graphique du taxon. Ces types sont insérés dans la hiérarchie des types de noeuds hypermédias sous les trois types correspondant au père du taxon. De la sorte, la hiérarchie initiale des types de noeuds hypermédias présentée sur la figure 4.2. est enrichie au fur et à mesure de l'acquisition incrémentale des taxons. La figure 4.3. ci-dessous présente un extrait de la hiérarchie de types de noeuds créée lors de la construction d'un hypermédia à base de connaissances en orchidologie ; elle correspond à la hiérarchie des types de concepts inférieurs au type ORCHIDALE sur la figure 1.11.

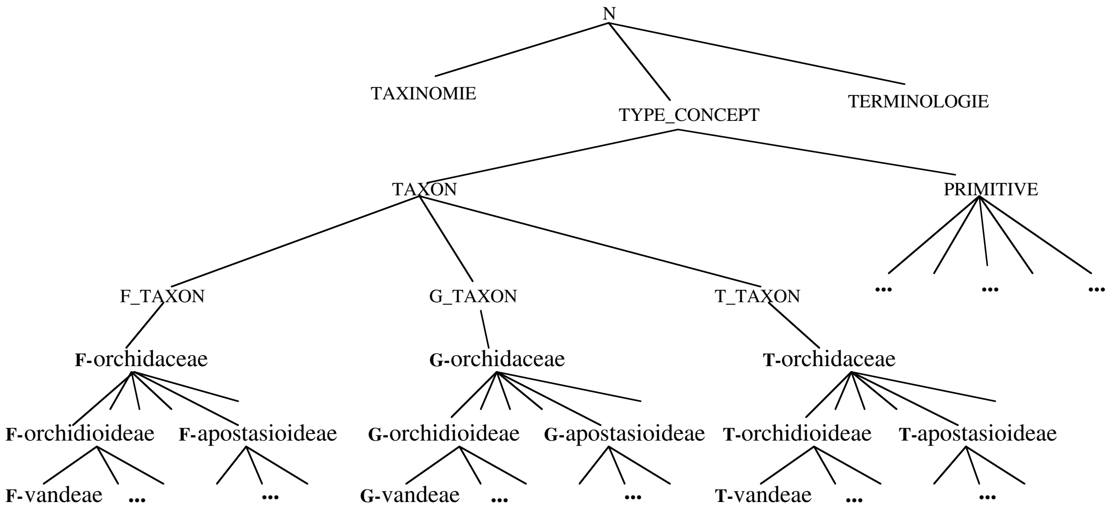


figure 4.3. Enrichissement de la hiérarchie des types de noeuds hypermédias par des types inférieurs à TAXON

Une fois créés les trois noeuds hypermédias de type TAXON supportant les descriptions formelle, textuelle et graphique d'un même taxon, des liens hypermédias sont créés à partir de chacun de ces trois noeuds. Ils représentent les relations taxinomiques qu'entretient ce taxon avec les autres taxons dont les descriptions sont déjà acquises :

- trois liens de type DESCENDANT_ANCETRE vers les noeuds supportant les descriptions des ancêtres du taxon décrit, c'est-à-dire dont le type est supérieur au type de ce dernier ;
- trois liens de type FRERE_FRERE vers les descriptions correspondantes des taxons frères ; si aucun taxon frère n'a encore été acquis, ces liens ne sont pas actifs ;
- trois liens de type PERE_FILS vers les descriptions correspondantes des taxons fils ; aucun taxon fils n'étant pour l'instant décrit, ces liens sont inactifs.

Les liens de type PERE_FILS issus chacun de l'un des trois noeuds supportant une description du taxon père et les liens de type FRERE_FRERE issus de chacun des noeuds supportant les différentes descriptions des taxons frères sont mis à jour.

Considérons l'état du réseau hypermédia à base de connaissances au moment de l'acquisition d'un taxon T_n. Soit T son père dans la taxinomie et soit LT = $\{T_1, \ldots\ T_{n-1}\}$ la liste de ses n-1 taxons frères déjà acquis. L'algorithme mis en oeuvre lors de la première phase de l'acquisition de T_n est décrit sur la table 4.4 ci-dessous.

```
Création des noeuds $F_n$, $T_n$, $G_n$ de types respectifs F-T, T-T et G-T.
Création des types de noeuds F-$T_n$, T-$T_n$ et G-$T_n$ inférieurs à F-T, T-T, et G-T.
Explicitation d'une description formelle de $T_n$ sur le noeud $F_n$.
Création des liens :
    $(F_n, T_n)$, $(F_n, G_n)$ et $(T_n, G_n)$ de type IN_TAXON.
    $(F_n, \{F_1,...F_{n-1}\})$, $(T_n, \{T_1,...T_{n-1}\})$ et $(G_n, \{G_1,...G_{n-1}\})$ de type FRERE_FRERE.
    $(F_n,\{\})$, $(T_n,\{\})$, $(G_n,\{\})$ de type PERE_FILS.
    $(F_n,\{F,...F\_ROOT\})$, $(T_n,\{T,...T\_ROOT\})$, $(G_n,\{G,...G\_ROOT\})$ de type DESCENDANT_ANCETRE.
Création d'une ancre $N_n$ de type N_TAXON sur le noeud de type TAXINOMIE
    et création d'un lien $(N_n,\{N_n, N_n, G_n\})$ de type N_TAXON_TAXON.
Pour i:=1..n-1 :
    $(F_i,\{F_j\}_{j=1..n-1,j\neq i}):=(F_i,\{F_j\}_{j=1..n,j\neq i})$
    $(T_i,\{T_j\}_{j=1..n-1,j\neq i}):=(T_i,\{T_j\}_{j=1..n,j\neq i})$
    $(G_i,\{G_j\}_{j=1..n-1,j\neq i}):=(G_i,\{G_j\}_{j=1..n,j\neq i})$
$(F,\{F_1,...F_{n-1}\}):=(F_i,\{F_1,...F_n\})$
$(T,\{T_1,...T_{n-1}\}):=(T_i,\{T_1,...T_n\})$
$(G,\{G_1,...G_{n-1}\}):=(G_i,\{G_1,...G_n\})$
```

table 4.4. Algorithme d'acquisition d'un taxon

La figure 4.4. suivante présente l'ensemble des liens nouvellement créés ou mis à jour lors de l'acquisition d'un taxon T_n ; ils sont issus ou arrivent aux noeuds Fn, Tn et Gn qui supportent les descriptions du taxon T_n. Les noeuds Fj, Tj et Gj supportent les descriptions d'un frère de T_n ; les noeuds Fi, Ti, et Gi les descriptions d'un ancêtre de T_n. Le noeud de forme triangulaire supporte la représentation de la taxinomie.

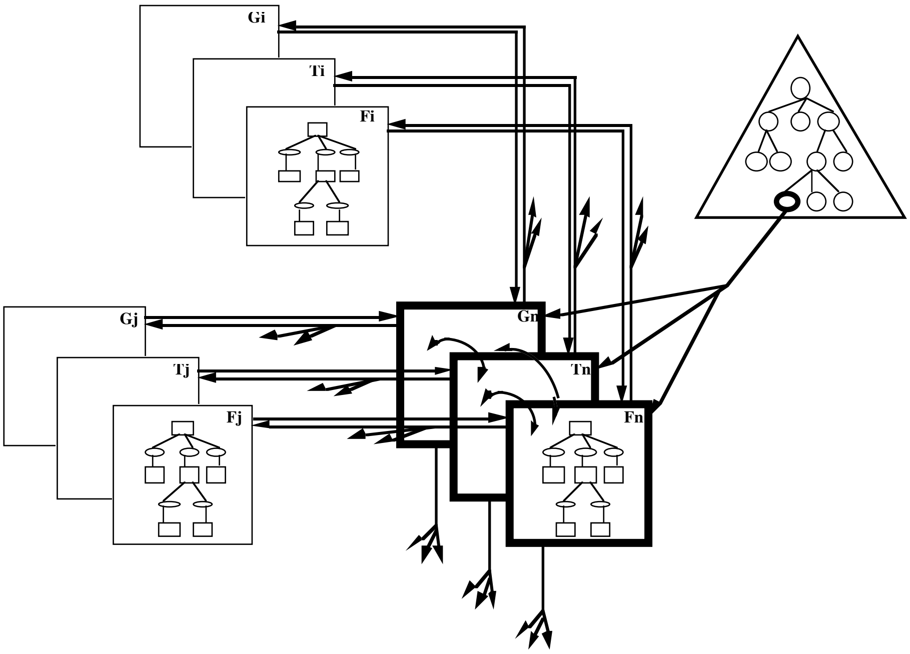


figure 4.4. Créations et mises à jour des liens du réseau hypermédia lors de l'acquisition d'un taxon T_n

4.4.2.3. Acquisition d'une hiérarchie de types de concepts hypermédia

Dans la seconde phase de l'algorithme d'acquisition - celle de complétion du graphe conceptuel qui décrit le taxon acquis (cf. § 3.2.) - de nouveaux types de concepts atomiques sont utilisés pour spécialiser ce graphe par restriction de concept. Dans la troisième phase de l'algorithme - la discrimination de la description formelle du taxon acquis de celles de ses frères (cf. § 3.3.) -, en cas d'échec de la discrimination, les descriptions des taxons sont spécialisées par joint avec de nouveaux arcs conceptuels puis par restriction de concept. Dans les deux cas, les nouveaux types de concepts utilisés sont introduits dans le support au fur et à mesure de l'acquisition des descriptions de taxons. En termes de construction du réseau hypermédia, l'occurrence de nouveaux types de concepts dans la description formelle d'un taxon se traduit d'abord par la mise à jour du noeud de type TERMINOLOGIE par l'ajout d'ancres portant le nom des nouveaux types de concepts. L'occurrence de nouveaux types de concepts s'accompagne également de la création de noeuds hypermédias destinés à supporter des informations textuelles ou graphiques sur ces primitives du langage de description. Pour chaque type de concept t nouvellement utilisé, il est proposé à l'utilisateur d'introduire dans le réseau hypermédia un noeud de type G_COMPOSANT ou G_CARACTERISTIQUE et un noeud de type T_COMPOSANT ou T_CARACTERISTIQUE, selon que le type de concept t est inférieur au type COMPOSANT ou CARACTERISTIQUE. Dans le même temps, des types de noeuds sont créés dont les noms sont ceux des types de concepts introduits, préfixés de la lettre T ou G. Ainsi, la hiérarchie des types de noeuds inférieurs au type PRIMITIVE reflète celle des types de concepts du support. La figure 4.5. ci-dessous présente un extrait de la hiérarchie des types de noeuds créée lors de la construction d'un hypermédia à base de connaissances en orchidologie ; elle correspond à la hiérarchie des types de concepts inférieurs aux types COMPOSANT et CARACTERISTIQUE sur la figure 1.11.

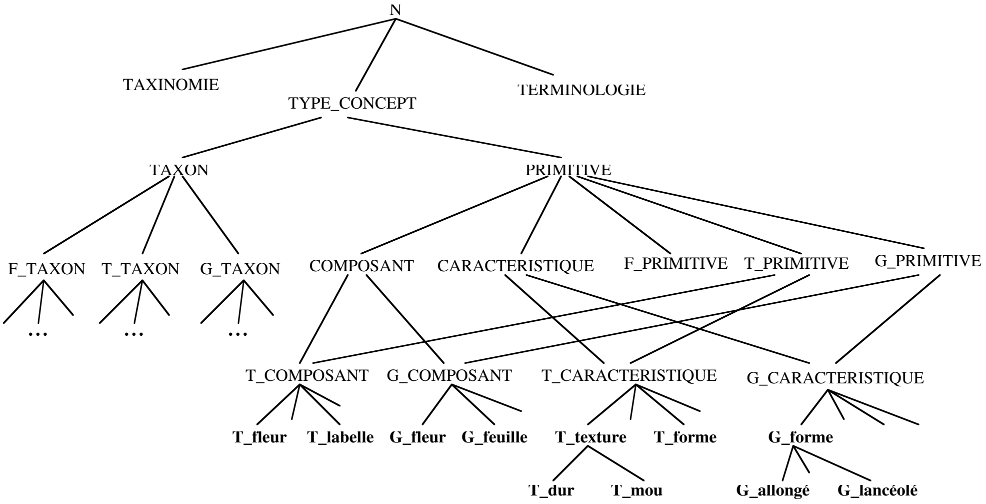


figure 4.5. Enrichissement de la hiérarchie des types de noeuds hypermédias par des sous-types de PRIMITIVE

Les noms des types de concepts qui apparaissent sur le noeud de type F_TAXON supportant la description formelle en cours d'explicitation, et pour lesquels existe un noeud de type PRIMITIVE supportant des informations à leur sujet constituent des ancres de type N_CONCEPT. A partir de ces ancres, des liens de type N_CONCEPT_PRIMITIVE sont créés vers les noeuds de type PRIMITIVE.

A partir de chaque nouvelle ancre de type N_PRIMITIVE du noeud de type TERMINOLOGIE, des liens de type N_PRIMITIVE_PRIMITIVE sont créés vers le noeud de type PRIMITIVE supportant des informations sur les primitives de langage désignées.

Enfin, vers ou à partir de chaque noeud de type PRIMITIVE nouvellement créé, un lien de type CARACTERISATION ou COMPOSITION est créé issu de ou allant vers le noeud de type PRIMITIVE qui supporte des informations sur le type de concept qui apparaît dans un même arc conceptuel de la description formelle du taxon en cours d'explicitation.

Soit P une primitive du langage de description des taxons, c'est-à-dire un type de concept atomique. Soit T le premier taxon dont la description requiert l'usage de P et soit

F le noeud hypermédia de type F_TAXON qui supporte T. Soit P' le type du concept source de l'arc conceptuel de la description de T dans lequel P apparaît et soient TP' et GP' (s'ils existent) les noeuds de type T_PRIMITIVE et G_PRIMITIVE qui supportent les informations associées à P'. L'algorithme mis en oeuvre lors de l'acquisition de P pour spécialiser la description de T est décrit sur la table 4.5 ci-dessous :

```
Création des ancres :
  NC de type N_CONCEPT sur le noeud F,
  NP de type N_PRIMITIVE sur le noeud de type TERMINOLOGIE.
Création des noeuds TP et GP de types T_PRIMITIVE et G_PRIMITIVE.
Création des types T_TP et G_GP :
  Si TP et GP sont de types T_COMPOSANT et G_COMPOSANT,
    T_TP et G_GP sont inférieurs à T_COMPOSANT et G_COMPOSANT ;
  Sinon, TP et GP sont de types T_CARACTERISTIQUE et G_CARACTERISTIQUE,
    T_TP et G_GP sont inférieurs à T_CARACTERISTIQUE et G_CARACTERISTIQUE.
Création des liens :
  (TP, GP) de type IN_PRIMITIVE,
  (NC, TP) de type N_CONCEPT_PRIMITIVE,
  (NP, TP) de type N_PRIMITIVE_PRIMITIVE,
  (TP', TP) et (GP', GP) de type COMPOSITION ou CARACTERISATION, selon
    que TP et GP sont de type COMPOSANT ou de type CARACTERISTIQUE.
```

table 4.5. Acquisition d'un nouveau type de concept atomique

La figure 4.6. suivante présente l'ensemble des liens nouvellement créés ou mis à jour lors de l'introduction d'un nouveau type de concept P dans la description formelle d'un taxon T ; ils sont issus ou arrivent aux noeuds TP et GP qui supportent les descriptions de P. Les noeuds TP' et GP' supportent les descriptions du type du concept source de l'arc conceptuel de la description formelle de T dans lequel P apparaît. Le noeud F supporte la description formelle de T. Le noeud de forme trapézoïdale est le noeud de type TERMINOLOGIE qui supporte la représentation de la hiérarchie des types atomiques.

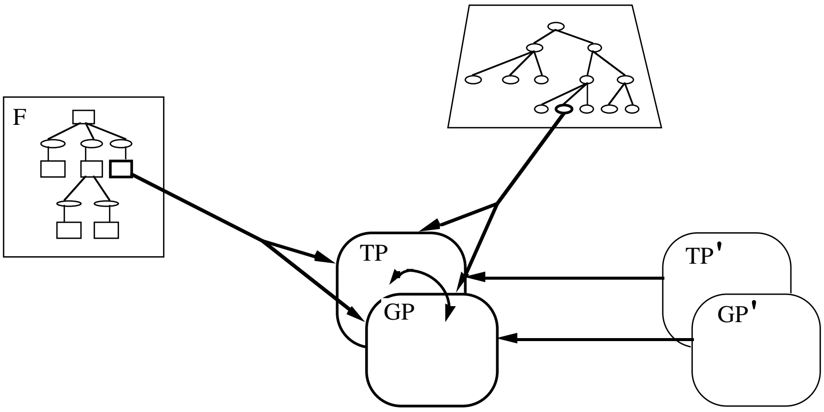


figure 4.6. Création de liens du réseau hypermédia lors de l'acquisition d'une primitive P

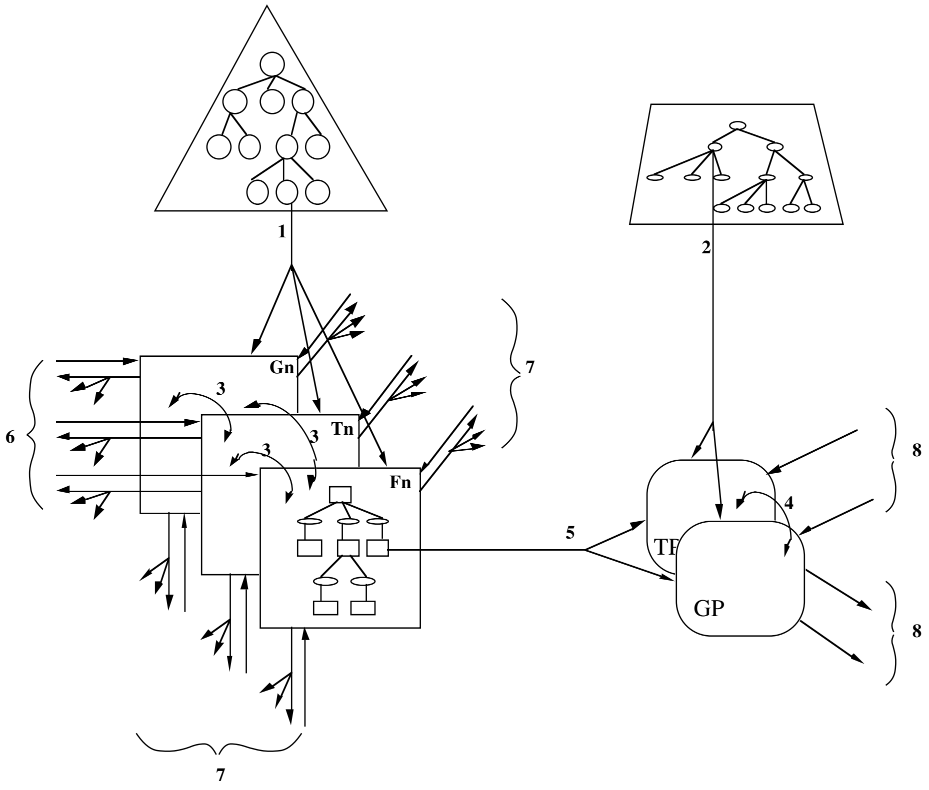
Nous décrivons dans la partie suivante le réseau hypermédia construit de façon semi-automatique selon la méthode que nous venons de présenter ; il s'appuie sur la structure logique de la base de connaissances sous-jacente qui est acquise. Nous montrons que ce réseau peut encore être enrichi manuellement, guidé par les types que nous avons définis.

4.4.3. Vers un hypermédia à base de connaissances taxinomiques

4.4.3.1. Réseau hypermédia initial

La structure initiale d'un réseau hypermédia construit selon la méthode que nous venons de décrire s'appuie sur la correspondance entre les noeuds, ancres et liens hypermédias et les objets du formalisme des graphes conceptuels. Les types de noeuds et d'ancres définis au cours du processus de construction du réseau hypermédia correspondent aux types de concepts de la base de connaissances ; les types de liens correspondent aux types de relations conceptuelles et aux relations de généralisation entre les types de concepts. De la sorte, les éléments du réseau hypermédia sont indexés à ceux de la base de connaissances sous-jacente.

La figure 4.7. ci-après présente la structure initiale d'un réseau hypermédia construit avec SATELIT. A ce stade, les noeuds de types T_TAXON, G_TAXON, T_PRIMITIVE et G_PRIMITIVE sont encore vides d'information ; il s'agit d'importer les textes, croquis ou images qu'ils sont destinés à supporter.



Types des liens :

1 : N_TAXON_TAXON
2 : N_PRIMITIVE_PRIMITIVE
3 : IN_TAXON
4 : IN_PRIMITIVE
5 : N_CONCEPT_PRIMITIVE
6: FRERE_FRERE
7 : PERE_FILS et DESCENDANT_ANCETRE
8 : COMPOSITION ou CARACTERISATION

figure 4.7. Structure initiale d'un réseau hypermédia construit avec SATELIT

4.4.3.2. Enrichissement du réseau hypermédia

A l'issue de cette phase principale de construction semi-automatique d'un réseau hypermédia, il est possible d'enrichir ce réseau initial en y introduisant manuellement de nouveaux liens et de nouveaux noeuds.

a) Création de liens hypermédias

Sur les noeuds de type T_TAXON qui supportent des descriptions textuelles de taxons, certaines chaînes de caractères du texte désignent des primitives du langage de description des taxons explicité. Avec ces chaînes, on peut créer des ancres de type N_CONCEPT et, à partir de ces ancres, des liens de type N_CONCEPT_PRIMITIVE. La création de tels liens permettra d'accéder aux informations textuelles ou graphiques relatives à un composant ou une caractéristique d'un taxon à partir de la description textuelle de ce taxon.

Sur les noeuds de type T_PRIMITIVE qui supportent des informations sur les primitives du langage de description explicité, certaines chaînes de caractère désignent d'autres primitives du langage. Avec ces chaînes, on peut créer des ancres et, à partir de ces ancres, des liens de type COMPOSITION ou CARACTERISATION.

Dans les légendes des croquis supportés par les noeuds de type G_TAXON et ceux de type G_PRIMITIVE, les chaînes de caractères peuvent également constituer des ancres à partir desquelles créer des liens, respectivement de type N_CONCEPT_PRIMITIVE et de type COMPOSITION ou CARACTERISATION. Des zones délimitées de ces croquis peuvent également faire office d'ancres pour de tels liens.

b) Création de noeuds

De nouveaux noeuds peuvent être introduits dans le réseau hypermédia, en particulier des noeuds supportant des index. Un index alphabétique de mots clefs est habituel dans les documents papier. Il est de ce fait utile de proposer à l'utilisateur d'un hypermédia, outre la hiérarchie des types de concepts du noeud de type TERMINOLOGIE, un index alphabétique des primitives du langage de description. Sur le noeud supportant cet index, les chaînes de caractères désignant les primitives constituent des ancres de type N_PRIMITIVE. De ces ancres, partent des liens symétriques de type IN_TERMINOLOGIE vers les ancres de même type du noeud de type TERMINOLOGIE et des liens de type N_PRIMITIVE_PRIMITIVE vers les noeuds de types PRIMITIVE supportant les informations textuelles ou graphiques sur les primitives désignées.

Également, un index alphabétique des différents taxons peut être constitué, qui offre un “point d'entrée” dans le réseau hypermédia alternatif au noeud de type TAXINOMIE supportant la représentation de la taxinomie. Sur le noeud supportant cet index, les chaînes de caractères désignant les taxons constituent des ancres de type N_TAXON dont partent des liens symétriques de type IN_TAXINOMIE vers les ancres de même type du noeud de type TAXINOMIE et des liens de type N_TAXON_TAXON vers les noeuds de type TAXON supportant les descriptions des taxons.

Enfin, en sciences naturelles, dans les descriptions de taxons, figurent toujours les noms des systématiciens qui les ont décrits. Un glossaire des systématiciens est donc utile dans un hypermédia sur un domaine des sciences naturelles. Pour cela, il faut introduire dans le réseau hypermédia un nouveau noeud sur lequel les chaînes de caractères qui désignent les systématiciens constituent des ancres pour des liens vers les noeuds de type TAXON supportant les descriptions des taxons.

Dans la partie suivante, nous décrivons les procédures d'accès à l'information sur un hypermédia à base de connaissances construit avec SATELIT.

4.4.4. Accès à l'information

Dans cette partie, nous décrivons les possibilités d'accès à l'information mémorisée dans un hypermédia à base de connaissances construit avec SATELIT. Ils consistent en une navigation *conceptuelle* sur le réseau hypermédia (§ 4.4.4.1.). La structure de la base de connaissances taxinomiques, aux éléments de laquelle sont indexés les éléments du réseau hypermédia, donne en effet un sens aux cheminements de l'utilisateur et le préserve du risque de désorientation. De plus, les descriptions formelles des taxons sous-jacentes au réseau hypermédia permettent de mettre en oeuvre des procédures dynamiques d'accès à l'information, essentiellement pour l'identification d'objets et plus généralement la recherche d'un taxon particulier à partir d'une description formelle construite avec les primitives de langage explicitées et en interrogeant l'utilisateur de SATELIT (§ 4.4.4.2.). Ces procédures sont fondées sur la manipulation formelle des connaissances sous-jacentes au réseau hypermédia ; elles consistent en des opérations sur les graphes conceptuels.

4.4.4.1. Navigation conceptuelle

L'indexation des éléments d'un réseau hypermédia à ceux d'une base de connaissances donne à celui-là une sémantique formelle qui le rend plus intelligible à l'utilisateur (cf. § 4.1.). En particulier, les informations que supporte un réseau hypermédia construit avec SATELIT sont organisées à l'aide de deux notions structurantes : la taxinomie et le langage de description des taxons (cf. § 4.4.2).

A partir du point d'entrée dans le réseau hypermédia que constitue son noeud de type TAXINOMIE, on accède aux descriptions (formelle, textuelle et graphique) d'un taxon choisi mémorisées sur les noeuds de type TAXON dédiés à ce taxon, en suivant les liens de type N_TAXON_TAXON issus de l'ancre de type N_TAXON désignant ce taxon. Puis, à partir d'un tel noeud, les autres noeuds (de même type) dédiés aux autres taxons sont accessibles en cheminant le long des liens de types DESCENDANT_ANCETRE, PERE_FILS et FRERE_FRERE qui expriment les relations de généralisation entre les descriptions formelles des taxons, les relations de généralisation entre les types de noeuds hypermédias, les relations taxinomiques entre les taxons.

Le second point d'entrée dans le réseau hypermédia est le noeud de type TERMINOLOGIE qui permet d'accéder à une primitive choisie du langage de description des taxons, c'est-à-dire de la terminologie du domaine modélisé. A partir des informations relatives à une primitive, on accède aux informations relatives à une autre primitive grâce aux liens de types COMPOSITION et CARACTERISATION entre les noeuds de type PRIMITIVE qui représentent les relations de même nom entre ces primitives. Dans les termes du formalisme des graphes conceptuels, deux primitives dont l'une désigne un composant ou une caractéristique de l'autre sont représentées par deux types de concepts présents dans un même arc conceptuel et le type de relation dans cet arc correspond au type de lien entre les noeuds de type PRIMITIVE.

Le passage entre ces deux espaces de navigation, la taxinomie et le langage de description, est assuré par les liens de type N_CONCEPT_PRIMITIVE entre un noeud de type TAXON supportant une description d'un taxon et un noeud de type PRIMITIVE supportant des informations relatives à une primitive de langage utilisée dans la description de ce taxon. Dans le cas de l'introduction d'un glossaire des systématiciens qui constitue un troisième point d'entrée dans le réseau hypermédia, le passage entre cet espace et les deux autres est assuré par les liens entre le noeud qui supporte le glossaire et le noeud supportant une description d'un taxon choisi.

Des espaces “secondaires” de navigation consistent chacun en l'ensemble des informations relatives à un seul taxon ou une seule primitive de langage. Il sont parcourus en cheminant le long de liens de types IN_TAXON ou IN_PRIMITIVE. Dans les termes du formalisme des graphes conceptuels, ces liens permettent d'accéder aux référents des types de concepts dans l'espace des descriptions textuelles et dans celui des descriptions graphiques.

Dans la partie suivante, nous montrons que la manipulation formelle des connaissances sous-jacentes au réseau hypermédia est un moyen d'accroître l'ensemble des liens disponibles à la navigation.

4.4.4.2. Accès dynamique à l'information

La manipulation formelle des connaissances permet de calculer dynamiquement des liens et dispense ainsi le réseau d'une trop grande complexité (cf. § 4.1.). Le but est d'accéder à un taxon quelconque à partir de la description qu’en fournit l'utilisateur, construite à l’aide des primitives de langage explicitées au cours de la construction de l'hypermédia à base de connaissances.

La procédure d'identification d'objets décrite dans le chapitre 3, interfacée sur un noeud du réseau hypermédia, est accessible à partir de tous les autres noeuds (cf. § 4.4.4.1.). Le résultat de la mise en oeuvre de cette procédure est la création dynamique d'un lien vers le noeud de type TAXON qui supporte la description du taxon dont l'objet décrit par l'utilisateur est reconnu être une instance.

Sur le même principe, on peut envisager la mise en place d'autres procédures de manipulation formelle des connaissances sur lesquelles sont indexées les éléments du réseau hypermédia. De telles procédures permettent d’accroître les possibilités d'accès à l'information mémorisée sur les noeuds du réseau hypermédia, grâce à la création dynamique de liens vers les noeuds qui supportent les informations relatives aux résultats des recherches. Par exemple, il peut être utile pour l'utilisateur de disposer d'un module de recherche des taxons (les plus généraux) qui possèdent certains traits descriptifs, ou encore, sur un noeud de type TAXON, d'un module de recherche des fils du taxon décrit qui possèdent certains traits descriptifs particuliers.

Dans tous les cas, la création dynamique de liens hypermédias vers les noeuds supportant l'information recherchée passe par des opérations sur les connaissances auxquelles sont indexées les noeuds hypermédias supportant cette information. Elles consistent en des appariements entre les graphes construits par l’utilisateur à partir des primitives de langage explicitées et les graphes décrivant les taxons. Le typage des éléments du réseau hypermédia qui reflète la structure de la base de connaissances sous-jacente assure un passage direct de celle-ci à celui-là. Une recherche d'information sur le réseau hypermédia est formulée en termes des éléments de la base de connaissances et son résultat est traduit en termes des éléments du réseau hypermédia.

4.5. Conclusion

Dans ce chapitre, après avoir brièvement présenté la problématique générale de l’intégration de connaissances aux réseaux hypermédias, nous nous sommes concentrés sur les approches de l’intégration de connaissances du domaine pour l’organisation d’un réseau hypermédia sémantiquement riche. La problématique associée à la conception d’hypermédias à base de connaissances du domaine est la représentation et l’intégration d’un modèle du domaine à un réseau hypermédia pour la réduction de la charge cognitive du concepteur.

Notre approche de la conception d'hypermédia à base de connaissances du domaine est comparable aux autres approches de la conception d’applications hypermédias à base de connaissances du domaine en ce que le concepteur est guidé dans sa tâche d’organisation d’un réseau hypermédia par la structure de la base de connaissances qui modélise le domaine et plus précisément par le typage des noeuds, ancres et liens hypermédias. De ce point de vue, l’originalité de notre approche réside dans le choix des graphes conceptuels pour représenter le modèle du domaine. En nous appuyant sur ce formalisme, nous avons défini un mécanisme de typage des éléments du réseau hypermédia qui reflète l'organisation de la base de connaissances à acquérir. Ce typage assure une organisation du réseau hypermédia sémantiquement plus riche et plus structurée que celle basée sur une hiérarchie de types ou un réseau sémantique. Le second aspect de notre approche est la définition d'un schéma de modèle à partir duquel définir la base canonique destinée à générer un modèle du domaine. A la différence des schémas proposés dans les autres approches, ce schéma est dépendant des caractéristiques communes aux domaines auxquels nous nous intéressons, caractérisés par une organisation taxinomique et des descriptions de taxons structurées. Nous pensons que ce que perd notre approche en généricité, le concepteur de l'hypermédia le gagne en terme d'aide à l'explicitation d'un modèle du domaine. A partir d'une base canonique construite sur ce schéma, nous mettons en oeuvre une méthode d'acquisition de connaissances. C'est le troisième aspect de notre approche : l'organisation d'un réseau hypermédia est confondue avec l'explicitation d'un modèle du domaine, guidée par la méthode d'acquisition mise en oeuvre. A l'inverse des autres approches où le concepteur pense en termes d'hypermédias et explicite un modèle hypermédia du domaine, la conception d'hypermédia est pensée dans notre approche en termes de base de connaissances. Les différentes étapes de la méthode d'acquisition sont traduites en termes de création de nouveaux noeuds, ancres et liens d'un certain type dans le réseau hypermédia. Du point de vue central qui est le notre de la conception d'hypermédias, cette approche assure au concepteur une charge cognitive réduite, la tâche d'organisation du réseau hypermédia étant partiellement automatisée. Du point de vue de l'utilisateur, l'intégration de connaissances du domaine dans un hypermédia offre un réseau hypermédia sémantiquement riche et permet de mettre en place des procédures d'accès dynamique à l'information basées sur la manipulation formelle des connaissances.

Cette méthode de construction d'hypermédias à base de connaissances que nous venons de présenter est mise en oeuvre dans le système SATELIT. Nous décrivons dans le chapitre suivant l'implantation que nous avons réalisée et présentons un exemple d'hypermédia à base de connaissances construit avec SATELIT, pour l'orchidologie.

Chapitre 5

Mise en oeuvre : SATELIT

Introduction

Nous avons présenté dans le chapitre précédent les aspects méthodologiques et algorithmiques du système SATELIT que nous avons développé pour la conception d'applications hypermédias à base de connaissances représentées dans le formalisme des graphes conceptuels. Nous précisons dans la première partie de ce chapitre les usages et les utilisateurs auxquels SATELIT est destiné (§ 5.1.). Nous décrivons ensuite l'implantation de ce système (§ 5.2.). Enfin, nous présentons la construction d'un hypermédia à base de connaissances avec SATELIT en prenant l'exemple d'une application à l'orchidologie (§ 5.3.).

5.1. Utilisations et utilisateurs de SATELIT

5.1.1. Pour quoi SATELIT ?

SATELIT est un mot valise pour Système d'Acquisition de Taxinomies par ELIcitation de Traits discriminants. Il désigne un système de conception d'hypermédias à base de connaissances du domaine et s'adresse aux domaines dotés d'une organisation taxinomique des concepts, notamment les domaines des sciences naturelles. Il répond donc à l'objectif que nous nous étions fixé dans l'introduction de cette thèse, celui d'offrir un outil de mémorisation et d'organisation de l'information pour les domaines des sciences naturelles. En effet, il permet de construire un réseau hypermédia dont les éléments sont indexés à une base de connaissances représentées dans le formalisme des graphes conceptuels ; comme nous l'avons développé dans le chapitre 1 de cette thèse, il est de ce fait capable de traiter des descriptions structurées des concepts du domaine. L'implantation actuelle de SATELIT est basée sur le support S du modèle des graphes conceptuels simples ; SATELIT ne permet pas encore de traiter des taxinomies de concepts comportant des exceptions. L'extension du système est en cours pour permettre de décrire les concepts du domaine sur un support $S_{\delta\varepsilon}$ (cf. chapitre 2).

La méthode de conception qu'implémente SATELIT repose sur la méthode d'acquisition des descriptions structurées des concepts du domaine que nous avons présentée dans le chapitre précédent. Le système permet la construction automatique d'un réseau hypermédia, guidée par l'explicitation des descriptions des concepts du domaine. L'utilisateur de l'hypermédia ainsi construit dispose en outre, grâce aux connaissances qu'intègre l'hypermédia, de moyens de navigation conceptuelle sur le réseau d'information et d'un accès dynamique à l'information basé sur la manipulation des descriptions formelles des concepts du domaine.

5.1.2. Pour qui SATELIT ?

Il convient de distinguer deux types d'utilisateurs : l'utilisateur proprement dit de SATELIT, c'est-à-dire le concepteur d'un hypermédia à base de connaissances, et l'utilisateur d'un hypermédia à base de connaissances construit avec SATELIT qui accède à l'information mémorisée grâce aux possibilités de navigation et de recherche d'information dont l'hypermédia est muni.

La construction d'un hypermédia à base de connaissances du domaine requiert d'une part des compétences à la fois en ingénierie des connaissances et en conception d'hypermédias et, d'autre part, les compétences d'un expert du domaine. Dans SATELIT, les compétences en ingénierie des connaissances et en conception d'hypermédias sont en quelque sorte "capturées" dans la méthode d'acquisition des connaissances et de construction d'un réseau hypermédia indexé sur ces connaissances. De ce fait, SATELIT permet à un expert du domaine de construire seul le réseau hypermédia à base de connaissances - notre approche n'excluant cependant pas un enrichissement ultérieur du réseau hypermédia ainsi construit (cf. § 4.4.3.2.). SATELIT est donc typiquement destiné aux systématiciens experts d'un domaine à modéliser. Il permet d'acquérir à partir de leur expertise une base de connaissances du domaine et de construire un réseau hypermédia indexé sur cette base avec les documents disponibles sur le domaine (classifications, descriptions, catalogues, planches, photos, etc.).

Un hypermédia à base de connaissances construit avec SATELIT est destiné à différents types d'utilisateurs. Un spécialiste du domaine consultera l'hypermédia construit au même titre qu'un ouvrage spécialisé sur le domaine pour rechercher une information précise. De plus, les connaissances mémorisées peuvent être manipulées pour raffiner l'expertise. Ce même hypermédia peut également être utile à un novice du domaine qui découvrira les concepts du domaine en naviguant sur le réseau. Des travaux sont en cours pour intégrer à SATELIT un système d'aide à la navigation basé sur la définition de profils d'utilisateur et l'apprentissage de leurs comportements sur le réseau d'information [Akoulchina et Ganascia, 97].

5.2. Implantation de SATELIT

Le système SATELIT se présente sous la forme d'une pile HyperCard ; il est écrit dans le langage HyperTalk. Notre choix d'HyperCard pour le développement de SATELIT a été motivé par deux considérations principales. Tout d'abord, HyperCard concilie deux styles de programmation : la programmation orientée objets et la programmation hypertexte. Il est de ce fait tout naturellement adapté pour la conception de réseaux hypermédias enrichis d'un typage de leurs éléments, tels que nous souhaitons en construire avec SATELIT. La seconde motivation de notre choix d'HyperCard concerne les utilisateurs auxquels est destiné SATELIT. Certains noeuds d'information du réseau hypermédia construit avec SATELIT attendent d'être enrichis ultérieurement, manuellement, par des documents textuels et graphiques scannés (cf. § 4.4.3.). Pour cela, HyperCard présente l'intérêt de disposer de différents niveaux d'utilisateurs et d'être ainsi accessible à des utilisateurs non informaticiens pour accomplir cette tâche. De plus, des utilisateurs non informaticiens disposent le plus souvent de micro-ordinateurs or HyperCard est précisément disponible sur Macintosh.

Une étape préliminaire de notre travail de programmation a été l'implantation dans le langage Macintosh Common Lisp (MCL) de la méthode d'acquisition des connaissances sur laquelle est fondée la méthode de conception d'hypermédias à base de connaissances de SATELIT. Nous avons ensuite implanté une première version de SATELIT en faisant appel aux fonctions MCL que nous avions écrites au travers de fonctions externes à

HyperCard pour implanter les procédures de manipulation des connaissances. Cependant, l'acquisition de données MCL interfacée avec HyperCard s'est avérée trop coûteuse en temps. En effet, l'incrémentalité du processus d'acquisition nécessite une mise à jour continuelle des données et donc des passages trop répétés d'HyperCard à MCL. Finalement, dans la version actuelle de SATELIT que nous présentons ici, les connaissances sous-jacentes au réseau hypermédia sont mémorisées sur des objets HyperCard : des champs ; elles sont manipulées par des scripts HyperTalk.

5.2.1. Architecture de SATELIT

Le système SATELIT est une *pile* HyperCard, c'est-à-dire un ensemble de *cartes*, supportant des *champs* et des *boutons*. Ces objets sont la traduction directe des éléments d'un réseau hypertexte : les noeuds d'information sont représentés par des cartes ; les informations textuelles sont contenues dans des champs (sous la forme de chaînes de caractères) ; les ancres des liens entre les noeuds d'information sont représentées par des boutons et des lignes de champs. Ce sont les zones activables des cartes : un clic de souris sur ces zones déclenchera l'activation d'un *script* (HyperTalk) commandant le remplacement de la carte courante par celle liée. La pile SATELIT est donc un ensemble initial de cartes, munies de boutons et de champs, qui représente le réseau hypermédia initial que nous avons défini au paragraphe 4.4.2.1.

Le typage que nous avons établi des éléments du réseau hypermédia (cf. § 4.4.1.) se traduit pour partie avec HyperCard par des *fonds* associés aux cartes : deux noeuds de même type sont représentés dans SATELIT par deux cartes de même fond. Cela ne permet cependant pas de typer les liens et les ancres, ni de représenter des hiérarchie de types. Les fonds des cartes de la pile SATELIT représentent donc les types de noeuds les plus généraux. Les hiérarchies de types de noeuds, d'ancres et de liens se déduisent de la structure de la base de connaissances aux éléments de laquelle sont indexés les éléments du réseau hypermédia ; les noms que nous donnons aux objets HyperCard que nous créons permettent de retrouver les types de concepts ou de relations de la base auxquels ils correspondent.

La pile SATELIT comporte des scripts HyperTalk attachés à ses différents objets HyperCard. Ces scripts implantent la méthode de construction d'un réseau hypermédia décrite dans le chapitre précédent (cf. § 4.4.2.). Leur activation se traduit d'une part par la création automatique de nouvelles cartes supportant de nouveaux champs et boutons et, d'autre part, par la mise en oeuvre de la méthode d'acquisition de connaissances qui guide le processus de création des éléments du réseau hypermédia (cf. § chapitre 3). Rappelons en effet que le système interroge l'expert du domaine pour construire incrémentalement une base de connaissances en même temps qu'un réseau hypermédia. Les connaissances acquises sont mémorisées dans des champs cachés des cartes créées. Ces champs de cartes de la pile SATELIT représentent le schéma de modèle conceptuel du domaine que nous avons défini au paragraphe 4.3.3.1.

Le système SATELIT est donc une "pile noyau", génératrice d'une pile particulière pour un domaine considéré. La figure 5.1. ci-dessous présente l'architecture générale de la pile SATELIT. On distingue en gris les cartes dont les fonds sont destinés à être partagés par plusieurs cartes dans une pile développée pour un domaine particulier. Les champs et les boutons des cartes et fonds de cartes seront précisés dans la partie suivante.

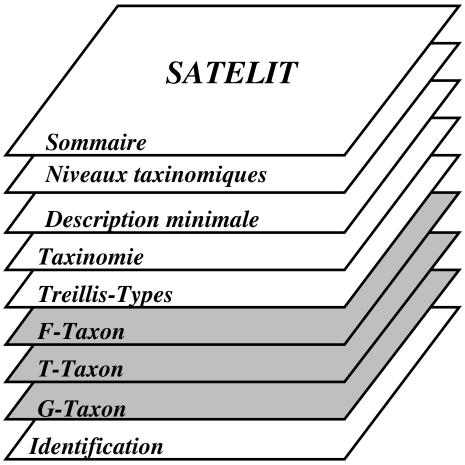


figure 5.1. Architecture générale de SATELIT

Pour construire un hypermédia à base de connaissances pour un domaine particulier, la pile SATELIT est d'abord copiée puis sa copie est développée par l'activation des scripts associés aux objets de la pile qui commandent la création de nouveaux objets. Dans la présentation de la pile SATELIT, nous distinguons donc deux sortes d'objets :

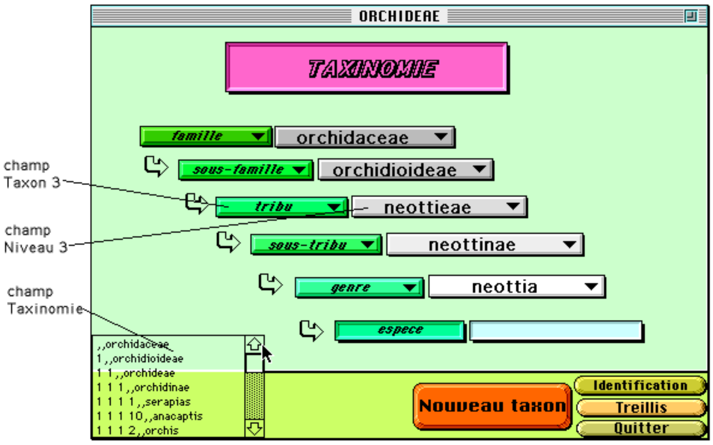
- des objets qui correspondent au schéma de modèle du domaine et constituent un point de départ pour créer de nouveaux objets (par copie) dans une pile particulière, pour un domaine particulier ;

- des objets qui correspondent aux différentes étapes du processus de construction d'un réseau hypermédia à base de connaissances. L'activation successive des scripts attachés à ces objets déroule le processus d'acquisition de connaissances auprès de l'expert du domaine et la création automatique des nouveaux objets dans la pile.

5.2.2. Les objets représentant le schéma du réseau hypermédia

5.2.2.1. La carte Taxinomie, ses champs et ses boutons

La carte *Taxinomie* est destinée à visualiser l'ensemble des taxons acquis et à accéder aux informations associées à l'un d'eux. Pour cela, elle supporte des champs auxquels sont associés des scripts. Ces champs seront remplis au moment de la construction d'un hypermédia pour un domaine particulier. La figure 5.2. ci-dessous présente la carte *Taxinomie* de la pile ORCHIDACEAE construite pour l'orchidologie.



famille: orchidaceae
sous-famille: orchidioideae
tribu: neottieae
sous-tribu: neottinae
genre: neottia
espece:

```
,,orchidaceae
1,,orchidioideae
1 1,,orchideae
1 1 1,,orchidinae
1 1 1 1,,serapias
1 1 1 10,,anacaptis
1 1 1 2,,orchis
```

figure 5.2. La carte *Taxinomie* de la pile ORCHIDACEAE.

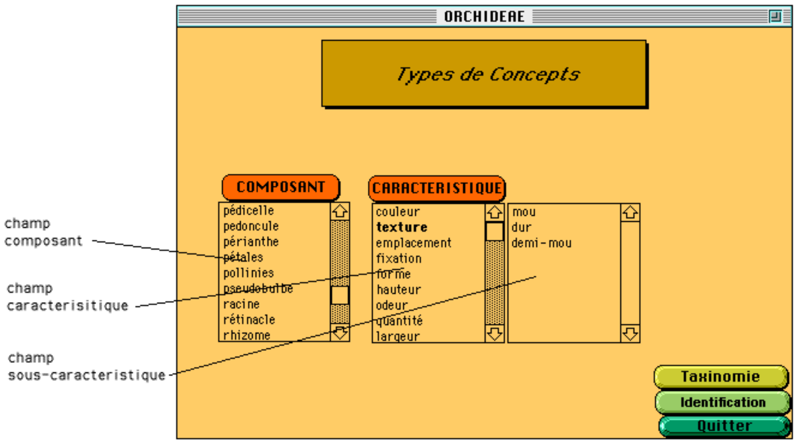
Les champs *Niveau i* contiennent les noms des différents niveaux taxinomiques et les champs *Taxon i* les noms des taxons de chaque niveau taxinomique. Le champ *Taxinomie* est normalement caché ; il contient un index des différents taxons tel que l'organisation hiérarchique des taxons puisse être reconstituée. Des scripts associés aux champs *Niveau i* permettent la visualisation de l'ensemble de la taxinomie. Lorsqu'un tel script est activé (par un clic de souris sur le champ), la liste de tous les taxons acquis de ce niveau taxinomique et descendant du taxon courant du niveau supérieur apparaît. Le choix de l'un d'eux entraîne la mise à jour des champs *Taxon j*, pour j≥i. Le calcul de la liste des frères d'un taxon d'un niveau taxinomique quelconque et la mise à jour des champs sont basés sur la manipulation du contenu du champ caché *Taxinomie*.

Les champs *Taxon i* contenant les noms des taxons sont les ancres des liens de type N_TAXON_TAXON qui permettent d'accéder aux informations associées aux taxons, contenues sur les cartes de la pile de fonds *F_Taxon*, *T_Taxon* et *G_Taxon*.

Le bouton *Treillis* permet d'accéder à la carte *Treillis_types* destinée à supporter une représentation de l'ensemble des types de concepts explicités lors de l'acquisition des taxons d'un domaine particulier.

5.2.2.2. La carte Treillis_Types, ses champs et ses boutons

La carte *Treillis_Types* est destinée à visualiser l'ensemble des types de concepts atomiques qui sont utilisés dans les descriptions des taxons et à accéder aux informations associées à l'un d'eux. La figure 5.3. ci-dessous présente la carte *Treillis_Types* de la pile ORCHIDACEAE.



pédicelle
pedoncule
périanthe
pétales
pollinies
pseudobulbe
racine
rétinacle
rhizome

couleur
texture
emplacement
fixation
forme
hauteur
odeur
quantité
largeur

mou
dur
demi-mou

figure 5.3. La carte *Treillis_Type* de la pile ORCHIDACEAE.

Les champs à menu déroulant, *caractéristiques* et *composants* permettent de visualiser les types de concepts inférieurs aux types primitifs CARACTERISTIQUE et COMPOSANT de la base canonique explicités au cours de la construction de l'hypermédia (cf. § 5.3.). Le champ *sous-caracteristiques* contient les types de concepts inférieurs au type sélectionné sur une ligne du champ *caracteristiques*. Le script de mise à jour de ce champ consiste en la manipulation d'un champ caché *Treillis* qui contient un index des différents types de concepts tel que l'organisation hiérarchique de ces types puisse être reconstituée.

Les lignes des champs *composants*, *caractéristiques* et *sous-caractéristiques* sont les ancres des liens de type N_PRIMITIVE_PRIMITIVE qui permettent d'accéder aux informations associées aux types de concepts, contenues sur les cartes de la pile de types *T_Primitive* et *G_Primitive*.

Le bouton *Taxinomie* permet d'accéder à la carte *Taxinomie* décrite ci-dessus.

5.2.2.3. Le fond de carte F_Taxon, ses champs et ses boutons

Le fond de carte *F_Taxon* est destiné à la représentation et la visualisation des descriptions formelles des taxons ; il est partagé par l'ensemble des cartes représentant des noeuds de type F_TAXON du réseau hypermédia construit. La figure 5.4. ci-dessous présente les champs et boutons du fond de carte *F_Taxon* défini dans la pile SATELIT.

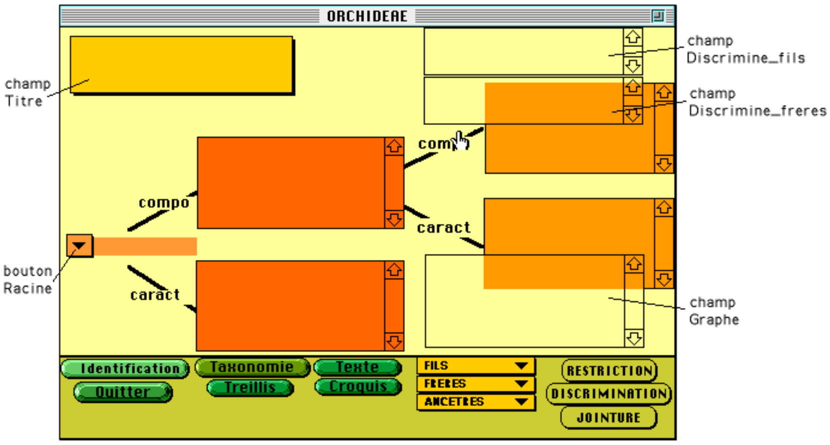


figure 5.4. Le fond de carte *F_Taxon*

Le champ *Titre* est destiné à contenir le nom du taxon décrit. Le champ *Graphe* est normalement caché ; il est destiné à contenir les arcs conceptuels qui constituent la représentation formelle de la description du taxon. Les champs *Discrimine_frères* et *Discrimine_fils* sont également cachés ; ils sont destinés à contenir la liste des arcs conceptuels de la description du taxon qui permettent de discriminer ce dernier respectivement de ses frères et entre ses fils. Les quatre autres champs de la carte sont destinés à visualiser deux niveaux de composition dans la description du taxon : une partie choisie avec le bouton *Racine* à menu déroulant, ses caractéristiques et ses sous-parties, et les caractéristiques et sous-parties d'une sous-partie sélectionnée. L'exploration de l'ensemble de la description est fondée sur l'activation (par des clicks) des scripts associés au bouton *Racine* et au champ contenant les sous-parties de la partie racine. Ces scripts sont fondés sur la manipulation du contenu du champ *Graphe*.

Le fond de carte *F_Taxon* implante également les primitives de navigation que nous avons définies dans le chapitre précédent (cf. figure 4.4.) :

- les liens de type N_CONCEPT_PRIMITIVE sont représentés par des scripts associés aux champs contenant les caractéristiques et composants du taxon décrit ; l'activation de ces scripts remplace la carte courante par la carte (si elle existe) de fond *T_Primitive* contenant les informations textuelles relatives à la primitive de langage sélectionnée ou par celle de fond *G_Primitive* contenant une description graphique ;

- les liens de type IN_TAXON sont représentés par les scripts associés aux boutons *Texte* et *Croquis* ; l'activation de ces scripts permet d'accéder aux cartes de fonds *T_Taxon* et *G_Taxon* qui supportent les descriptions textuelle et graphique du taxon ;

- les liens de types respectifs T_FRERE_FRERES, T_PERE_FILS et T_DESCENDANT_ANCETRES sont représentés par les scripts associés aux boutons *Frères*, *Fils* et *Ancêtres* à menu déroulant ; l'activation de ces scripts permet d'accéder aux cartes de fond *F_Taxon* supportant la description sous forme de graphe du taxon sélectionné ;

- le lien de type N_TAXON_TAXON est représenté par le script associé au bouton *Taxinomie* ; l'activation de ce script permet d'accéder à la carte *Taxinomie*.

5.2.2.4. Les fonds de carte T_Taxon, G_Taxon, T_Primitive et G_Primitive

Les fonds de carte *T_Taxon*, *G_Taxon*, *T_Primitive* et *G_Primitive* sont destinés à la visualisation des descriptions textuelles et graphiques des taxons et des primitives de langage qui seront mémorisées pour un domaine particulier. Ils sont partagés par l'ensemble des cartes représentant les noeuds hypermédias de types, respectivement, T_TAXON, G_TAXON, T_PRIMITIVE et G_PRIMITIVE.

Les fonds *T_Taxon* et *T_Primitive* comportent un champ à menu déroulant destiné à contenir des informations textuelles associées à un taxon ou à une primitive de langage.

Les fonds *T_Taxon* et *G_Taxon* comportent des boutons à menu déroulant similaires à ceux du fond *F_Taxon* ; ils représentent les ancres des liens de type FRERE_FRERES, PERE_FILS et DESCENDANT_ANCETRES et des liens de type IN_TAXON.

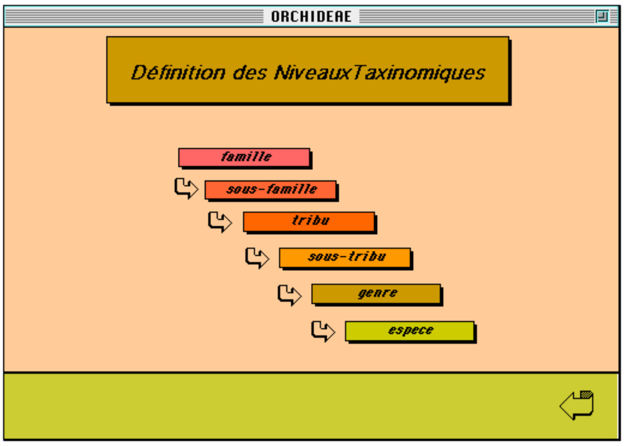
Les fonds *T_Primitive* et *G_Primitive* comportent des boutons qui représentent les ancres des liens de type IN_PRIMITIVE.

5.2.3. Les objets déclenchant le processus de construction du HBC

5.2.3.1. Les cartes Niveaux_Taxinomiques et Description_Minimale

Les cartes *Niveaux_Taxinomiques* et *Description_Minimale*, leurs champs et leurs boutons concernent la phase d'initialisation du processus d'acquisition des connaissances qui guide ensuite la construction du réseau hypermédia.

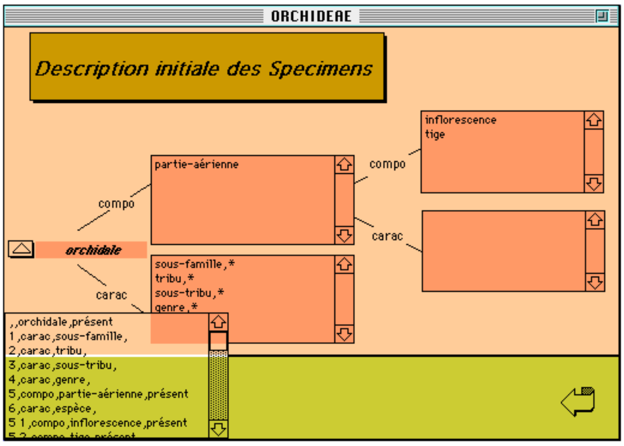
La carte *Niveaux_Taxinomiques* d'une pile SATELIT est dédiée à l'acquisition des niveaux taxinomiques à considérer pour un domaine particulier. La figure 5.5. ci-après présente la carte *Niveaux_Taxinomiques* de la pile ORCHIDACEAE. L'expert du domaine saisit les différents niveaux dans des champs disposés les uns au-dessous des autres. Dans le même temps, un champ caché est rempli automatiquement pour contenir la liste de ces niveaux ; ceux-ci sont utilisés comme paramètres de différents scripts de la pile.



famille
sous-famille
tribu
sous-tribu
genre
espece

figure 5.5. La carte *Niveaux Taxinomiques* de la pile ORCHIDACEAE.

La carte *Description_Minimale* est dédiée à l'acquisition d'une description minimale des objets du domaine considéré, c'est-à-dire à la construction du graphe canonique à partir duquel seront dérivées toutes les descriptions de taxons (cf. § 1.4.2.1.). La figure 5.6. ci-dessous présente la carte *Description_Minimale* pour la pile ORCHIDACEAE. Cette carte supporte des champs similaires à ceux du fond de carte *F_Taxon* (cf. figure 5.4.) destinés à la représentation (dans un champ caché) et à la visualisation de la description construite, au fur et à mesure de sa construction. Les boutons *compo* et *caract* permettent d'activer des scripts pour l'acquisition d'une nouvelle partie ou d'une nouvelle caractéristique (cf. § 5.3.1.).



partie-aérienne

sous-famille,*
tribu,*
sous-tribu,*
genre,*

inflorescence
tige

```
,,orchidale,présent
1,carac,sous-famille,
2,carac,tribu,
3,carac,sous-tribu,
4,carac,genre,
5,compo,partie-aérienne,présent
6,carac,espèce,
5 1,compo,inflorescence,présent
5 2,compo,tige,présent
```

figure 5.6. La carte Description_Minimale de la pile ORCHIDACEAE.

5.2.3.2. Les boutons Nouveau des cartes Taxinomie et Treillis_Types

Le bouton *Nouveau* de la carte *Taxinomie* permet d'activer le script qui lui est associé qui déclenche la création automatique de nouvelles cartes dans la pile SATELIT. Ces cartes sont créées à partir des fonds de cartes *F_Taxon*, *T_Taxon* et *G_Taxon* (cf. § 5.2.2.) ; les champs de la carte de fond *F_Taxon* sont initialisés avec le contenu des champs de la carte associée au taxon père. Le script du bouton *Nouveau* de la carte *Taxinomie* assure également la mise à jour de différents champs cachés des cartes de la pile : le champ *Taxinomie* de la carte *Taxinomie*, les champs associés aux boutons *Freres* et *Fils* des cartes relatives aux taxons frères et au taxon père du nouveau taxon, etc.

Le bouton *Nouveau* de la carte *Treillis_Types* permet l'activation du script qui lui est associé qui déclenche la création automatique de nouvelles cartes dans la pile à partir des fonds de cartes *T_Primitive* et *G_Primitive*. Ces cartes sont destinées à la mémorisation d'informations associées à la primitive de langage sélectionnée sur la carte *Treillis_Types*.

5.2.3.3. Les boutons Restriction, Discrimination et Jointure du fond de carte F_Taxon

La méthode d'acquisition incrémentale des descriptions de taxons (cf. chapitre 3) est implantée dans les scripts associés aux boutons *Restriction*, *Discrimination* et *Jointure* du fond de carte *F_Taxon*. L'activation du script associé au bouton *Restriction* déclenche un processus d'interrogation de l'expert du domaine afin de spécialiser les arcs conceptuels génériques de la description du taxon décrit sur la carte ; les champs de la carte sont mis à jour automatiquement au fur et à mesure ainsi que le champ caché *Treillis* de la carte *Treillis_Type* qui contient les types de concepts utilisés dans les descriptions de taxon.

L'activation du script associé au bouton *Discrimination* déclenche la procédure de discrimination du taxon décrit sur la carte. Ce script gère la mise à jour de la liste des traits discriminants des taxon frères qui est mémorisée dans un champ caché de la carte de fond *F_Taxon* associée au taxon père. Ce script contient le déclenchement du script associé au bouton *Jointure* dans le cas où les descriptions des taxons frères doivent être précisées pour être discriminées.

L'activation du script associé au bouton *Jointure* déclenche un processus d'interrogation de l'expert afin d'introduire de nouveaux traits descriptifs. Ce script gère la mise à jour des champs des cartes de fond *F_Taxon* relatives aux taxons frères et du champ caché *Treillis* de la carte *Treillis_Type*.

Dans la partie suivante, nous décrivons une session d'utilisation de SATELIT. Elle est vue comme l'activation successive des scripts associés aux différents objets HyperCard que nous venons de présenter.

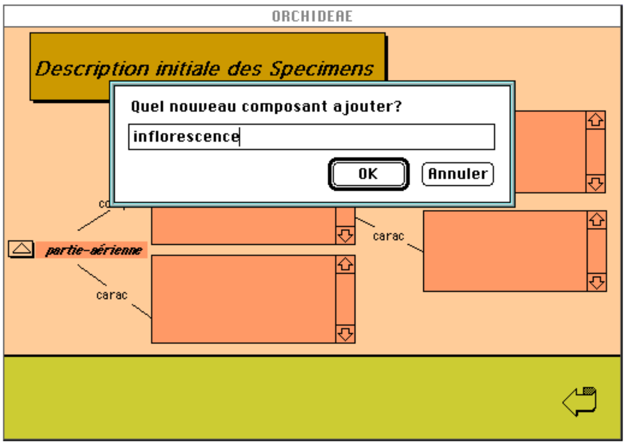
5.3. Déroulement d'une session d'utilisation de SATELIT

L'utilisation de la pile SATELIT pour un domaine particulier commence par sa copie et le renomage de la pile copiée. Nous présentons dans cette partie la construction et la consultation d'une pile ORCHIDACEAE à partir de la copie de la pile SATELIT. A l'ouverture de la pile, un sommaire succède à la carte de présentation. Trois menus sont disponibles : initialisation, acquisition de taxons et consultation de la taxinomie. Nous décrivons dans cette partie l'initialisation de la pile ORCHIDACEAE et la construction d'un réseau hypermédia fondé sur l'acquisition des descriptions des taxons qui apparaissent dans l'extrait de la taxinomie des orchidées que nous avons présenté dans le chapitre 1 de cette thèse ; nous décrivons ensuite les possibilités d'accès aux informations ainsi mémorisées.

5.3.1. Phase d'initialisation

En activant le bouton *Initialisation* de la carte *Sommaire*, on accède à la carte *Niveaux Taxinomiques* destinée à la saisie des noms des différents niveaux taxinomiques définis dans le domaine considéré. Pour la pile ORCHIDACEAE, on indique donc les niveaux sous-famille, tribu, sous-tribu, genre et espèce (cf. figure 5.5.).

En activant ensuite le bouton de validation, on accède à la carte *Description Minimale* destinée à l'acquisition d'une description commune à l'ensemble des objets du domaine (cf. § 5.2.3.1.). Comme le montre la figure 5.7. ci-dessous, un clic sur le bouton *compo* ou *caract* déclenche une procédure d'ajout d'un arc conceptuel représentant un trait de composition ou de caractérisation de la partie de la description visualisée comme la racine courante. Les champs de visualisation et de mémorisation de cette description sont mis à jour au fur et à mesure de l'acquisition de nouveaux traits descriptifs, ainsi que ceux de la carte *Treillis_Type* au fur et à mesure de l'introduction de nouveaux composants ou de nouvelles caractéristiques.



Quel nouveau composant ajouter?

inflorescence

figure 5.7. Ajout de l’arc [partie-aérienne]→(compo)→[inflorescence]

Une fois construite la description initiale des objets, l’activation du bouton de validation déclenche :

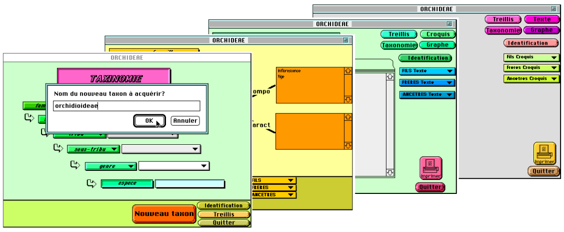
- l’initialisation du champ *Taxinomie* de la carte *Taxinomie* avec “orchidaceae”- le nom de la famille dont on souhaite acquérir les taxons inférieurs - ;
- la création d’une carte de fond *F_Taxon* et de nom *F_orchidaceae* ;
- initialisation du champ *Taxon* de cette carte avec la description initiale des objets du domaine par copie du contenu du champ caché de la carte *Description Minimale* ;
- la création des cartes *T_orchidaceae* et *G_orchidaceae* de fonds respectifs *T_Taxon* et *G_Taxon*.

Après cette phase initiale d'acquisition d'une base canonique, le processus d'acquisition des descriptions des taxons peut débuter qui guide la création du réseau hypermédia à base de connaissances.

5.3.2. Construction du réseau hypermédia à base de connaissances

En activant le bouton *Acquisition de taxons* de la carte *Sommaire* ou bien le bouton de validation de la carte *Description Minimale*, on accède à la carte *Taxinomie* à partir de laquelle créer de nouvelles cartes pour acquérir les descriptions de nouveaux taxons, grâce au bouton *Nouveau*. En mode d’acquisition, le bouton *Identification* est caché.

En sélectionnant sur la carte *Taxinomie* la famille des *Orchidaceae* (qui devient en gras), puis en activant le bouton *Nouveau*, l’expert déclenche le développement de la pile ORCHIDACEAE par l’acquisition de la description d’une première sous-famille et la mémorisation d’informations relatives à celles-ci. Prenons l’exemple de la sous-famille des *Orchidioideae* (cf. figure 5.8.).



Nom du nouveau taxon à acquérir?

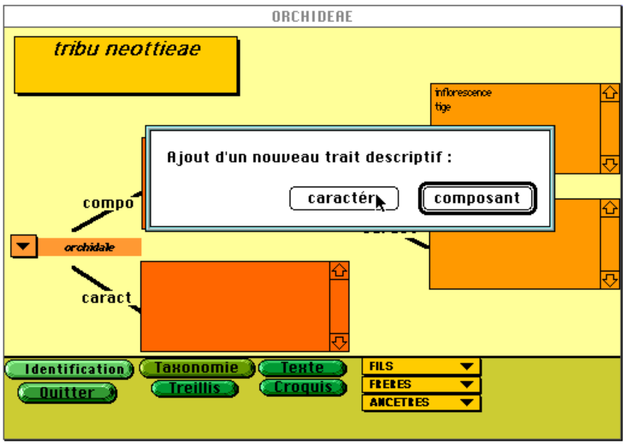
orchidioideae

figure 5.8. Activation du bouton *Nouveau* de la carte *Taxinomie* de la pile ORCHIDACEAE.

Les cartes *F_orchidioideae*, *T_orchidioideae* et *G_orchidioideae* sont construites automatiquement et le champ *Taxon* de la carte *F_orchidioideae* est initialisé par copie du champ *Taxon* de la carte *F_orchidaceae* (relative au taxon père). La carte *F_orchidioideae* devient alors la carte courante et le bouton *Restriction* est automatiquement activé. Aucun trait descriptif contenu dans le champ taxon n'est générique de sorte qu'aucune restriction de concept n'est opérée. La dernière instruction du script associé au bouton *Restriction* est l'activation du bouton *Discrimination*. Aucun taxon frère n'est encore acquis de sorte qu'aucun test de discrimination n'est déclenché. A l'issue de ce processus d'acquisition, les cartes *T_orchidioideae* et *G_orchidioideae* nouvellement créées sont prêtes à recevoir les informations textuelles et graphiques relatives aux *Orchidioideae*.

En retournant à la carte *Taxinomie* puis en sélectionnant la sous-famille des *Orchidioideae* et en indiquant la tribu des *Vandeae* après avoir activé le bouton *Nouveau*, on déclenche de la même manière la création des cartes *F_vandeae*, *T_vandeae* et *G_vandeae* et l'initialisation de la description de cette tribu dans le champ *Taxon* de la carte *F_vandeae* par copie du contenu du champ *Taxon* de la carte *F_orchidioideae*. De même qu'après l'acquisition de la description des *Orchidioideae*, aucune procédure de restriction ni de discrimination n'est déclenchée.

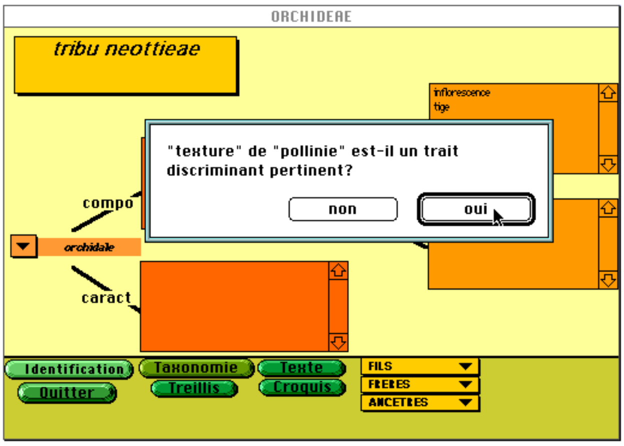
En sélectionnant de nouveau la sous-famille des *Orchidioideae* sur la carte *Taxinomie* et en indiquant la tribu des *Neottieae*, on déclenche la création des cartes *F_neottieae*, *T_neottieae* et *G_neottieae* et l'initialisation de la description de cette tribu dans le champ *Taxon* de la carte *F_neottieae* par le contenu du champ *Taxon* de la carte *F_orchidioideae*. L'activation automatique du bouton *Restriction* de la carte *F_neottieae* provoque l'activation du bouton *Discrimination* qui déclenche cette fois une procédure de discrimination des descriptions des tribus *Vandeae* et *Neottieae*. Ces descriptions étant identiques, le bouton *Jointure* est automatiquement activé ; il déclenche une procédure d'interrogation de l'expert pour enrichir la description de ces taxons (cf. figure 5.9.).



Ajout d'un nouveau trait descriptif :

figure 5.9. Activation du bouton *Jointure* de la carte *F_neottieae*

L’expert indique la texture des pollinies, dure chez les *Vandeae* et molle chez les *Neottieae* (cf. § 1.3.2.). L’arc [pollinie]→(de_texture)→[texture] est alors ajouté dans le champ *Taxon* de la carte *F_orchidioideae* ; il est spécialisé en l’arc [pollinie]→(de_texture)→[dur] dans le champ *Taxon* de la carte *F_vandeae* et en l’arc [pollinie]→(de_texture)→[mou] dans le champ *Taxon* de la carte *F_neottieae*. Le script associé au bouton *Discrimination* de la carte *F_neottieae* est alors automatiquement à nouveau déclenché et l’arc [pollinie]→(de_texture)→[texture] est retenu comme trait discriminant (cf. figure 5.10.).

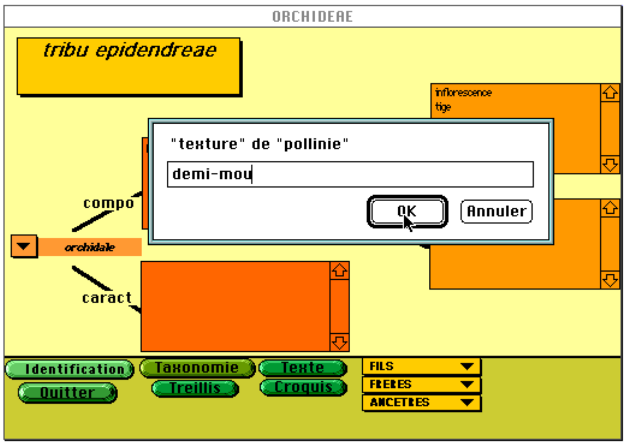


"texture" de "pollinie" est-il un trait discriminant pertinent?

figure 5.10. Activation du bouton *Discrimination* de la carte *F_neottieae*

L’arc [pollinie]→(de_texture)→[texture] est introduit dans le champ *Discrimination_fils* de la carte *F_orchidioideae*. Les arcs [pollinie]→(de_texture)→[dur] et [pollinie]→(de_texture)→[mou] sont introduits dans les champs *Discrimination_frères* des cartes *F_vandeae* et *F_neottieae*, respectivement.

En sélectionnant de nouveau la sous-famille des *Orchidioideae* sur la carte *Taxinomie* et en indiquant la tribu des *Epidendreae* après avoir activé le bouton *Nouveau*, on déclenche la création des cartes *F_epidendreae*, *T_epidendreae* et *G_epidendreae* et l'initialisation de la description de cette tribu dans le champ *Taxon* de la carte *F_epidendreae* par le (nouveau) contenu de la carte *F_orchidioideae*. L'activation automatique du bouton *Restriction* déclenche cette fois une procédure d'interrogation de l'expert, pour spécialiser l'arc [pollinie]→(de_texture)→[texture] hérité de la description du taxon père (cf. figure 5.11.).



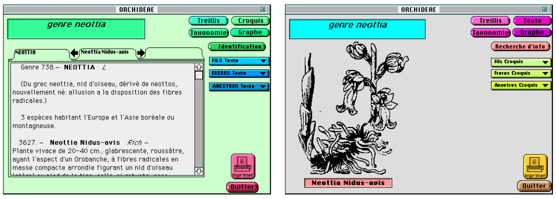
"texture" de "pollinie"

demi-mou

figure 5.11. Activation du bouton *Restriction* de la carte *F_epidendreae*

Conformément à la réponse de l'expert, le type de concept "texture" est remplacé par "demi_mou". Le bouton *Discrimination* est alors activé ; le champ *Discrimination_fils* de la carte *F_orchidioideae* relative au père des taxons à discriminer n'étant plus vide, la procédure débute par un test de discrimination à l'aide du trait discriminant [pollinie]→(de_texture)→[texture] entre la description des *Epidendreae* qui vient d'être acquise et celles des tribus soeurs - les *Vandeae* et les *Neottieae*. Ce test étant positif, le script associé au bouton *Jointure* n'est pas déclenché et le trait [pollinie]→(de_texture)→[demi_mou] est introduit dans le champ *Discrimination_frères* de la carte *F_epidendreae* après confirmation de sa pertinence par l'expert.

En retournant à la carte *Taxinomie*, on peut de nouveau acquérir une tribu de la sous-famille des *Orchidioideae*, ou bien une autre sous-famille de la famille des *Orchidaceae*, ou encore des sous-tribus des tribus acquises puis des genres et des espèces. La même procédure est chaque fois déclenchée qui entraîne la création automatique de nouvelles cartes de fonds *F_Taxon* ,*T_Taxon* et *G_Taxon*. A tout moment, chaque fois qu'une procédure d'acquisition de la description d'un taxon est achevée, on peut accéder aux cartes de fonds *T_Taxon* et *G_Taxon* qui ont été créées, destinées à supporter des informations textuelles ou graphiques relatives aux taxons. On mémorise des informations sur ces cartes sous la forme de textes introduits dans des champs ou d'images scannées qui peuvent être enrichies de légendes activables. La figure 5.12. ci-dessous présente la carte *G_neottia* sur laquelle figure une image d'un spécimen du genre *Neottia*.



Genre 738.– **NEOTTIA** *L.*

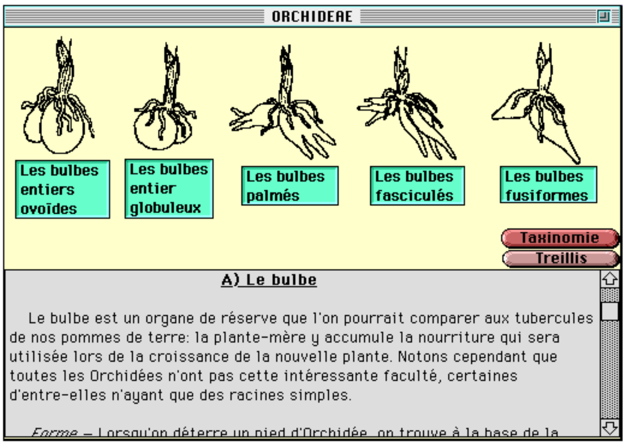
(Du grec neottia, nid d'oiseau, dérivé de neottos, nouvellement né: allusion a la disposition des fibres radicales.)

3 espèces habitant l'Europe et l'Asie boréale ou montagneuse.

3627. – **Neottia Nidus-avis** *Rich* –
Plante vivace de 20-40 cm., glabrescente, roussâtre, ayant l'aspect d'un Orobanche, à fibres radicales en masse compacte arrondie figurant un nid d'oiseau

figure 5.12. Les cartes *G_neottia* et *G_neottia*.

On accède également, à partir de la carte *Taxinomie*, à la carte *Treillis_Types* dont le bouton *Nouveau* permet le déclenchement d'un script de création automatique de cartes destinées à la mémorisation d'informations relatives aux primitives du langage de description des taxons explicitées. De même que pour créer un noeud d'information relative à un taxon, on sélectionne d'abord une primitive (qui devient en gras) puis on active le bouton *Nouveau*. On accède alors aux cartes de fond *T_Primitive* ou *G_Primitive* créées qui sont enrichies par l'introduction de textes dans les champs prédéfinis ou d'images scannées. La figure 5.13. ci-dessous présente la carte de fond *T_Primitive* qui supporte des informations relatives au bulbe des orchidées (textes et croquis ont été regroupés sur une même carte ; il n'y a pas de carte *G_bulbe*).



A) Le bulbe

Le bulbe est un organe de réserve que l'on pourrait comparer aux tubercules de nos pommes de terre: la plante-mère y accumule la nourriture qui sera utilisée lors de la croissance de la nouvelle plante. Notons cependant que toutes les Orchidées n'ont pas cette intéressante faculté, certaines d'entre-elles n'ayant que des racines simples.

Forme – Lorsqu'on déterre un pied d'Orchidée, on trouve à la base de la

figure 5.13. La carte *T_bulbe* de la pile ORCHIDACEAE.

La pile SATELIT permet donc bien une méthode de création semi-automatique d'un réseau hypermédia à base de connaissances telle que nous l'avons définie dans le chapitre précédent. Les cartes, champs et boutons sont créés automatiquement, au fur et à mesure du processus d'acquisition de connaissances auprès de l'expert du domaine ; les liens entre les cartes sont calculés automatiquement en manipulant les connaissances mémorisées dans les champs cachés des cartes créées. En revanche, la mémorisation des informations textuelles et graphiques sur les cartes créées à cet effet n'est pas guidée et la création de liens à partir des mots significatifs dans les textes ou dans les légendes de croquis est manuelle. On pourrait envisager un calcul automatique d'une partie de ces liens en comparant les mots des textes et des légendes avec les noms des cartes de la pile. Cela relève des techniques d'analyses de textes que nous avons présentées au paragraphe 4.2. et qui dépassent le cadre de cette thèse.

Dans la partie suivante, nous présentons les possibilités d'accès à l'information dans une pile créée à partir de SATELIT.

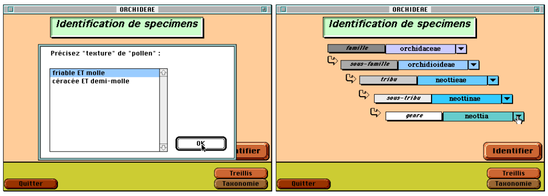
5.3.3. Consultation des informations mémorisées

L'activation du bouton *Consultation* de la carte *Sommaire* rend inactifs les boutons dédiés à la création de nouvelles cartes ou à l'acquisition de nouvelles descriptions : les boutons *Nouveau*, *Restriction*, *Discrimination* et *Jointure*. En activant le bouton *Consultation*, on accède à la carte *Taxinomie* à partir de laquelle les possibilités d'accès aux autres cartes de la pile sont les suivantes :

- choisir un taxon pour accéder à ses descriptions textuelle, graphique ou sous forme de graphe mémorisées sur des cartes de fonds *T_Taxon*, *G_Taxon* et *F_Taxon* ;
- activer le bouton *Treillis* pour accéder à la carte *Treillis_Types* à partir de laquelle choisir une primitive de langage pour accéder aux informations relatives à celle-ci mémorisées sur des cartes de fonds *T_Primitive* et *G_Primitive* ;
- activer le bouton *Identification* pour accéder à la carte *Identification* sur laquelle mettre en oeuvre une procédure d'identification de spécimen.

A partir des cartes relatives aux taxons ou aux primitives du langage de description explicité, les possibilités de navigation sont celles que nous avons décrites dans le chapitre précédent (cf. § 4.4.4.). Les primitives de navigation que nous avons définies sont implantées par les scripts associés aux boutons et à certains champs construits automatiquement lors de la phase de construction du réseau (cf. § 5.2.).

Sur la carte *Identification*, l'activation du bouton *Identifier* déclenche une procédure d'interrogation de l'utilisateur à qui il est demandé de décrire le spécimen qu'il cherche à identifier. Ses réponses permettent une exploration pas à pas du graphe de détermination. La structure de celui-ci est celle de la taxinomie acquise et ses noeuds sont représentés par les champs *Discrimination_fils* des cartes de fond *F_Taxon*. A l'issue de cette procédure, on accède à la carte de fond *F_Taxon* qui supporte la description du taxon déterminé. La figure 5.14. ci-dessous présente la carte *Identification* de la pile ORCHIDACEAE.



Précisez "texture" de "pollen" :

friable ET molle
céracée ET demi-molle

famille : orchidaceae
sous-famille : orchidioideae
tribu : neottieae
sous-tribu : neottinae
genre : neottia

figure 5.14. La carte *Identification* de la pile ORCHIDACEAE.

Cette procédure d'identification a été développée parce que, d'une part, elle est essentielle aux domaines auxquels s'adresse SATELIT et que, d'autre part, elle nous permet de montrer l'opérationnalité des connaissances intégrées dans un réseau hypermédia construit avec SATELIT. Dans le chapitre précédent, nous avons abordé des procédures de recherche plus générales qui peuvent être mises en oeuvre sur un réseau hypermédia construit avec SATELIT en manipulant les connaissances acquises (cf. § 4.4.4.2.). Cela dépasse cependant le cadre de cette thèse dans laquelle nous nous sommes concentrée sur la construction d'un hypermédia à base de connaissances. Des travaux sont en cours dans notre équipe sur l'accès à l'information dans un hypermédia à base de connaissances construit avec SATELIT. Plus précisément, Akoulchina s'intéresse à concevoir et développer un agent d'interface pour SATELIT qui aide l'utilisateur à accéder à l'information recherchée en apprenant des profils d'utilisateurs [Akoulchina et Ganascia, 97].

5.4. Conclusion

Dans ce chapitre, nous avons présenté l'implantation du système SATELIT et son application à la construction d'un hypermédia à base de connaissances en orchidologie. L'intérêt de SATELIT est double. Premièrement, il permet l'organisation automatique d'un réseau hypermédia, au fur et à mesure de l'acquisition auprès d'un expert du domaine de connaissances structurées sur lesquelles sont indexées les éléments du réseau. Deuxièmement, les connaissances qu'intègre ainsi un hypermédia construit avec SATELIT sont d'une part le fondement des primitives de navigation conceptuelle construites automatiquement ; d'autre part, elles sont manipulables et permettent de mettre en oeuvre des procédures de recherche d'information.

La pile ORCHIDACEAE construite avec SATELIT supporte la description de 150 taxons ; l'ensemble de ces descriptions a requis 340 types de concepts. Cette expérimentation en orchidologie témoigne de la faisabilité de notre approche de la conception d'hypermédias à base de connaissances avec SATELIT. Elle montre également les limites d'une implantation de SATELIT sous HyperCard. Au fur et à mesure que le nombre de taxons acquis et la taille des descriptions des taxons augmentent, l'exécution des scripts de manipulation de ces descriptions contenues dans les champs de cartes n'est plus instantanée. En particulier, certaines recherches d'information peuvent requérir plusieurs minutes. Les travaux d'Akoulchina en recherche d'information l'ont amenée à implanter en JAVA une nouvelle version d'un hypermédia à base de connaissances que construirait SATELIT [Akoulchina et Ganascia, 97]. Reste à implanter le module d'acquisition des connaissances et de construction automatique des pages HTML qui constituent les noeuds d'information du réseau hypermédia.

Conclusion

Nos travaux s'inscrivent dans le cadre général de la mémorisation d'information et la construction de bases de connaissances ; ils se situent au croisement des domaines de la conception d'hypermédias et de la représentation et l'acquisition de connaissances. De fait, notre problématique est double. Elle porte d'une part sur la représentation de connaissances taxinomiques, et plus précisément sur la formalisation de descriptions de concepts structurées et organisées hiérarchiquement au sein de taxinomies. Dans ce cadre, nous avons abordé le problème particulier de la représentation des exceptions dans des taxinomies. D'autre part, notre problématique concerne la conception d'hypermédias à base de connaissances. Nous avons proposé une méthode de conception d'hypermédias à base de connaissances taxinomiques, fondée sur une méthode d'acquisition de connaissances et sur le typage et l'indexation des éléments du réseau hypermédias à ceux de la base des connaissances acquises.

Nos Contributions

Un premier travail, préalable à celui de représentation de connaissances taxinomiques proprement dit, a été de caractériser de telles connaissances. Nous avons dégagé les spécificités des domaines qu'elles modélisent et leurs utilisations pour les traduire ensuite en termes de contraintes dans le choix d'un formalisme de représentation. Sur cette base, nous avons montré l'adéquation des graphes conceptuels à cette tâche particulière de représentation de connaissances. Tout d'abord, ce formalisme est dédié à la représentation de connaissances déclaratives : il permet de formaliser des descriptions structurées de concepts à l'aide de graphes conceptuels et les relations de généralisation entre ces graphes rendent compte de l'organisation taxinomique des concepts. Ensuite, des connaissances représentées dans le formalisme des graphes conceptuels, parce qu'elles sont déclaratives et structurées, se prêtent de ce fait particulièrement bien à l'organisation d'un réseau hypermédia. De plus, la formalisation des descriptions des concepts sous forme de graphes garantit une certaine intelligibilité de ces descriptions et leur visualisation est facile à réaliser, précisément sur les noeuds d'information du réseau hypermédia qu'elles organisent. Enfin, les opérations définies sur les graphes conceptuels munissent ce formalisme de représentation de capacités de raisonnement. La manipulation formelle des graphes permet notamment de comparer et discriminer des descriptions de concepts, de déterminer les relations de subsomption entre concepts, de comparer différentes taxinomies, d'expliciter les traits descriptifs des concepts qui rendent compte de leur organisation taxinomique, etc. . Cependant, en adoptant les graphes conceptuels simples [Sowa, 84][Mugnier et Chein, 96] pour représenter des connaissances taxinomiques, nous avons été confrontés à l'imperfection des taxinomies naturelles qui se traduit par des exceptions à l'héritage entre taxons. Partant, nous avons proposé d'étendre les graphes conceptuels simples qui, dès lors que les taxons ne sont plus décrits par des propriétés toutes strictes, ne permettent plus de formaliser leurs descriptions.

Notre contribution dans le domaine de la représentation des connaissances a consisté à introduire une représentation des défauts et des exceptions dans les graphes conceptuels simples. Pour cela, nous avons adapté une démarche définitionnelle proposée par Coupey et Fouqueré dans le cadre des logiques de descriptions [Coupey et Fouqueré, 97]. Nous avons d'abord montré l'inadéquation pour les graphes conceptuels des approches inférentielles propres à la logique des défauts, aux réseaux sémantiques et à la plupart des travaux dans les logiques de description. Nous avons mis en évidence les similarités qui existent entre la TBox d'un langage de description et le support du modèle des graphes conceptuels, dotés chacun d'un mécanisme de définition - de termes ou de types. Sur le modèle du langage de description $AL_{\delta\varepsilon}$, nous avons alors défini le support $S_{\delta\varepsilon}$ comme le support S du modèle des graphes conceptuels simples enrichi de deux connecteurs dont la sémantique et les propriétés sont comparables à celles des connecteurs de $AL_{\delta\varepsilon}$. Les différences qui apparaissent entre $S_{\delta\varepsilon}$ et $AL_{\delta\varepsilon}$ trouvent leur origine dans les natures différentes des mécanismes de définition de type associés à ces deux formalismes. A l'issue de ce travail, nous pensons qu'une comparaison plus formelle des graphes conceptuels et des logiques de description ouvrirait une perspective plus large de faire bénéficier les graphes conceptuels de certains résultats établis pour les logiques de description.

A la différence des approches inférentielles des exceptions à l'héritage basées sur l'introduction d'implications “faibles”, dans l'approche définitionnelle que nous avons adoptée, les défauts et les exceptions sont gérés au travers même des définitions des descriptions des concepts. Un des avantages d'une telle démarche est que les opérations de spécialisation définies sur les graphes conceptuels simples ne sont pas modifiées et restent le fondement des relations de subsomption entre concepts. Seuls sont modifiés les graphes sur lesquels ces opérations s'accomplissent. Nous avons ainsi pu définir le processus d'exception entre deux concepts comme une succession d'opérations *standard* de spécialisation de graphe. De même, en ce qui concerne l'acquisition des descriptions des concepts, nous avons montré que le passage du support S au support $S_{\delta\varepsilon}$ ne modifie pas le principe de la méthode descendante d'explicitation des connaissances que nous avons adoptée, qui procède par spécialisation et est de fait basée sur l'utilisation des opérations de spécialisation de graphe.

Notre seconde contribution s'inscrit dans la problématique générale de l'intégration de connaissances dans les hypermédias pour les doter de moyens d'accès conceptuels à l'information mémorisée. Dans ce cadre, nous nous sommes intéressés à l'intégration de *connaissances du domaine* dans les hypermédias en mettant en évidence l'intérêt d'un modèle du domaine pour l'organisation d'un réseau hypermédia sémantiquement riche. Plus précisément notre problématique concerne l'intégration d'un modèle du domaine à un réseau hypermédia et la réduction de la charge cognitive du concepteur du réseau que représente cette tâche d'explicitation de connaissances. Nous avons proposé une méthode de conception d'hypermédias à base de connaissances taxinomiques représentées dans le formalisme des graphes conceptuels et explicitées selon une méthode d'acquisition de connaissances.

La méthode que nous proposons consiste à organiser le réseau hypermédia à l'aide d'un mécanisme de typage des noeuds, ancres et liens hypermédias, à partir d'un schéma de modèle du domaine et selon une méthode d'acquisition de connaissances. Le choix des graphes conceptuels pour représenter les connaissances du domaine assure un typage des éléments du réseau et donc une organisation du réseau sémantiquement riches et structurés. Le schéma de modèle que nous avons défini est dépendant du type de domaines que nous cherchons à modéliser - à savoir les domaines des sciences naturelles caractérisés par une organisation taxinomique et des descriptions structurées de taxons. Ce schéma sert de guide pour l'expert dans l'explicitation d'un modèle du domaine représenté par une base canonique de graphes. A partir de celle-ci, nous mettons en oeuvre une méthode d'acquisition de connaissances fondée sur les opérations de spécialisation de graphes. La tâche d'organisation du réseau hypermédia est confondue avec celle d'acquisition de la base de connaissances qui modélise le domaine ; les éléments du réseau sont automatiquement créés, typés et indexés aux connaissances au fur et à mesure de leur acquisition. Ainsi, à l'inverse des approches de la conception d'hypermédias vue comme une méthode d'acquisition de connaissances pour l'explicitation d'un modèle hypermédia du domaine, nous envisageons dans notre approche la conception d'hypermédias en termes d'acquisition de connaissances. L'implantation du système SATELIT et son application à la construction d'un hypermédia à base de connaissances en orchidologie témoignent de la faisabilité de notre approche.

Limites et perspectives

Améliorations du système SATELIT

Un premier ensemble de perspectives est lié à l'amélioration du système SATELIT. Tout d'abord, à l'issue des expérimentations que nous avons menées en orchidologie, il apparaît clairement que pour traiter une taxinomie telle que celle de la famille des *Orchidaceae* qui comporte plusieurs centaines de taxons, le développement de SATELIT en JAVA est indispensable. Cette nouvelle implantation permettra la manipulation d'une large base de connaissances.

Une autre perspective liée au développement de SATELIT est de modifier la méthode d'acquisition de connaissances mise en oeuvre, qui s'avère par trop rigide. En effet, les descriptions de taxons qui sont explicitées ne comportent que des traits descriptifs pertinents pour les *discriminer*. Or d'autres traits de caractérisation apparaissent dans les descriptions textuelles qui sont indexées aux graphes ainsi construits. Nous envisageons de modifier le processus d'interrogation de l'expert du domaine afin que celui-ci puisse acquérir les traits descriptifs qu'il souhaite. De plus, telle qu'elle est conçue, la méthode d'acquisition impose qu'un arc conceptuel générique dans la description d'un taxon soit spécialisé pour chacun des taxons fils. Or les descriptions de taxons ne sont pas toujours homogènes de sorte qu'il peut parfois être difficile à l'expert de fournir les spécialisations attendues en réponse aux questions que lui pose le système. Nous envisageons donc d'assouplir encore la méthode d'acquisition des connaissances mise en oeuvre en acceptant que l'expert puisse ne pas répondre à certaines questions. Dans cette perspective, on pourrait tirer partie des travaux sur la gestion d'information incomplète.

Une autre perspective, également liée au développement de SATELIT, est d'intégrer d'autres procédures d'hypertextualisation en aval du processus de construction automatique du réseau hypermédia fondée sur l'acquisition de connaissances. En effet, une fois qu'un réseau hypermédia initial a été automatiquement organisé au fur et à mesure du processus d'acquisition de connaissances, il est dans un deuxième temps enrichi manuellement par la mémorisation d'informations textuelles et graphiques sur les noeuds d'information créés lors de la première phase de construction automatique et par la création de nouveaux liens à partir des ancres que constituent certains mots des textes ou certaines zones des croquis. Typiquement, des techniques d'analyse de textes telles que celles citées dans le chapitre 3 trouveraient ici leur place, qui permettent de créer automatiquement des liens sémantiques entre les documents hypermédias.

Enfin, il nous semble important d'explorer les techniques d'interface homme-machine pour améliorer la convivialité et l'ergonomie du système SATELIT. A terme, notre but est en effet de faire de SATELIT un outil qui puisse être proposé aux systématiciens pour expliciter leur expertise, informatiser les catalogues botaniques ou zoologiques et permettre d'exploiter l'information mémorisée grâce aux connaissances expertes explicitées qui modélisent le domaine. Dans cette perspective, les travaux menés dans notre équipe pour développer un système d'aide à la navigation et à la recherche d'information dans SATELIT sont tout à fait complémentaires des nôtres [Akoulchina et Ganascia, 97]. Nous envisageons également de coupler SATELIT avec le système COING de regroupement conceptuel [Bournaud, 96] afin d'offrir aux systématiciens un outil d'investigation qui, à partir de l'explicitation de leur expertise, leur permette de réviser les structures taxinomiques établies du domaine.

Perspectives théoriques

Un premier type de perspectives théoriques concerne nos travaux de représentation des connaissances. Tout d'abord, nous avons l'intention d'établir l'interprétation logique de l'extension du modèle des graphes conceptuels simples que nous avons proposée. Il s'agit de traduire en calcul des prédicats les graphes construits sur le support $S_{\delta\varepsilon}$ pour en expliquer formellement la sémantique.

Nous avons également l'intention de poursuivre la comparaison des graphes conceptuels et des logiques de descriptions, dans un cadre plus formel. Nous pensons que le fait d'avoir pu adapter les travaux de Coupey et Fouqueré en logiques de descriptions [Coupey et Fouqueré, 97] pour représenter des défauts et exceptions dans le modèle des graphes conceptuels ouvre des perspectives prometteuses de faire bénéficier les graphes conceptuels d'autres résultats déjà établis dans le cadre des logiques de descriptions.

Une autre perspective qui a trait à la représentation des exceptions sur les taxinomies concerne la révision de taxinomies. Il nous semble en effet intéressant pour mieux comprendre certains modèles de domaines de pouvoir en modifier les structures taxinomiques en reconsidérant la place des concepts dont les descriptions comportent des exceptions aux traits descriptifs des concepts qui les subsument. Symétriquement, des structures taxinomiques strictes peuvent être révisées en faisant apparaître dans la description de certains concepts des exceptions aux traits descriptifs d'autres concepts, ceux-ci subsumant alors ceux-là. Dans cette perspective, nous nous intéressons aux travaux de Bournaud sur le regroupement conceptuel [Bournaud, 96] et à ceux de Coupey et Fouqueré sur le classement d'instances exceptionnelles [Coupey et Fouqueré, 94b].

Enfin, une perspective à plus long terme est de pouvoir construire des graphes conceptuels quelconques pour décrire les concepts. Une des limites de notre approche est liée au choix d'un langage de description des concepts sous la forme d'arcs conceptuels. C'est ce qui nous a permis de limiter la complexité de la discrimination des concepts fondée sur l'appariement de graphes. La contrepartie de ce choix est un pouvoir d'expression affaibli du langage de description des concepts. Une approche permettant d'augmenter progressivement la richesse des appariements entre graphes est celle proposée dans le système REMO [Zucker, 96]. Cependant, il est clair que les graphes que nous considérons ne sont pas quelconques et partagent une certaine structure. Un mécanisme plus simple pourrait permettre de pallier le problème particulier de l'appariement de graphes décrivant des taxons. Une voie de recherche qui nous semble prometteuse est d'utiliser des schémas de graphes pour guider l'appariement. C'est l'approche proposée par Vismara [Vismara, 96].

Bibliographie

[Aïmeur, 94] E. Aïmeur. *METIS : un système et une Méthode d'Explicitation de Taxinomies destinées à l'Identification de Structures conceptuelles*. Thèse de doctorat de l'université Paris VI, 1994.

[Aïmeur et Ganascia, 93] E. Aïmeur, J.G. Ganascia. Elicitation of Taxonomies based on the use of Conceptual Graph Operators. In *Proceedings of ICCS'93, 1st International Conference on Conceptual Structures. Lecture Notes in AI, 699*. Springer Verlag, Berlin, pages 361-379, Québec, 1993.

[Akoulchina et Ganascia, 97] I. Akoulchina, J.G. Ganascia. SATELIT-Agent: an Adaptive Interface Based on Learning Interface Agent Technology. In *Proceedings of UM'97, 6th International Conference on User Modeling*. pages 21-32, Sardinia, 1997.

[Auroi, 92] C. Auroi. *La diversité biologique : la vie en péril*. Georg, Genève, col. Dossier de l'Environnement, 7, 1992.

[Balpe et al, 96] J.P Balpe, A. Lelu, F. Papy, I. Saleh. *Techniques avancées pour l'hypertexte*. Hermès, Paris, 1996.

[Baader et Hollunder, 92] Baader, B. Hollunder. Embedding Defaults into Terminological Knowledge Representation Formalisms. In *Principles of Knowledge Representation and Reasoning: Proceedings of the 3rd International Conference (KR'92)*. Morgan Kaufmann Publishers, San Mateo, pages 306-317, Cambridge, 1992.

[Baader et Hollunder, 94] Baader, B. Hollunder. Computing Extension of Terminological Default Theories. In *Foundations of Knowledge Representation and Reasoning. Lecture Notes in Artificial Intelligence, 810*. Springer Verlag, Berlin, pages 30-52, 1994.

[Bareiss et Osgood, 93] R. Bareiss, R. Osgood. Applying AI Models to the Design of Exploratory Hypermedia Systems. In *Proceedings of Hypertext'93, 4th ACM Conference on Hypertext.* ACM Press, New York, pages 94-104, Seattle, 1993.

[Berge, 70] C. Berge. *Graphes*. Dunod, Paris, 1970.

[Biébow et Chaty, 93] B. Biébow, G. Chaty. A Comparison Between Conceptual Graphs and KL-ONE. In *Proceedings of ICCS'93, 1st International Conference on Conceptual Structures. Lecture Notes in AI, 699*. Springer Verlag, Berlin, pages 75-89, Québec, 1993.

[Bournaud, 96] I. Bournaud. *Regroupement conceptuel pour l'organisation de connaissances*. Thèse de doctorat de l'université Paris VI, 1996.

[Bournaud et Ganascia, 95] I. Bournaud, J.G. Ganascia. Conceptual clustering of complex objects: a generalization space based approach. In *proceedings of ICCS'95, 3rd International Conference on Conceptual Structues. Lecture Notes in Artificial Intelligence, 954*. Springer-Verlag, Berlin, pages 173-187, Santa Cruz, CA, 1995.

[Boy, 91] G. Boy. Indexing Hypertext Documents in Context. In *Proceedings of Hypertext'91, 3rd ACM Conference on Hypertext.* ACM Press, New York, pages 51-61, San Antonio, 1991.

[Brachman, 85] R.J. Brachman. 'I Lied About the Trees' or Defaults and Definitions in Knowledge Representation System. In *The AI Magazine, 6 (3)*, pages 80-93, 1985.

[Brachman et Schmolze, 85] R.J. Brachman, J.G Schmolze. An Overview of the KL-ONE Knowledge Representation System. In *Cognitive Science, 9*, pages 171-216, 1985.

[Chang, 93] D. Chang. HieNet: a User-Centered Approach for Automatic Link Generation. In *Proceedings of Hypertext'93, 4th ACM Conference on Hypertext.* ACM Press, New York, pages 145-158, Seattle, 1993.

[Chauvet et Olivier, 93] M. Chauvet, L. Olivier. *La biodiversité, enjeu planétaire : préserver notre patrimoine génétique*. Sang de la terre, Paris, 1993.

[Chein et Leclère, 94] M. Chein, M. Leclere. A Cooperative Program for the Construction of a Concept Type Lattice. In *Supplement Proceedings of ICCS'94, 2nd International Conference on Conceptual Structures*, pages 16-30, Maryland, 1994.

[Chein et Mugnier, 92] M. Chein, M.L. Mugnier. Conceptual Graphs : Fundamental Notions. In *Revue d'Intelligence Artificielle, 6 (4)*, pages 365-406, 1992.

[Conruyt, 94] N. Conruyt. *Amélioration de la robustesse des systèmes d'aide à la description, à la classification et à la détermination des objets biologiques*. Thèse de doctorat de l'université Paris IX, 1994.

[Coupey et Fouqueré, 94a] P. Coupey, C. Fouqueré. Définir un concept avec des propriétés par défaut et des exceptions. In *Actes de RFIA'94, Conférence Reconnaissance des Formes et Intelligence Artificielle*, pages 197-208, Paris, 1994.

[Coupey et Fouqueré, 94b] P. Coupey, C. Fouqueré. Classer des concepts définis avec des défauts et des exceptions In *Actes de LMO'94, Conférence Langages et Modèles à Objets*, pages 69-80, Grenoble, 1994.

[Coupey et Fouqueré, 97] P. Coupey, C. Fouqueré. Extending Conceptual Definitions with Default Knowledge. In *Computational Intelligence, 13(2)*, pages 258-299, 1997.

[Croft et Turtle, 89] W. Croft, H. Turtle. A retrieval model for incorporating hypertext links. In *Proceedings of Hypertext'89 , 2nd ACM Conference on Hypertext*. ACM Press, New York, pages 213-224, Pittsburgh, 1989.

[Darricau et al., 97] G. Darricau, H. Hadj-Mabrouk, J.G. Ganascia. Acquisition et représentation des connaissances impliquées dans les spécifications des logiciels de sécurité - Application au pilotage automatique. Rapport INRETS n° 97-29, 1997.

[Etherington et Reiter, 83] D. Etherington, R. Reiter. On Inheritance Hierarchies With Exceptions. In *Proceedings of AAAI'83*, *National Conference of the American Association for Artificial Intelligence*, pages 104-108, Washington, 1983.

[Euzenat, 95] J. Euzenat. Building Consensual Knowledge Bases: Context and Architecture. In *Nicolaas Mars (éd.), Proceedings of KBKS'95, 2nd International Conference on Building and Sharing of Very Large-Scale Knowledge Bases*. IOS Press, Amsterdam, pages 143-155, Enschede (NL), 1995.

[Euzenat, 96] J. Euzenat. Corporate Memory Through Cooperative Creation of Knowledge Bases and Hyper-Documents. In Proceedings of 10th Knowledge Acquisition Workshop, pages (36)1-18, Banff (CA), 1996.

[Falhman, 79] C. Falhman. *NETL : A System for Representing and Using Real-World Knowledge*. MIT Press, Cambridge, 1979.

[Falhman et al, 81] C. Falhman, D.S. Touretzsky, W. van Rogger. Cancellation in a Parallel Semantic Network. In *Proceedings of IJCAI'81, 7th International Joint Conference on Artificial Intelligence*, pages 257-263, Vancouver, 1981.

[Faron et Kieu, 95] C. Faron, Q. Kieu. SATELIT: un outil d'explicitation de connaissances hypermédias. In *Actes des JAC'95, Journées d'Acquisition des connaissances*, pages 187-198, Grenoble, 1995.

[Faron et al, 96] C. Faron, I. Akoulchina, J.G. Ganascia. Integrating Knowledge in Hypermedias: the SATELIT System. In *Proceedings of CESA'96, Computational Engineering in Systems Applications, Symposium "Cybernetic and Robotic", vol. 4*, pages 94-98, Lille, 1996.

[Faron et Ganascia, 97a] C. Faron, J.G. Ganascia. Representation of Defaults and Exceptions in Conceptual Graphs Formalism. In *Proceedings of ICCS'97, 5th International Conference on Conceptual Structures. Lecture Notes in Artificial Intelligence, 1257.* Springer Verlag, Berlin, pages 153-167, Seattle, 1997.

[Faron et Ganascia, 97b] C. Faron, J.G. Ganascia. Acquisition and Maintenance of Taxonomic Knowledge Embedding Defaults when Represented in Conceptual Graphs Formalism. In *Proceedings of KRUSE'97, 2nd International Symposium on Knowledge Representation, Use and Storage*, pages 136-150, Vancouver, 1997.

[Frisse et Cousins, 89] M.E. Frisse, S.B. Cousins. Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook Project. In *Proceedings of Hypertext'89, 2nd ACM Conference on Hypertext*. ACM Press, New York, pages 199-212, Pittsburgh, 1989.

[Froidevaux, 86] C. Froidevaux. Taxonomic Default. In *Proceedings of ECAI'86, European Conference on Artificial Intelligence. Advances in Artificial Intelligence*. North-Holland, Amsterdam, pages 305-311, Brighton, 1986.

[Furtado Vasco et al, 96] J. Furtado Vasco, C. Faucher, E. Chouraqui. A Knowledge Acquisition tool for Multi-Perspective Concept Formation. In *Proceedings of EKAW'96, European Knowledge Acquisition Workshop*. Springer Verlag, Berlin, pages 229-244, 1996.

[Gallaire, 87] H. Gallaire. La représentation des connaissances. In *La Recherche en intelligence artificielle*. Editions du Seuil, Paris, pages 265-284, 1987.

[Haake et al, 94] J. Haake, C. Neuwirth, N. Streitz. Coexistence and Transformation of Informal Structures: Requirements for more Flexible Hypermedia Systems. In *Proceedings of ECHT'94, 3rd ACM European Conference on Hypertext Technology*, ACM Press, New York, pages 1-12, Edinburgh, 1994.

[Haussler, 89] D. Haussler. Learning conjunctive concepts in structural domains. In *Machine learning, 4*, pages 7-40, 1989.

[Hofmann et al, 90] M. Hofmann, U. Schreiweis, H. Langendörfer. An Integrated Approach of Knowledge Acquisition by the Hypertext System CONCORDE. In *Proceedings of ECHT'90, 1st European Conference on Hypertext Technology. Hypertext: Concepts, Systems and Applications*. Cambridge Press University, pages 166-179*,* INRIA-France, 1990.

[Kheirbek et Chiaramella, 95] A. Kheirbek, Y. Chiaramella. Integrating Hypermedia and Information Retrieval with Conceptual Graphs Formalism. In *Proceedings of HIM'95, Hypertext - Information Retrieval - Multimedia.* Universitatsverlag Konstanz, pages 47-60, 1995.

[Lalande, 56] A. Lalande. *Vocabulaire technique et critique de la philosophie*. Presse Universitaire de France, Paris, 1956.

[Leclère, 95] M. Leclère. *Les connaissances du niveau terminologique du modèle des graphes conceptuels : construction et exploitation*. Thèse de doctorat de l'université Montpellier II, 1995.

[Leclère, 96] M. Leclère. C-CHiC Construction coopérative de hiérarchies de catégories. In *Revue d'Intelligence Artificielle, 10 (1)*, pages 57-100, 1996.

[Lecoufle, 81] Lecoufle. *Orchidées exotiques*. La Maison rustique, Paris, 1981.

[Lehman, 92] F. Lehman, éditeur. *Semantic Networks in Artificial Intelligence*. Pergamon Press, Oxford, 1992.

[Lemaire et Rechenmann, 96] F. Lemaire, F. Rechenmann. Intégration de connaissances terminologiques dans les grandes bases d'objets :exemples en biologie moléculaire. In

La banque des mots. Revue de Terminologie Française, Numéro spécial 7, "Terminologie et Intelligence Artificielle", pages 105-112, 1996.

[LeRobert, 94] *Dictionnaire Historique de la langue française*. Les Dictionnaires Le Robert, Paris, 1994.

[Lucarella, 90] D. Lucarella. A model for hypertext-based information retrieval. In *Proceedings of ECHT'90, 1st European Conference on Hypertext Technology. Hypertext: Concepts, Systems and Applications*. Cambridge University Press, pages 81-94, INRIA-France, 1990.

[Lucas, 91] L. Lucas. Visually Designing the Computer Learner Interface. In *Educational Technology*, 31 (7), pages 56-58, 1991.

[Marshall et al, 91] C. Marshall, F.G. Halasz, R.A. Rogers, W.C. Jansen. Aquanet : a Hypertext Tool to Hold your Knowledge in Place. In *Proceedings of Hypertext'91, 3rd ACM Conference on Hypertext*. ACM Press, New York, pages 261-274, San Antonio, 1991.

[Marshall et Rogers, 92] C. Marshall, R.A. Rogers. Two Years before the Mist: Experiences with Aquavet. In *Proceedings of ECHT'92, 2nd ACM European Conference on Hypertext*. ACM Press, New York, pages 23-32, Milano, 1992.

[Marteau et Krumeich, 95] P.F. Marteau, C. Krumeich. Semantic Analysis of Text Applied to Competitive Intelligence. In *Proceedings of idt'95, 12th Conference on information industry and market*. Paris, pages 258-265, 1995.

[Martin, 95] P. Martin. Links between Electronic Documents and a Knowledge Base of Conceptual Graphs. In *Proceedings Supplement of ICCS'95, 3rd International Conference on Conceptual Structures,* pages 112-125, Santa Cruz, 1995.

[Matile et al, 87] L. Matile, P. Tassy, D. Goujet, Société Française de Systématique. Introduction à la systématique zoologique: concepts, principes, méthodes. In *Biosystema 1, Société Française de Systématique,* Paris, 1987.

[Mayr, 89] E. Mayr. *Histoire de la biologie*. Fayard, Paris, 1989.

[Mineau, 90] G.W. Mineau. *Structuration des bases de connaissances par généralisation.* Ph.D. d'informatique de l'université de Montréal, 1990.

[Mugnier et Chein, 93] M.L. Mugnier, M. Chein. Characterization and Algorithmic Recognition of Canonical Conceptual Graphs. In *Proceedings of ICCS'93, 1st International Conference on Conceptual Structures. Lecture Notes in Artificial Intelligence, 699*. Springer Verlag, Berlin, pages 294-311, Québec, 1993.

[Mugnier et Chein, 96] M.L. Mugnier, M. Chein. Représenter des connaissances et raisonner avec des graphes. In *Revue d'Intelligence Artificielle, 10 (1)*, pages 7-56, 1996.

[Myaeng, 92] S. Myaeng. Conceptual Graphs as a Framework for Text Retrieval. In *Conceptual Structures. Current research and practice*. Ellis Horwood, New York, pages 559-573, 1992.

[Nanard et Nanard, 91] J. Nanard et M. Nanard. Using Structured Types to Incorporate Knowledge in Hypertext. In *Proceedings of Hypertext'91, 3rd ACM Conference on Hypertexts*. ACM Press, New York, pages 329-342, San Antonio, 1991.

[Nanard et Nanard, 93] J. Nanard, M. Nanard. Should Anchors Be Typed Too? An Experiment with MacWeb. In *Proceedings of Hypertext'93, 4th ACM Conference on Hypertexts*. ACM Press, New York, pages 51-62, Seattle, 1993.

[Nanard et al, 93] M. Nanard, J. Nanard, A.M. Massotte, J. Chauché, A. Djemaa, A. Joubert, H. Betaille. La métaphore du généraliste : Acquisition et utilisation de la connaissance macroscopique sur une base de documents techniques. In *actes des JAC'93, Journées d'Acquisition de Connaissances*, Saint Raphaël, 1993.

[Napoli et Volle, 93] A. Napoli, P. Volle. Une introduction aux logiques terminologiques. Technical Report CRIN 93-R-033, Nancy, 1993.

[Normark et Osterbye, 95] K. Normark, K. Osterbye. Rich Hypertext: a Foundation for Improved Interaction Techniques. In *International Journal of Human-Computer Studies, 43*, pages 301-321, 1995.

[Padgham, 89] L. Padgham. *NonMonotonic Inheritance for an Object Oriented Knowledge Base*. Ph.D. thesis 213, Linköping University, 1989.

[Padgham et Zhang, 93] L. Padgham, T. Zhang. A Terminological Logic with Defaults: A Definition and an Application. In *Proceedings of the 13th IJCAI, International Joint Conference on Artificial Intelligence*, pages 662-668, Chambéry, 1993.

[Padgham et Nebel, 93] L. Padgham, B. Nebel. Combining Classification and Nonmonotonic Inheritance Reasoning: A First Step. In *Proceedings of the 7th International Symposium on Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence, 689*. Springer Verlag, Berlin, pages 132-141, Norway, 1993.

[Quantz et Royer, 92] J. Quantz, V. Royer. A Preference Semantics for Defaults in Terminological Logics. In *Principles of Knowledge Representation and Reasoning: Proceedings of the 3rd International Conference (KR'92).* Morgan Kaufmann Publishers, San Mateo, pages 294-305, Cambridge, 1992.

[Quint et Vatton, 92] V. Quint, I. Vatton. Combining Hypertext and Structured Documents in Grif. In *Proceedings of ECHT'92, 2nd ACM European Conference on Hypertext.* ACM Press, New York, pages 23-32, Milano, 1992.

[Rechenmann, 93] F. Rechenmann. Building and sharing large knowledge bases in molecular genetics. In *Proceedings of KBKS'93, 1st International Conference on Building and Sharing of Very Large-Scale Knowledge Bases*, pages 291-301, Tokyo, 1993.

[Reiter, 78] R. Reiter. On Reasoning by Default. In *Proceedings of TINLAP, Theoretical Issues in Natural Language Processing,* University of Illinois, 1978. In *Readings in Knowledge Representation.* Morgan Kaufmann Publishers, Los Altos, pages 401-410, 1985.

[Reiter, 80] R. Reiter. A Logic for Default Reasoning. In *Artificial Intelligence, 13(1, 2)*, pages 81-132, 1980.

[Salton et Allan, 93] G. Salton, J. Allan. Selective Text Utilization and Text Traversal. In *Proceedings of Hypertext'93, 4th ACM Conference on Hypertext.* ACM Press, New York, pages 131-144, Seattle, 1993.

[Satoh et al, 93] S. Satoh, A. Takasu, E. Katsura. A Collaborative Supporting Method between Document Processing and Hypertext Construction. In *Proceedings of ICDAR'93, International Conference on Document Analysis and Recognition.* IEEE Computer Society Press, vol 2, pages 533-537, 1993.

[Schank, 82] R.C. Schank. *Dynamic Memory : a Theory of Reminding and Learning in Computers and People.* Cambridge University Press, 1982.

[Schmolze et Lipkis, 83] J. Schmolze, T. Lipkis. Classification in the KL-ONE Knowledge Representation System. In *Proceedings of IJCAI'83, 8th International Joint Conference on Artificial Intelligence*, pages 330-332, 1983.

[Schütt et Streitz, 90] H. Schütt, N. Streitz. HyperBase: a Hypermedia Engine Based on a Relational Database Management System. In *Proceedings of ECHT'90, 1st European Conference on Hypertext Technology. Hypertext: Concepts, Systems and Applications*. Cambridge University Press, pages 95-108, INRIA-France, 1990.

[Shibata et Katsumono, 93] Y. Shibata, M. Katsumono. Dynamic hypertext and knowledge agent systems for multimedia information networks. In *proceedings of Hypertext'93, 4th ACM Conference on Hypertext*. ACM Press, New York, pages 82-92, Seattle, 1993.

[Streitz et al, 92] N. Streitz, J. Haake, A. Lemke, W. Sculer, H. Schütt, M. Thüring. SEPIA : a Cooperative Hypermedia Authoring Environment. In *Proceedings of ECHT'92, 2nd ACM European Conference on Hypertext*. ACM Press, New York, pages 11-22, Milano, 1992.

[Sowa, 84] J.F. Sowa. *Conceptual Structures : Information Processing in Mind and Machine*. Addison-Wesley Publishing Co, Reading, 1984.

[Sowa, 91] J.F. Sowa, éditeurr. *Principles of Semantic Networks : Exploration in the Representation of Knowledge*. Morgan Kaufmann Publishers, San Mateo, 1991.

[Vercoustre, 90] A.M. Vercoustre. Structured Editing - Hypertext Approach: Cooperation and Complementary. In *Proceedings of EP'90 , International Conference on Electronic Publishing. Document Manipulation and Typography*. Cambridge University Press, 1990.

[Vismara, 96] P. Vismara. Appariement dirigé pour le raisonnement par classification sur des hiérarchies de graphes. In *Actes de LMO'96, Confeérence Langages et Modèles à Objets*, pages 201-214, 1996.

[Woodhead, 91] N. Woodhead. *Hypertext and Hypermedia: Theory and Applications*. Sigma Press, Wilmslow, Addison-Wesley Publishing Co, Nokingham, 1991.

[Woods, 75] W. Woods. What's in a link? Foundation for Semantic Networks. In *Representation and Understanding: Studies in Cognitive Science*. Academic Press, New York, pages 35-82, 1975.

[Zizi et Beaudouin-Lafont, 95] M. Zizi, M. Beaudouin-Lafont. Hypermedia Exploration with Interactive Dynamic Maps. In *International Journal of Human-Computer Studies 43(3)*. pages 441-464, 1995.

[Zucker, 96] J.D. Zucker. *Appariements et changements de représentation pour l'apprentissage symbolique*. Thèse de doctorat de l'université Paris VI, 1996.