

INTRODUCTION

Il paraîtra très paradoxal d'accorder une grande importance à l'observation dans cette partie des sciences mathématiques qu'on appelle généralement les mathématiques pures, puisqu'on estime couramment que l'observation concerne seulement les objets qui impressionnent les sens. Puisque nous devons rattacher les nombres à l'intellect pur, nous avons peine à comprendre comment des observations et des quasi-expériences peuvent être de quelque utilité dans l'étude de leur nature. Et pourtant, en fait, comme je le montrerai ici par des arguments très sûrs, les propriétés des nombres, connues aujourd'hui, ont été découvertes principalement par l'observation et elles l'ont été bien longtemps avant d'être confirmées par des démonstrations rigoureuses. Nombreuses sont même les propriétés des nombres qui nous sont familières mais que nous ne sommes pas encore capables de prouver; seule l'observation nous a conduit à les connaître. Ainsi voyons-nous que dans la théorie des nombres, théorie encore très imparfaite, nous pouvons fonder sur l'observation les espoirs les plus grands; elle nous conduira à de nouvelles propriétés que nous entreprendrons de prouver par la suite. Cette sorte de connaissance qui s'appuie seulement sur l'observation et dont la validité n'est pas encore confirmée, doit être soigneusement distinguée de la vérité; on dit habituellement qu'on l'atteint par induction. Néanmoins nous avons rencontré des cas où la simple induction conduisait à l'erreur. Aussi devons nous avoir grand soin de ne pas accepter comme vraies des propriétés des nombres que nous avons découvertes par observation et qui s'appuient sur l'induction seule. Nous devons voir là l'occasion d'étudier de façon plus précise les propriétés découvertes, de les prouver ou de les réfuter; dans les deux cas nous apprendrons certainement quelque chose d'utile.

Euler¹ (1707-1783)

¹ Euler, *Opera Omnia*, 1^{ère} série, vol. 2, p. 459, *Specimen de usu observationum in mathesi pura*.

Cette citation exprime très clairement la démarche d'expertise d'un domaine, que ce soient les mathématiques, la biologie, ou un autre domaine.

Le sujet de cette thèse est : “Amélioration de la robustesse des systèmes d’aide à la description, à la classification et à la détermination des objets biologiques”. En tant que biologiste et informaticien de formation, et étant arrivé au terme de cette thèse, je suis très sensible à ces écrits d'un mathématicien du XVIII^{ème} siècle certainement influencé par les recherches fructueuses des systématiciens de l'époque. Ces idées ont été reprises par G. Pólya dans son livre “Les mathématiques et le raisonnement plausible” (1957) dont je joins en annexe le premier chapitre sur l'induction et l'analogie en mathématiques. En découvrant ces écrits, je suis heureux de constater que les naturalistes et certains mathématiciens adoptent finalement la même approche pour tirer le meilleur parti possible de leurs expériences : leur démarche commune semble basée sur l'observation intime des faits et un raisonnement “plausible” de type inductif et analogique.

L'expert (mathématicien ou biologiste) qui explore son domaine bâtit des hypothèses (ou conjectures) à partir d'interprétations de ses observations qui indiquent les relations *familiales* qu'il voit entre les différents objets. Aux nombres des mathématiciens correspondent les spécimens dans une collection pour le biologiste systématicien.

En écrivant ces relations, il aboutit à des formes de description plus ou moins bien établies qui lui permettent de comprendre son domaine. Il peut utiliser pour cela des manières différentes de représenter ses observations : descriptions monographiques, dessins, photos, ces dernières étant certainement les plus délicates pour véhiculer son expertise. Sans relâche, il confronte ses descriptions à la réalité des nouveaux individus pour mieux les connaître. En développant sa familiarité avec les spécimens, c'est un peu comme si l'expert utilisait les mots de sa propre langue et était capable de «lire les spécimens» (comme le mathématicien «lit les nombres») dans une meilleure compréhension de leur structure [Aubé, 1991], ainsi que dans la résolution de problèmes tels que la classification et la détermination en biologie.

Aujourd'hui, avec l'omniprésence des ordinateurs dans les laboratoires de recherche, il devient opportun que l'informaticien coopère avec l'expert dans sa démarche de familiarisation avec les objets de son travail journalier, en lui procurant des outils d'aide :

- 1) à la modélisation de son savoir (pour structurer ses connaissances),
- 2) à la mise au point de descriptions d'objets comparables entre elles,
(permettant de systématiser un processus de description suivant une même structure descriptive),

- 3) au traitement de ces descriptions (pour élaborer un savoir compréhensif ou en tirer des hypothèses ou des règles de décision valides).

Tous ces outils n'ont qu'un seul but pour l'expert : valoriser son expérience et mettre à l'épreuve ses opinions. La méthode d'apprentissage exposée dans cette thèse est une première réponse à cet objectif d'un point de vue de la *classification* en biologie (qui est une démarche inductive comme expliqué au chapitre 3).

L'autre aspect de cette aide est de permettre une meilleure transmission du savoir de l'expert à d'autres personnes du domaine moins qualifiées que lui (ou encore de manière plus ambitieuse au grand public). Le modèle descriptif et les descriptions sont les deux parties obligatoires et prioritaires du savoir à transmettre, les règles n'étant pas nécessaires. En effet, l'utilisateur du "système expert" doit être avant tout capable de suivre le schéma mental de l'expert au travers d'un guide du "savoir observer" (calqué sur son modèle) afin de proposer une description qualitativement équivalente à celle de l'expert. Celle-ci, comparée aux descriptions de l'expert, conditionne la justesse des résultats du système expert à un problème de *détermination* d'une nouvelle observation.

La biologie est un domaine où foisonnent les faits particuliers difficilement exprimables par des règles générales. C'est pourquoi un naturaliste expérimenté n'accepte pas facilement une hypothèse. Contrairement aux mathématiques, les objets d'observation en biologie ne sont pas issus de notre intelligence mais nous sont donnés par la nature sans qu'on les ait choisis. Ils sont des spécimens ou individus très divers, ce qui nécessite de la part de l'expert de nombreuses vérifications de ses hypothèses, qui se traduisent par des adaptations successives de la classification des spécimens. Mais les clés de détermination ou les règles qui en résultent ne peuvent en aucun cas constituer des preuves à démontrer au sens des mathématiques : la variabilité naturelle est trop importante et une exception est si vite trouvée que l'on est obligé de tolérer une certaine latitude dans les classifications naturelles.

Alors, que peut bien signifier la robustesse des systèmes de classification et de détermination, si nous ne sommes pas capables d'apporter la preuve que les règles apprises sont sûres et authentiques ? (autrement dit, le savoir peut-il progresser en l'absence de certitudes ?).

Parmi les quelques réponses que l'on va développer dans le cadre de cette thèse, il est clair en premier lieu que la biologie n'est pas un domaine sujet aux démonstrations, du fait de la multitude des exceptions aux règles induites pour certaines classes, exceptions qui ne sont pas nécessairement connues au moment de leur élaboration. Il suffit de penser aux nouvelles maladies apparaissant chaque année en pathologie végétale ou animale et qui ont des symptômes très

similaires à ceux de certaines autres maladies déjà répertoriées (variabilité inter-classe faible).

De même, la variabilité intra-classe (intra-spécifique ou à l'intérieur d'un même groupe biologique) pouvant être très importante, l'approche que nous trouvons raisonnable consiste à fournir toutes les descriptions possibles de spécimens permettant de couvrir la classe et de les identifier en leur adjoignant l'étiquette du nom de la classe à laquelle chacun de ces spécimens appartient. Le nombre de descriptions est proportionnel à la capacité de l'expert d'appréhender la diversité d'une classe plus ou moins finement. Les descriptions sont en effet le reflet de son expérience à un moment donné. Nous pensons qu'il est préférable de multiplier les descriptions dans chaque classe, car il sera plus simple par la suite de changer le nom de l'étiquette de la classe plutôt que de modifier la description elle-même sans perte d'information. Il convient aussi d'éviter au maximum les descriptions trop généralisantes (les regroupements de descriptions) lors de la constitution d'une base de cas conforme à la *couverture réelle* de la classe, et de déléguer la tâche de généralisation aux outils d'induction.

Inversement, les descriptions biologiques devraient recueillir le maximum d'information significative (c'est-à-dire avec une précision suffisante) dans le but de les classer et les déterminer dans les détails. Idéalement, les descriptions portent sur une collection d'individus physiquement répertoriés et accessibles à de nouvelles interprétations, ce qui donne la possibilité aux descriptions de rester toujours conformes à la réalité des individus. Notre objectif à terme est de fournir des outils **d'aide à la classification et à la détermination de spécimens** du présent et à venir, guidés en cela par le travail quotidien des systématiciens sur des échantillons biologiques.

Ces premières conclusions sont issues de notre démarche de terrain, c'est-à-dire que nous avons été confrontés à des données d'observation réelles et complexes sur différents sujets (diagnostic ou détermination de maladies en pathologie végétale à l'INRA², classification de spécimens au MNHN³), et à des acteurs variés qui n'ont pas la même manière d'observer et de décrire. Avant d'en arriver à proposer des outils informatiques, il a fallu aller "au charbon" pendant deux années pour expérimenter différentes solutions auprès des utilisateurs : systèmes experts de diagnostic, questionnaires papiers, informatisés ensuite pour la saisie des descriptions sur minitel ou micro-ordinateur. Le **chapitre 1** de cette thèse met en avant la démarche d'observation de *l'utilisation* des différentes solutions qui est la seule manière permettant de connaître réellement le problème posé. Nous allons montrer à chaque étape l'évolution de notre cheminement conceptuel qui a justifié la proposition du sujet sur la robustesse des systèmes dans l'aide à la description, la classification et la détermination d'objets biologiques. Il met en

² Institut National de la Recherche Agronomique.

³ Muséum National d'Histoire Naturelle.

valeur la *demande* et les *besoins* des utilisateurs pour des outils d'aide à l'acquisition des connaissances.

Alors que les recherches en informatique se sont surtout axées sur la robustesse statistique des deux aspects du traitement (classification et détermination), peu de travaux ont été réalisés sur la **robustesse des descriptions** elles mêmes en amont du traitement. Et pourtant, les bonnes descriptions conditionnent la pertinence des règles apprises. Nous souhaitons dans cette thèse rétablir l'équilibre en faveur de l'acquisition de bonnes descriptions à apprendre, ce qui correspond de plus à une aspiration essentielle de la part des systématiciens. L'acquisition des connaissances passe donc par des descriptions robustes avant d'appliquer des méthodes de traitement adaptées pour la classification et la détermination.

Pour notre domaine d'expérimentation en biologie, nous avons choisi de bien dissocier le terme de détermination de celui de classification dont les significations seront précisées au chapitre 3. Ici, la détermination concerne un individu dont on cherche le nom de sa classe d'appartenance ce que certains appellent une identification⁴ (voir § 3.4). La classification concerne plutôt un concept dont on cherche à expliciter les caractères distinctifs à l'aide à la fois des descriptions des individus qui appartiennent au concept et des descriptions des individus qui, au contraire, n'y appartiennent pas. Dans ce sens, la classification est le processus qui permet de déterminer un concept, c'est-à-dire d'expliquer les caractères compréhensifs du concept [Petit-Robert]. La détermination possède donc un double sens en fonction de l'objet sur lequel il porte (concept ou individu). Nous emploierons la détermination dans le sens de détermination d'un individu et la classification dans le sens de détermination d'un concept.

Dans ce contexte, la **robustesse** n'est pas statistique mais plutôt **empirique**, c'est-à-dire liée aux objectifs (description, classification et détermination) et aux conditions d'utilisation des outils (nature des utilisateurs et contexte des données). Nous développerons cette notion de robustesse dans le **chapitre 2** du point de vue théorique et pratique et nous confronterons notre vision avec celle des différents utilisateurs.

Définition : la robustesse des systèmes d'aide à la description, à la classification et à la détermination en biologie est l'ensemble des facteurs qualitatifs qui améliore l'acquisition et le traitement des connaissances sur le domaine (compréhension, précision, cohérence, exhaustivité, redondance, fiabilité, facilité de mise à jour, ergonomie, tolérance aux bruits). Elle donne la possibilité de :

- 1) valoriser le travail de l'expert (l'aider à mieux maîtriser son domaine),
- 2) transmettre et utiliser ses connaissances,

⁴ Au sens anglo-saxon du terme.

3) mieux comprendre les systèmes naturels.

La robustesse empirique s'appuie sur l'observation familière du travail quotidien de l'expert qui décrit des spécimens ainsi que sur la prise en compte des interprétations de ses observations et de son vocabulaire (désambiguation) par les autres utilisateurs de son système. Le concept de robustesse n'est jamais acquis définitivement dans les domaines biologiques, il s'adapte et s'améliore progressivement en ayant la connaissance plus intime du problème posé : il est le fruit de l'observation du terrain.

Ainsi, pour améliorer la robustesse générale, nous avons mis au point une méthode d'acquisition de connaissances descriptives dont nous évaluerons l'adéquation à la démarche des naturalistes. En quelques mots, la méthode procède ainsi :

- 1) observation de la diversité des spécimens et modélisation de l'observable dans un modèle descriptif,
- 2) construction d'un questionnaire issu du modèle descriptif et description de l'observé dans une base de cas,
- 3) induction de règles à partir de la base, détermination de nouveaux faits,
- 4) validation de l'observé (les cas), validation de l'observable (le modèle),
- 5) itération.

De plus, nous avons étudié un autre aspect de la robustesse au niveau de la consultation, celui de déterminer un spécimen face aux non-réponses (ou réponses «inconnu») de l'utilisateur. Des outils ont été implantés pour répondre à ces différentes faiblesses (voir plus loin).

Pour acquérir la robustesse escomptée, l'informaticien doit être en mesure de comprendre le domaine étudié et d'apprécier les différences conceptuelles inévitables au niveau du vocabulaire employé par les biologistes. Le **chapitre 3** est ainsi une sorte de glossaire où nous confrontons les différentes acceptions de certains mots clés tels que la classification, l'identification, le concept, l'objet, l'individu, etc..

La clé de voûte de notre étude est le concept de "description" : les deux communautés de chercheurs (en biologie et en informatique) doivent se mettre d'accord sur ce qu'il représente d'un point de vue quantitatif (nombre de descriptions) et qualitatif (niveau des descriptions (espèces ou spécimens), valeur des descriptions) afin que l'informaticien puisse proposer une *offre* adaptée à la sémantique du domaine. Une **description** est par exemple considérée comme **réelle** lorsqu'elle concerne un seul spécimen et **virtuelle** lorsqu'elle "synthétise" les descriptions de plusieurs spécimens ou d'une population considérée comme homogène.

La difficulté est qu'il a fallu travailler à partir de l'existant dans un univers non paramétrique : les données disponibles sont riches, complexes, hétérogènes en qualité et en quantité. Outre leur caractère réel ou virtuel, nous sommes confrontés à des descriptions morphologiques de spécimens "sur la table", ou extraites des livres anciens, ces descriptions étant parfois incomplètes.

L'incomplétude des descriptions résulte soit :

- 1) de l'état de l'échantillon récolté,
- 2) d'opinions préconçues du biologiste qui n'a pas jugé opportun d'apporter une plus grande précision à certaines de ses observations,
- 3) du fait de nouveaux critères inaccessibles aux moyens techniques d'une époque déterminée (critères microscopiques, biochimiques, génétiques, etc.).

Après discussion avec les systématiciens, nous montrons ce que devrait être une "bonne" description de spécimen(s) dans le **chapitre 4**. Une bonne description s'appuie sur trois facultés à bien distinguer :

Un *savoir observer*, ce qui implique de la part de l'expert de modéliser **l'observable** sous la forme d'un modèle descriptif structuré,

Un *savoir décrire* afin d'acquérir **l'observé** sous forme de descriptions de qualité, à l'aide d'un questionnaire bâti selon l'architecture du modèle descriptif,

Un *savoir raisonner*, afin de traiter les connaissances (l'observable et l'observé) selon un certain but : le raisonnement inductif pour obtenir une classification (modélisation de l'observé), le raisonnement déductif et/ou analogique pour obtenir une détermination.

Ces trois savoir-faire sont la clé de voûte de la méthode mise au point au chapitre 2. Modéliser l'observable est la phase la plus critique pour l'obtention des descriptions observées : nous illustrons ce constat à l'aide de notre application sur les éponges marines.

Cette analyse faite, nous exposons le formalisme mathématique retenu pour décrire les objets biologiques observés. Cet effort d'abstraction est nécessaire à la compréhension du domaine pour les informaticiens qui peuvent ensuite développer des méthodes et algorithmes efficaces tenant compte de la sémantique du domaine. Le formalisme retenu est décrit au **chapitre 5**. Il est emprunté à Diday (1987) et a été adapté pour traiter nos descriptions complexes sous forme d'objets de synthèse booléens. Il met en valeur la nature composite (dépendante), plus ou moins précise (spécialisable) et itérative (multi-instanciable) des objets dans les descriptions naturelles. Nous n'utilisons pas les

objets modaux du fait que nous décrivons plutôt des spécimens que des espèces et que les experts préfèrent fonder leurs décisions sur des valeurs descriptives tranchées (ils n'ont pas d'état d'âme pour juger des spécimens !).

Les “bonnes” descriptions de spécimens sont déjà un premier objectif à résoudre pour constituer un système robuste. Il s'agit de représenter et de stocker dans des fichiers informatiques les “images” les plus fidèles possibles des individus étudiés. Ces données descriptives sont instanciées à partir du questionnaire et du modèle descriptif et peuvent alors être transmises telles quelles ou bien traitées par des algorithmes pour exploiter le savoir qu'elles véhiculent.

Pour que la méthode exposée plus haut apporte la robustesse escomptée, il faut tenir compte de la qualité des acteurs (spécimen(s), expert(s), utilisateur(s), outils d'acquisition de connaissances, algorithmes d'apprentissage) pouvant intervenir aux différentes étapes de la chaîne, de manière à augmenter la qualité des descriptions et des règles apprises. En outre, contrairement au naturaliste amateur qui cherche à justifier son modèle par des exemples conformes, l'expert ne dédaigne pas la recherche de contre-exemples pour invalider ses hypothèses. Cette démarche scientifique de remise en cause systématique du modèle descriptif et des règles apprises permet à l'expert d'avancer dans sa recherche de règles plus vraisemblables qui éliminent les contre-exemples (le perfectionnisme du chercheur !). Elle implique néanmoins que l'informaticien lui fournisse des outils d'aide à la mise à jour de ses connaissances, notamment pour assurer la cohérence des anciennes descriptions (certains caractères ne sont plus valides dans le nouveau modèle) ou leur complétude (quand de nouveaux descripteurs sont apparus dans le modèle).

Dans cette thèse, nous avons réalisé une grande partie de ces outils permettant de constituer une chaîne complète depuis la modélisation jusqu'au traitement des connaissances descriptives. En amont de la phase de traitement, nous avons implanté les deux premiers aspects de la méthode (acquérir l'observable dans un modèle descriptif, acquérir l'observé dans un questionnaire) avec un outil baptisé HyperQuest, que nous présentons au **chapitre 6**. L'originalité réside dans le choix de l'approche hypertexte qui permet de respecter une étroite correspondance entre les objets conceptuels décrits au chapitre 5 (objet de synthèse, assertion composite, horde composite, objet classifié, objet muni de méthodes ou de propriétés) et les entités hypertextes (pile, fond, carte, bouton, champs). Cela permet de s'affranchir d'une application particulière et d'atteindre un niveau de généralité inter-applications, de manière à construire automatiquement un questionnaire à partir de la connaissance du modèle descriptif. Ce questionnaire peut être ensuite personnalisé et illustré par l'expert lui-même très facilement de manière à procurer à l'utilisateur les moyens de “savoir observer” et “savoir décrire”. L'ergonomie et la convivialité sont alors des paramètres importants de la robustesse du système pour acquérir des descriptions de qualité.

Ensuite, dans notre méthode opérationnelle, nous avons choisi deux modes de traitement des descriptions en fonction des objectifs poursuivis que nous formalisons au **chapitre 7**. Nous sommes partis de l'algorithme **KATE** [Manago, 1991] qui construit une classification à partir des connaissances structurées du modèle observable et des descriptions observées. En voulant utiliser cette classification comme une clé de détermination, on s'est aperçu que le résultat de la consultation était moins fiable face à de nouvelles observations incomplètes. Ceci est dû au fait que l'on raisonne de manière déductive à partir d'un arbre de décision (une caractérisation des descriptions) dans un contexte empirique, et que les descriptions initiales ne sont plus intégralement accessibles par ce raisonnement.

Nous avons donc développé un processus particulier de raisonnement analogique basé sur la comparaison de toutes les descriptions entre elles et que nous avons baptisé **CaseWork**. Mais au lieu de *comparer les exemples entre eux* en utilisant une mesure de similitude globale et polythétique⁵ coûteuse pour le traitement (comptage des attributs par rapport aux exemples), nous préférons *comparer les attributs entre eux* en reprenant la même mesure de calcul d'entropie que pour l'induction avec KATE (c'est-à-dire compter les exemples par rapport aux attributs). Cette dernière méthode est monothétique⁶. Elle est plus facile à justifier et à expliquer à l'utilisateur, qui peut connaître les attributs ordonnés en fonction de leur pouvoir de séparation des exemples. CaseWork produit des résultats de détermination par comparaison meilleurs que KATE en appliquant le principe du raisonnement à partir des cas, c'est-à-dire à partir de la base d'expériences passées. En effet, CaseWork tient compte de toute la base d'exemples, contrairement à la détermination déductive avec KATE qui utilise un arbre de décision extrait à partir de la base de cas. Notons que ce "savoir raisonner" n'est pas nouveau dans les systèmes de détermination en biologie. On les trouve dans la littérature sous forme de clés à accès multiple ou encore sous forme de programmes de détermination polyclaves [Pankhurst, 1991]. De même, en analyse des données, des mesures de proximité expriment par un nombre les ressemblances ou les dissemblances existant entre toutes les variables qui caractérisent les exemples pris deux à deux [Chandon & Pinson, 1981]. Ces indices sont utilisés pour des problèmes de classification ou de catégorisation (voir § 3.3.2).

Inversement, le raisonnement par cas (ou encore à partir de cas) utilise la mesure de similarité en phase de détermination : c'est aussi un processus de remémoration et d'adaptation en fonction du contexte de la nouvelle observation [Lieber, 1993]. L'intérêt que nous lui portons tient à son aspect complémentaire

⁵ Évaluant les ressemblances et différences entre exemples sur l'ensemble des attributs (méthode d'appariement).

⁶ Basée sur la distribution relative des exemples par rapport aux valeurs possibles de chaque attribut pris séparément (avec élimination des exemples non conformes à la valeur choisie).

par rapport à l'induction. Un projet d'intégration des deux approches est en cours (entre 1992 et 1995) dans le cadre du projet INRECA⁷ (ESPRIT III n° 6322) dont nous joignons en annexe 5 le descriptif sous forme d'un article paru à EWCBR-93, le premier congrès sur le raisonnement par cas en Europe.

Les limites et mérites respectifs des deux approches seront évalués expérimentalement sur les descriptions des éponges marines fournies lors de cette thèse, ainsi que sur d'autres applications industrielles (diagnostic de pannes) qui ne nous concernent pas directement.

Mais pour l'heure en attendant ces résultats, nous affirmons que pour l'objectif de classification en biologie, l'expert peut utiliser la méthode inductive avec KATE afin de découvrir des conjectures par caractérisation des classes des exemples (sous forme d'arbre ou de règles de décision). La plausibilité de ces hypothèses peut ensuite être testée directement par l'expert en examinant les règles apprises, ou par détermination de nouveaux individus (par déduction à partir de l'arbre).

Si l'objectif est uniquement la détermination d'objets (le diagnostic par exemple), il est préférable d'utiliser un outil de recherche par comparaison comme CaseWork qui tient compte de toutes les descriptions et permet d'éviter les impasses dues à la rencontre d'observations localement impossibles (échantillons incomplets, mal conservés, etc.).

Dans tous les cas, une justification des erreurs de classement (voir § 3.3) peut être mise en évidence et révéler l'insuffisance de la base d'exemples (l'observé) ou du modèle descriptif (l'observable). A tout moment, l'expert doit être capable de trouver les raisons des "fautes" commises dans les descriptions. Il devra utiliser les outils de manière interactive, ce qui est une condition d'acceptabilité du service rendu par l'informatique. Les autres conditions sont liées aux moyens mis en œuvre pour acquérir ces descriptions avec l'éditeur de modèle descriptif et de cas (le questionnaire). Ils doivent être à la fois facile d'accès (souples et ergonomiques), scientifiquement rigoureux dans la démarche de description (logique de description, cohérence entre l'observable et l'observé) et permettre d'exploiter toute la richesse informative du domaine naturel (liaisons intrinsèques entre caractères, exhaustivité et précision des données descriptives).

Ce n'est que par cet effort de compréhension de la complexité d'un domaine naturel que l'informaticien est un interlocuteur utile pour l'expert : il sera disposé à adapter le modèle à la réalité des connaissances et non l'inverse.

⁷ INduction and REasoning from CAseS.