MATHÉMATIQUES ET SCIENCES HUMAINES

DENIS PIERRE

Les schémas mentaux : représenter et maintenir une connaissance apprise

Mathématiques et sciences humaines, tome 139 (1997), p. 37-68

<http://www.numdam.org/item?id=MSH_1997__139__37_0>

Math. Inf. Sci. hum., (35e année, n°139, *1997*, pp. 37-68)

LES SCHÉMAS MENTAUX : REPRÉSENTER ET MAINTENIR UNE CONNAISSANCE APPRISE

Denis PIERRE[1]

RÉSUMÉ — *Nous nous intéressons à l'apprentissage à partir d'exemples et à la résolution de problème dans un univers évolutif représenté par une base de connaissances incomplète. Nous formalisons un cadre de représentation de connaissances susceptible d'être élaboré et critiqué par des humains comme par des machines. Cette représentation des connaissances est appelée théorie semi-empirique car cette forme de théorie n'est pas complètement axiomatique. Nous avons formalisé la gestion de la croissance incrémentale de la connaissance dans notre système autour de l'idée de noyau central, issue des recherches en psychologie sociale, en définissant un sous-ensemble cohérent de la base de connaissances comme le concept à apprendre, et en obligeant l'apprenti à persévérer dans cette croyance. La propriété essentielle de ce sous-ensemble est la monotonie du raisonnement.*

SUMMARY — Mental schemes : how to represent and maintain a learned knowledge.
We are interested in learning from examples and problem solving in an evolutive universe represented by an incomplete knowledge base. We formalize a knowledge representation framework that could be built and criticized by human and/or artificial agents. This knowledge representation is called a semi-empirical theory because this kind of theory is not completely axiomatic. We formalize a system called a mental scheme modelling the knowledge increase during the learning process. We deal with the dynamic characteristic of the learning acquisition process through reasoning mechanisms, proof building and the definition of a knowledge core.

1 INTRODUCTION

L'apprentissage est la production et l'évaluation de modèles devant être exploités après leur validation par un expert du domaine. Il existe plusieurs types d'apprentissage répondant à différents critères dont les principaux sont énumérés ci-dessous :

- *Phase d'apprentissage* : en général, les méthodes d'apprentissage se décomposent en deux étapes, apprendre et décider.
- *Connaissance du domaine* : certaines méthodes nécessitent une connaissance *a priori* du domaine, qu'elles adapteront par la suite à de nouveaux contextes.
- *Incrémentalité* : les éléments constituant la connaissance du domaine peuvent être présentés progressivement. Ils sont alors analysés suivant l'ordre de présentation.
- *Résistance au bruit*: la notion de bruit est liée aux erreurs de description ou de mesure. En général, les méthodes incrémentales ne résistent pas au bruit. On admet qu'un grand nombre d'exemples est, en apprentissage, un facteur prépondérant de résistance au bruit.

Les deux grands domaines de recherche en apprentissage en relation avec le problème qui nous intéresse sont l'apprentissage à partir d'explications, ou EBL, et l'apprentissage par détection de similarités, ou SBL.

1. Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), UMR 9928 Université Montpellier II/CNRS, 161 Rue ADA, 34392 Montpellier Cedex 5, FRANCE

L'apprentissage par explications, ou EBL, est un type d'apprentissage dans lequel la généralisation doit être justifiée par des explications. Les systèmes de type EBL expliquent pourquoi une instance particulière, décrite par une collection de littéraux représentant ses caractéristiques et ses valeurs, est un élément d'un concept (un prédicat sur un univers d'instances). Le résultat de l'apprentissage est la description d'un ou de plusieurs concepts. Le raisonnement est basé sur l'explication des observations, et repose sur une connaissance du domaine souvent représentée par une base de règles.
L'apprentissage par explications comprend une phase d'*explication*, qui prouve qu'une instance est exemple d'un concept, en déterminant les caractéristiques de l'exemple pertinentes pour le concept et une phase de *généralisation* qui permet de déterminer un ensemble de conditions suffisantes pour lesquelles la structure d'explication est valide. L'approche EBL est donc une technique déductive qui repose sur une théorie complète du domaine. De plus, ces méthodes nécessitent une définition d'un *concept but*, ou concept à apprendre, disponible avant l'apprentissage.

L'apprentissage par détection de similarités, ou SBL, est une approche dans laquelle un système d'apprentissage reçoit des informations relatives à un domaine d'application sous forme de faits ou de formules. Sur la base de ces informations, le système induit des règles générales caractérisant le domaine, et émet des hypothèses sur des éléments inconnus du domaine. Les informations apportées par la suite peuvent permettre l'*induction* de nouvelles hypothèses, tant qu'elles ne falsifient pas les hypothèses existantes. La consistance de l'ensemble d'hypothèses doit être maintenue lorsque de nouvelles informations sont apportées.

L'objet de notre recherche se caractérise par la formalisation d'un modèle d'apprentissage à partir d'exemples de type SBL, une *théorie semi-empirique*. Une théorie semi-empirique est une forme d'expression des connaissances susceptible d'être élaborée et critiquée par un agent humain ou programmé. En formalisant dans un *schéma mental* l'ensemble des relations qui lient les énoncés d'une théorie semi-empirique, nous souhaitons pouvoir décrire comment un apprenti peut formuler des hypothèses en découvrant progressivement le domaine d'application. Cet apprenti est en situation de résolution de problème initié par une question relative à un domaine sur lequel il n'a aucune connaissance a priori. Il enrichit incrémentalement ses connaissances, au gré des interactions avec un expert, et il dispose de méthodes qui lui permettent de formuler des hypothèses et produire des argumentations.
Les théories semi-empiriques sont une alternative aux formulations logiques puisqu'elles ne présupposent pas l'existence d'un cadre a priori mais permettent l'émergence d'une argumentation, sous la forme d'une preuve, lorsqu'un état stable de la connaissance est atteint. L'objectif de l'apprenti que nous cherchons à formaliser est d'atteindre un état dans lequel l'ensemble de ses connaissances (données, hypothèses et méthodes) sont stables.

Cet article présente préalablement un état de l'art succint des principaux thèmes liés à notre objet d'étude. Puis nous énoncerons la formalisation du schéma mental : son principe de représentation de l'incertitude et ses mécanismes de formulation d'hypothèses et de construction de preuves. Enfin nous présenterons la méthode de gestion de la croissance incrémentale de la connaissance dans un schéma mental, inspirée de la théorie du noyau central issue des recherches en psychologie sociale.

2 APPRENDRE ET REPRÉSENTER

Le processus de classification d'entités observées est une tâche complexe qui précède souvent le développement d'une théorie sur ces entités. Ce processus est une forme d'apprentissage à partir d'observations, c'est-à-dire un apprentissage non supervisé. Les méthodes les plus classiques calculent des mesures de similitude entre entités et les classes composées sont des ensembles d'objets avec une forte similarité intra-classe et une faible similarité inter-classes. Ces méthodes supposent que tous les attributs pertinents décrivant les entités sont connus *a priori* et qu'ils sont présents en nombre suffisant pour créer une classification. De plus, elles ne prennent pas en considération la connaissance (*background knowledge*) qui porte sur les relations entre des attributs ou des concepts généraux qui pourraient être utilisés pour caractériser des configurations d'objets. En particulier, les attributs sont considérés avec une importance égale, aucune distinction n'est faite relativement à leur degré de pertinence au regard du problème traité. Ainsi, s'il existe un accord entre les valeurs d'un nombre suffisamment grand d'attributs non pertinents, des objets sensiblement différents peuvent être classés comme étant similaires.

Les travaux autour de la notion de regroupement en concepts (*conceptual clustering*) tentent de pallier ces

insuffisances [20 ; 21 ; 28 ; 7 ; 11 ; 8]. Ces approches font l'hypothèse que les entités peuvent être arrangées en classes représentant des concepts simples plutôt qu'en classes basées uniquement sur une mesure de similitude.

Lorsque plusieurs classifications sont possibles, le choix de l'une parmi d'autres peut être orienté en désignant le *but* de la classification. Ceci définit une sorte de point de vue qui guide le processus de formation de concepts en désignant les attributs significatifs pour un but donné et en décomposant ce but en une succession de sous-buts subordonnés à la vérification de ces attributs. Stepp et Michalski [28] construisent ainsi un graphe de dépendance de buts (*Goal Dependency Network*, noté GDN). Les connaissances générales, contraintes et critères généraux, sont utilisées pour guider une classification. Les connaissances spécifiques au domaine, règles d'inférence appliquées sur le GDN, sont utilisées pour inférer les attributs significatifs pour un but donné. Le GDN permet donc de générer des attributs dérivés liés au but de la classification. Plusieurs buts peuvent être présents dans un GDN, c'est alors la précédence d'un but sur un autre qui guidera la classification.
Ce principe est illustré sur un exemple de classification de trains décrits à partir d'attributs tels que la couleur des roues, le nombre de voitures, la forme des voitures, la forme des objets transportés ou leur type : produit toxique ou non. Si la classification est guidée par la simplicité des formes géométriques les groupes formés se distinguent par des expressions du type "*les trains de cette classe sont composés de quatre voitures*" ou "*les trains de cette classe ont tous les roues blanche*". Si le but est de survivre et qu'un graphe de dépendance de buts lie le fait de survivre à la surveillance du transport de produits toxiques, la classification sera guidée par un attribut significatif différent qui décrit si une voiture transporte des produits toxiques ou non.

Cet exemple illustre assez bien le fait qu'un ensemble de descriptions contient des informations de nature différentes qui ne vont pas toutes dans le sens de l'illustration d'un concept unique. Ainsi, dans cet exemple, certains attributs illustrent la notion de survie, d'autres (mais non nécessairement un ensemble disjoint du premier) illustrent la forme. Il semble alors raisonnable de convenir d'apprendre l'un ou l'autre de ces deux concepts, et non les deux simultanément, chacun émettant du "bruit" dans la description de l'autre. En effet, tenter de déterminer des régularités sur des attributs avec lesquels aucune relation ne prend sens (*i.e.* ces attributs ne sont pas liés par une relation d'implication, de causalité, de hiérarchie, d'équivalence, de ressemblance, d'antagonisme...) est une entreprise aléatoire. Ce choix s'impose lorsque les problèmes de révision de connaissance sont abordés et qu'il devient nécessaire de maintenir une hypothèse contre une autre. Connaître le but de la classification aide alors à la révision.

Nous montrerons dans la partie 4 comment nous isolons, dans notre modèle de représentation, un *noyau* de connaissances stable, but hypothétique du processus d'apprentissage, et comment cette hypothèse facilite la révision des connaissances.

2.1 Raisonnement non monotone

L'utilisation d'une règle d'inférence pour produire une conclusion à partir d'une base de connaissances dépend implicitement d'une propriété générale des logiques classiques (incluant la logique propositionnelle et la logiques du premier ordre) que l'on appelle la *monotonie*. Une logique est monotone si, lorsque nous ajoutons un nouvel élément à une base de connaissances, les propositions originales de la première base sont présentes dans la nouvelle. Soit, pour deux ensembles de formules T et T' et une formule F,

$$\text{si } T \subset T' \text{ et } T \models F, \text{ alors } T' \models F.$$

Cette propriété n'est plus vérifiée en *raisonnement non monotone*. Ceci signifie que lorsque de nouvelles informations apparaissent, des conclusions précédentes peuvent être rétractées. Supposons par exemple qu'il soit demandé, dans un premier temps, de compléter la suite *1,3,5,7...* puis, dans un second temps, que l'on précise la question en complétant la suite par *1,3,5,7,11...* Nous constatons que l'ajout d'une information peut invalider une conclusion obtenue précédemment et pourtant acceptable.

Ce thème intéresse les chercheurs en IA pour diverses raisons :

- La plupart des raisonnements humains qui impliquent un raisonnement inductif ont cette propriété de non monotonicité.

- La complexité de la tâche de raisonnement peut être réduite en utilisant une base de connaissances restreinte avec des règles d'inférence plus efficaces, mais non monotones.

- Les logiques classiques imposent à leurs résultats d'être *certains*, et ceci s'avère souvent excessif. Il est souvent possible de se satisfaire d'une réponse *raisonnable*, mais incertaine, qui peut donc être contredite.

Nous cherchons une solution intermédiaire entre l'expression logique d'une théorie et son expression par un ensemble de descriptions. Les théories semi-empiriques sont une conceptualisation de la connaissance indépendante du langage. La connaissance de l'apprenti est empirique. Nous souhaitons exprimer des règles obtenues par une inférence guidée par les données et représentatives des parts d'exemples et de contre-exemples observées. Nous jugeons donc l'évaluation d'une expression par une croyance plus adaptée à la forme de représentation exigée. Les problèmes rencontrés demeurent cependant : comment représenter des connaissances contradictoires ou exceptionnelles, et comment, suivant quels critères, isoler un sous-ensemble de la base de connaissances afin de minimiser les opérations de révision ?

2.2 Révision de théorie

Le problème de la construction d'une base de connaissances (acquisition de connaissances) peut être vu comme un processus décomposé en deux phases : dans la première phase il est construit un modèle initial qui est révisé pour devenir une base de connaissances plus performante, au cours d'une seconde phase. Au cours de l'utilisation de cette base de connaissances, la modification dynamique de l'environnement peut provoquer son invalidité pour les raisons possibles suivantes :

- De nouveaux développement peuvent conduire à de nouveaux problèmes non couvert par la base de connaissances. Le système ne peut donc pas résoudre ces problèmes.

- Des connaissances intégrées à la base peuvent devenir obsolètes et ne doivent plus être utilisées pour calculer des solutions dans la mesure où elles ne sont plus valides dans l'environnement courant.

Dans la première situation, nous avons un nouveau cas d'application (un exemple positif) qui ne peut pas être dérivé par la base de connaissances. Dans le second cas il est possible de dériver une solution de la base de connaissances qui n'est pas admissible, par les nouvelles lois de l'environnement. Ceci est donc appelé, par conséquence, un exemple négatif.

D'un point de vue plus formel, ceci signifie qu'une base de connaissances B doit être révisée en utilisant les exemples positifs E^+ et/ou négatifs E^- de sorte que tous les exemples positifs et aucun des exemples négatifs soient couverts par la base de connaissances résultat B'.

Prenons une base de connaissances définie comme une théorie de Horn $T = F \cup R$ composée de faits F et de règles R satisfaisant un ensemble de contraintes d'intégrité CI, la *tâche d'exploration* de la révision de théorie est de transformer T en T' telle que

$$\begin{aligned} & T' \models e, \forall e \in E^+ \\ \text{et} \quad & T' \not\models e, \forall e \in E^- \end{aligned}$$

La théorie résultante T' doit, bien sûr, satisfaire les contraintes d'intégrité, i.e. $CI \cup T'$ doit être consistant. Cette vérification d'intégrité représente la *tâche de vérification* de la révision de théorie.

La tâche principale demeure cependant comment obtenir la théorie révisée. En principe, il existe deux possibilités

- premièrement, nous pouvons modifier les règles, par exemple en utilisant des techniques de généralisation ou de spécialisation

- ou nous pouvons étendre l'ensemble des faits, ou les faits additionnels peuvent être trouvés par abduction.

FORTE [24] est un système appliquant une recherche "ascendante" dans un espace d'opérateurs de spécialisation ou de généralisation afin de déterminer une révision minimale d'une théorie, en l'occurrence une base de connaissances décrite sous forme d'un programme Prolog sans fonction, qui reste consistante avec un ensemble d'apprentissage. Le processus général de révision identifie, à chaque itération, des points

dans la théorie, appelés points de révision, qui peuvent être des sources d'erreurs, et donc pour lesquels une révision a le potentiel d'augmenter la justesse de la théorie. Un ensemble de révisions est généré, basé sur ces points de révision, le meilleur d'entre eux est sélectionné et exécuté. Ce processus est exécuté jusqu'à ce qu'aucune révision n'améliore la théorie.

Il est à noter que Richards et Mooney définissent dans FORTE un module désigné sous l'appellation de *"théorie fondamentale du domaine"* qui contient les prédicats que l'utilisateur souhaite écarter du processus de révision. Deux raisons peuvent conduire à un tel choix. Premièrement, placer les éléments de la théorie qui sont connus comme étant corrects dans la théorie fondamentale du domaine permet de réduire l'espace de révisions. Deuxièmement, la théorie fondamentale du domaine peut fournir les définitions intensionnelles des relations fondamentales utilisées pour définir un domaine. Ce module a une fonction similaire à ce que nous définirons comme le *noyau* de la base de connaissances dans la section 4.

2.3 Gestion de l'incertitude

L'*inférence* est une action qui permet à un humain ou une machine d'accroître ses connaissances. Cette personne ou cette machine "fait une inférence", c'est-à-dire qu'elle infère un résultat à partir d'un ensemble de données. Dans une situation où un agent, humain ou programmé, raisonne, l'incertitude du raisonnement peut être attachée à l'inférence appliquée, ses prémisses, sa conclusion, ou ces trois éléments à la fois. Ainsi, face à une information, la croyance apportée aux conclusions engendrées par cette information et l'application d'une règle d'inférence peut être influencée par l'incertitude que l'on accorde à l'information elle-même ou à l'inférence appliquée dans une telle situation, déduction ou induction. Ces deux possibilités sont illustrées par la notation suggérée par Hempel [12] et complétée par Kyburg dans [15; 16] avec des prémisses d'incertitude:

Schéma I

$$\frac{\begin{array}{c}\textit{Prémisse d'incertitude}_1 \\ \textit{Prémisse}_2\end{array}}{\textit{P(Conclusion)=r}}$$

Ce premier schéma représente une inférence qui contient deux prémisses, dont une exprime une incertitude (probabilité, croyance ...), et dont la conclusion est un énoncé exprimant explicitement une incertitude.

$$\frac{\begin{array}{c}\textit{La probabilité que les éléphants savent nager est } [0.2,0.3] \\ \textit{Si les éléphants savent nager, ils seront sauvés}\end{array}}{\textit{P(Les éléphants seront sauvés)=}[0.2,1.0]}$$

L'inférence est strictement déductive, la conclusion ne peut être vraie que si les prémisses sont vraies. Le deuxième schéma retenu par Hempel est le suivant:

Schéma II

$$\frac{\begin{array}{c}\textit{Prémisse d'incertitude}_1 \\ \textit{Prémisse}_2\end{array}}{\textit{Conclusion}}\ p$$

Ce schéma représente une inférence dans laquelle la conclusion est catégorique, mais dont la règle d'inférence n'est pas simplement déductive. Le paramètre qualificatif p caractérise cette fois la règle d'inférence et non plus la conclusion: p représente la valeur de probabilité minimale pour exécuter cette inférence.

$$\frac{\begin{array}{c}\textit{La probabilité que les éléphants savent nager est } 0.97 \\ \textit{Si les éléphants savent nager, ils seront sauvés}\end{array}}{\textit{Les éléphants seront sauvés}}\ 0.95$$

La première prémisse de cette inférence contient la connaissance de l'incertitude, la seconde contient la connaissance de base, catégorique, non compromise par l'incertitude, tant que le contexte de son argument lui convient. Dans ce cas, l'inférence n'est plus simplement déductive, il n'y a plus préservation de la vérité des prémisses. Ce pourrait être une inférence inductive, probabiliste, statistique ou une inférence par analogie. La conclusion de cette inférence n'est plus un énoncé exprimant l'incertitude d'un

autre énoncé, mais simplement ce dernier énoncé, sous forme de conclusion catégorique. Dans l'exemple, la conclusion n'est pas que les éléphants seront *probablement* sauvés, mais qu'ils *seront* sauvés. Ceci ne signifie pas que la règle d'inférence est correcte, en particulier elle pourrait être remise en cause par des événements plus récents, mais ceci impliquerait une autre inférence.

Notre modélisation tente de considérer ces deux types de représentation de l'incertitude, à travers des valeurs de *croyance* et la définition de règles d'inférences hypothétiques à partir de prémisses non certaines. Nous représentons l'incertitude par des valeurs dans un ensemble de croyances, défini comme un *bitreillis* au sens de Ginsberg.

2.4 La logique multi-valuée de Ginsberg

Une croyance est une forme de valuation de la connaissance. Ce terme désigne une extension de la notion de valeur de vérité attribuée à une proposition. L'idée de Ginsberg [9] est de définir, sur un ensemble de valeurs désignées sous le terme de *croyances*, deux relations d'ordre permettant d'exprimer le fait qu'une croyance est *moins vraie* ($\leq_T$) ou *moins connue* ($\leq_K$) qu'une autre. Les valeurs *faux* (F) et *vrai* (V) désignent les bornes inférieures et supérieures de la relation $\leq_T$, *silence* (S) et *contradictoire* (C) celles de la relation $\leq_K$.
Nous souhaiterions de plus définir une opération de négation sur cet ensemble de sorte que si $a \geq_T b$ alors $\neg a \leq_T \neg b$ (la négation inverse le sens de la relation de vérité) mais aussi que si $a \geq_K b$ alors $\neg a \geq_K \neg b$ (si nous en savons moins sur b que sur a, nous en savons également moins sur la négation de b que sur la négation de a). Enfin nous souhaitons la négation idempotente : $\neg\neg a = a$.
Un tel ensemble définit un bitreillis. Ginsberg énonce cette définition à partir des opérateurs, lesquels permettent de définir les relations d'ordre.

DÉFINITION 2.1 (Ginsberg) Un *bitreillis* est un sextuplet $(B, \wedge, \vee, \cdot, +, \neg)$ tel que :

1. $(B, \wedge, \vee)$ et $(B, \cdot, +)$ sont tous les deux des treillis
2. $\neg$: B $\rightarrow$ B est définie par :
   (a) $\neg^2 = 1$
   (b) $\neg$ est un homomorphisme de treillis de $(B, \wedge, \vee)$ dans $(B, \vee, \wedge)$ et de $(B, \cdot, +)$ dans lui-même.

Le bitreillis non trivial minimal au sens de Ginsberg est réduit aux bornes inférieures et supérieures des deux relations d'ordre distinctes $\leq_T$ et $\leq_K$, soit $\{F, V, S, C\}$. Celles-ci sont liées aux opérations du bitreillis de la façon suivante :

1. $x \leq_T y \Leftrightarrow x \vee y = y$ et $x \wedge y = x$
2. $x \leq_K y \Leftrightarrow x + y = y$ et $x \cdot y = x$

Il est toujours possible de construire un bitreillis à partir du produit cartésien d'un treillis avec lui-même [9]. Ainsi, le bitreillis minimal peut être défini par le produit du treillis ({0,1},$\wedge$,$\vee$) avec lui-même :

- $(x, y) \vee (z, t) = (x \vee y, z \wedge t)$, et $(x, y) \wedge (z, t) = (x \wedge y, z \vee t)$
- $(x, y) + (z, t) = (x \vee y, z \vee t)$, et $(x, y) \cdot (z, t) = (x \wedge y, z \wedge t)$
- $\neg(x, y) = (y, x)$

2.4.1 Clôture

M.L. Ginsberg définit un mécanisme d'inférence utilisant ces valeurs de vérité sur son modèle représentationnel, la *clôture*, qui construit, à partir d'un ensemble de propositions, toutes les conséquences de ces propositions.
L'interprétation d'un ensemble de formules bien formées est définie classiquement comme une fonction dans un bitreillis. Pour une interprétation ϕ et une proposition p, ϕ contient une information explicite sur p dans la valeur de vérité $\phi(p)$, mais aussi des informations implicites sous la forme de valeurs de vérité sur des expressions logiquement liées à p. Ces deux types d'information doivent être obtenus à travers l'opération de clôture qui produit une nouvelle interprétation cl(ϕ) qui détient les informations explicites

et implicites sur p dans une nouvelle valeur $\mathrm{cl}(\phi)(p)$. Nous attendons de l'opération de clôture qu'elle ait les propriétés suivantes :

1. Elle doit être une construction de bitreillis, dépendant uniquement des ordres et de la négation définissant le bitreillis original.
2. $\mathrm{cl}(\mathrm{cl}(\phi)) = \mathrm{cl}(\phi)$, pour tout ϕ.
3. $\mathrm{cl}(\phi)(p) \geq_K \phi(p)$, pour tout p.

La notion d'inférence doit donc incorporer l'information dérivée de la structure du bitreillis mais ne doit pas faire appel à des notions ou des techniques dépendant du domaine. Les deux dernières conditions précisent que l'opération de clôture ne nécessite d'être réalisée qu'une seule fois afin d'extraire les informations désignées comme implicites données par ϕ, et que si ϕ attribue la valeur x à une proposition p, la clôture de ϕ ne modifie en rien cette information explicite. Ginsberg définit une méthode de calcul itératif de la clôture d'une interprétation pour une expression p en déterminant un ensemble de valeurs x_i correspondant aux dérivations de p à partir d'autres énoncés du langage, de sorte que $\mathrm{cl}(\phi)(p) \geq_K x_i$ pour tout i. Il s'agit en fait de déterminer les x_i tels que

$$\mathrm{cl}(\phi)(p) = \sum_i x_i$$

2.4.2 Le bitreillis des mondes possibles

Une classe intéressante de bitreillis peut être obtenue à partir d'une collection de *mondes possibles*. Supposons que W est un ensemble de mondes, et que (U, V) est un couple de sous-ensembles de W. Dire qu'une proposition p a une valeur de croyance (U, V) signifie que p est fausse dans les mondes de U et vraie dans les mondes de V. Il n'est pas exigé que l'ensemble $U \cap V$ soit vide, un monde w appartenant à cet ensemble précisant que p est à la fois vraie et fausse dans w. Les relations d'ordre peuvent être définies de la façon suivante

$$(U, V) \leq_T (S, T) \text{ ssi } U \supseteq S \text{ et } V \subseteq T$$

une proposition p est moins vraie qu'une proposition q si p est fausse au moins dans les mondes où q est fausse et si q est vraie au moins dans les mondes où p est vraie ; et

$$(U, V) \leq_K (S, T) \text{ ssi } U \subseteq S \text{ et } V \subseteq T$$

une proposition q est plus connue qu'une proposition p si q est connue comme vraie (resp. fausse) au moins lorsque p est connue comme vraie (resp. fausse).

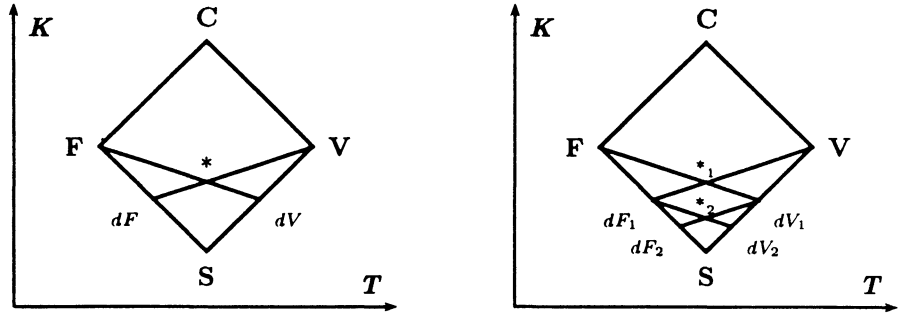
Les bornes des relations d'ordre sont données respectivement par $F = (W, \varnothing)$, $V = (\varnothing, W)$, $S = (\varnothing, \varnothing)$ et $C = (W, W)$. Et les opérateurs $\vee$, $\wedge$, $+$, et $\cdot$ sont définis par :

- $(U, V) \wedge (S, T) = (U \cup S, V \cap T)$
- $(U, V) \vee (S, T) = (U \cap S, V \cup T)$
- $(U, V) \cdot (S, T) = (U \cap S, V \cap T)$
- $(U, V) + (S, T) = (U \cup S, V \cup T)$
- $\neg(U, V) = (V, U)$

L'ensemble de tous les couples de sous-ensembles de W étant désigné par B_W, soit $B_W = \mathcal{P}(W) \times \mathcal{P}(W)$, pour tout ensemble W, $(B_W, \wedge, \vee, \cdot, +, \neg)$ est un bitreillis.

2.4.3 Le bitreillis de la logique des défauts

Un autre exemple de bitreillis est représenté sur la figure 1(a) ; il permet le traitement de la logique des défauts. En plus des valeurs F, V, S, et C, une proposition peut être évaluée comme *fausse par défaut* (dF) ou *vraie par défaut* (dV). La valeur $* = dF + dV$ désigne une proposition qui est à la fois vraie et fausse par défaut. Ce bitreillis est obtenu par une expansion du point S en un autre bitreillis reflétant



(a) Le bitreillis de la logique des défauts

(b) Le bitreillis de la logique des défauts prioritaires

FIG. 1 – *Exemples de bitreillis pour des logiques de défauts*

l'existence d'un autre ordre de croyance. Ici encore, la négation correspond à une symétrie selon l'axe (S, C).

Pour le traitement du célèbre exemple de Tweety, le manchot non volant, Ginsberg pose que notre croyance dans le fait que Tweety vole provient de la conséquence de la règle par défaut *les oiseaux volent*. En attribuant à cette règle la valeur de croyance dV, et aux conclusions de celle-ci, une valeur de croyance jamais plus connue que dV, la contradiction engendrée par le nouveau fait précisant que Tweety ne vole pas est levée car cette proposition est alors plus connue que la conclusion de la règle.

Par expansion de la valeur S du bitreillis précédent en un autre bitreillis, nous pouvons également définir un bitreillis pour la logique des défauts prioritaires (figure 1(b)) désignant un défaut du second ordre. Cette idée peut être étendue de façon à définir une hiérarchie infinie de défauts, discrets ou continus.

Nous pouvons noter que ces deux bitreillis partagent les valeurs F, V, S, et C avec le premier. En fait, tout bitreillis possédera quatre éléments distincts correspondant aux bornes inférieures et supérieures des deux relations d'ordre.

2.4.4 Éléments fondamentaux

Une notion intéressante introduite par Ginsberg est celle d'éléments *fondamentaux* (grounded). Elle provient du fait qu'il existe, dans un bitreillis, des éléments qui correspondent à des informations "primitives". Dans la perspective de construire la clôture d'une interprétation arbitraire en accumulant les informations de sources différentes, c'est une notion qu'il semble intéressant de formaliser.

Considérons le bitreillis de la logique des défauts, représenté sur la figure 1(a). On peut aisément imaginer, à partir d'un ensemble d'exemples, qu'une proposition soit vraie ou fausse, ou même vraie par défaut ou fausse par défaut. Il est plus difficile d'imaginer qu'une proposition ait pour valeur de croyance * ou *contradictoire*. Ces deux points se distinguent intuitivement des autres dans le bitreillis par le fait qu'ils ne sont pas situé sur la "bordure inférieure" du bitreillis. Cette "bordure inférieure" est en fait divisée en deux parties : celle joignant S à F et celle joignant S à V. Nous dirons d'un élément x situé entre S et V qu'il est *fondamentalement* V, et qu'il est *fondamentalement* F s'il est situé entre S et F.

DÉFINITION 2.2 (Ginsberg) Un élément x d'un bitreillis B est *fondamentalement* V si, pour tout $y \in B$, $y \geq_T x \Rightarrow y \geq_K x$.
Il sera *fondamentalement* F si, pour tout $y \in B$, $y \leq_T x \Rightarrow y \geq_K x$.
x est *fondamental* s'il est à la fois fondamentalement V et fondamentalement F.

LEMME 2.1 Pour un bitreillis B, $x \in B$ est fondamental ssi $x = \inf_K(B)$.

Supposons en effet l'existence d'un élément $x \in B$ différent de $S = \inf_K(B)$ tel que x soit fondamental dans B. Alors quelle que soit la position de x par rapport à S selon l'ordre $\leq_T$ ($S \geq_T x$ ou $S \leq_T x$), la relation $S \geq_K x$ doit être vérifiée. Ceci impose, par définition de S, $x = S$.
Le silence est donc l'unique élément fondamental de tout bitreillis.

Nous avons présenté les références les plus importantes, les plus proches des thèmes que nous abordons dans ce travail. À savoir : l'expression de régularités observées sur une base de connaissances, l'expression d'une valeur de croyance permettant la description des éléments de la base et leur comparaison en terme de connaissance et de vérité. Enfin nous avons présenté certains des travaux en raisonnement non monotone et en révision de théorie. Chacun de ces thèmes correspond à un des aspects de ce travail. Nous cherchons en effet à formaliser un modèle d'acquisition et de représentation de connaissances, le problème de l'évaluation de la connaissance sera donc notre premier souci avant d'évoquer les différents problèmes liés à la découverte de régularités et la construction d'hypothèses dans une base de connaissances.

3 LES THÉORIES SEMI-EMPIRIQUES : UNE FORMALISATION

Les théories semi-empiriques (TSE) [25 ; 26] sont une conceptualisation de la connaissance indépendante du langage. Elles précisent la manière dont une connaissance se formule, s'expérimente et se diffuse. La figure 2 présente une taxinomie des termes employés pour exprimer la connaissance en TSE. Cette taxinomie est basée sur le travail de Addis [2]. Les connaissances s'organisent en données, mécanismes et méthodes. Les *données* constituent la connaissance. Les *mécanismes* sont des méthodes de base : l'abduction crée des données, l'induction les organise et la déduction propage les contraintes. Enfin les *méthodes* sont liées aux interactions avec un agent externe capable de produire des critiques qui permettent d'examiner la pertinence des expressions *être une objection*, *être une conjecture*, *être un lemme* ou *être une preuve*.

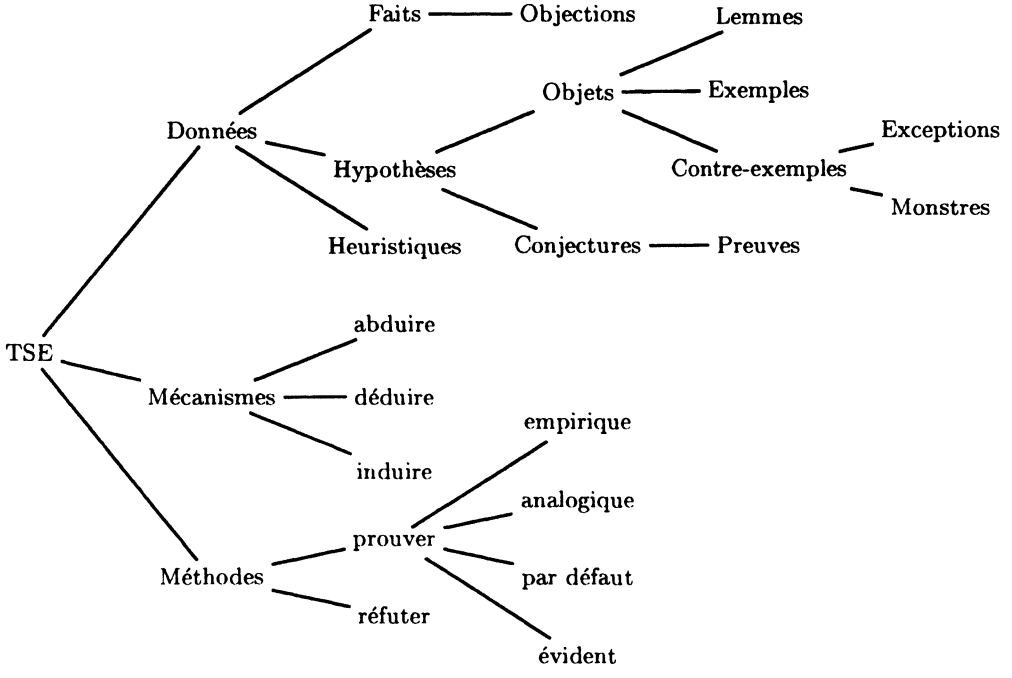


FIG. 2 – *Taxinomie des termes intervenant dans la formalisation et l'évolution des connaissances pour les TSE.*

Un *fait* est un élément d'une description, c'est-à-dire une relation entre deux énoncés A et X du type A *possède la propriété* X. Les TSE doivent permettre à un apprenti d'émettre des hypothèses sur des faits, illustrées par des exemples et des contre-exemples produits par un oracle, et qui forment l'échantillon (ou connaissances de base) de l'apprenti.
Les énoncés d'une Théorie Semi-Empirique, éléments d'un langage dénotant une description ou un attribut, se distinguent selon les interprétations qui leurs sont associées. Un énoncé est un *objet* quand on peut lui associer un ensemble de faits. Un énoncé est un *concept* quand on peut lui associer un ensemble

d'exemples (ou descriptions).

L'apprenti dispose des méthodes d'abduction et d'induction pour produire des hypothèses (ou *règles*) et des faits. En posant l'hypothèse que la connaissance exploitée par l'apprenti est productrice d'erreurs, il faut lui fournir les éléments nécessaires à son contrôle et la modification de ses connaissances : un énoncé construit par l'apprenti et reconnu par l'expert comme caractérisant les exemples d'un concept est nommé un *lemme* ; un énoncé produit par l'apprenti et admis par l'expert comme pouvant correspondre à un concept est une *preuve*. La *réfutation* d'un lemme est la production d'un *contre-exemple*. Une *objection* est un énoncé dont la vérification par un objet exprime une différence entre l'objet objecté et un exemple satisfaisant la conjecture.
L'apprenti utilise la preuve pour prédire si un objet est un exemple d'une conjecture. Un objet qui critique la preuve est aussi un contre-exemple de la conjecture. Quand l'apprenti reconnaît l'objet étudié comme un contre-exemple, il propose une objection vérifiée par cet objet et soumise à la critique de l'usager [4].

Notre proposition est de définir une structure appelée *schéma mental*, formalisant les relations entre énoncés élaborées dans une théorie semi-empirique.
Dans un schéma mental, les descriptions (exemples) sont représentées par une relation entre énoncés (des objets et des concepts) associés à une valeur dans un ensemble de *croyances*, destinées à mesurer la connaissance acquise en évaluant un fait ou une règle en terme de connaissance et de vérité.
Nous attendons donc d'un schéma mental qu'il permette d'énoncer des hypothèses soutenues par des exemples décrits et d'en établir une preuve. La connaissance dans un schéma mental progresse par exploitation des mécanismes d'abduction et d'induction.

3.1 Représenter les connaissances

Le schéma mental est donc la structure représentative des théories semi-empiriques. Il contient les descriptions des éléments observables (ou objets) suivant un ensemble d'attributs (ou concepts), à partir desquelles des régularités vont être recherchées. Il est donc la représentation d'une relation entre objets et concepts d'une base de connaissances. Cette relation peut être intuitivement interprétée comme "l'objet A possède la propriété définie par le concept X".

3.1.1 Langage

Objets et concepts sont des *énoncés*. Un énoncé est classiquement un élément d'un *langage*. Nous ajoutons à une définition minimale d'un langage un ordre partiel sur les énoncés, exprimant une relation de subsomption entre énoncés comme une connaissance a priori du domaine. Cette relation est notée $\leq_L$ et interprétée par *"moins général que"* (ou *"plus spécifique que"*.
Les énoncés dits "atomiques" sont des chaînes de caractères. Il est possible de construire des énoncés composés par la disjonction ou la conjonction d'énoncés (appelés énoncés composants). La composition engendre bien évidemment une relation de subsomption entre les énoncés composés et les énoncés composants : si e_1 et e_2 sont deux énoncés de L, alors

- $e_1 \vee e_2$ est un énoncé de L, et on a les relations $e_1 \leq_L e_1 \vee e_2$ et $e_2 \leq_L e_1 \vee e_2$.
- $e_1 \wedge e_2$ est un énoncé de L, et on a les relations $e_1 \geq_L e_1 \wedge e_2$ et $e_2 \geq_L e_1 \wedge e_2$.

La négation n'est pas une propriété nécessaire à la définition d'un langage, elle n'est donc pas nécessairement élément du langage. Elle peut être ajoutée au langage à la condition qu'elle vérifie la propriété suivante : si $a \leq_L b$, alors $\neg b \leq_L \neg a$
La négation définit alors un automorphisme qui inverse la relation d'ordre. En effet, une propriété générale de la négation est qu'elle inverse le sens de la relation de spécificité entre deux énoncés : en interprétant cette relation par un lien *est-un* entre énoncés, alors, par exemple, si la couleur *or clair* est un *or*, *non-or* est un *non-or clair* (et non l'inverse).

Exemple 3.1 La figure 3 illustre un exemple de langage arborescent qui définit donc un sup-demi treillis. On peut y voir (par exemple) que l'énoncé *or* est plus spécifique que l'énoncé *couleur* (soit *or* $\leq_L$ *couleur*).

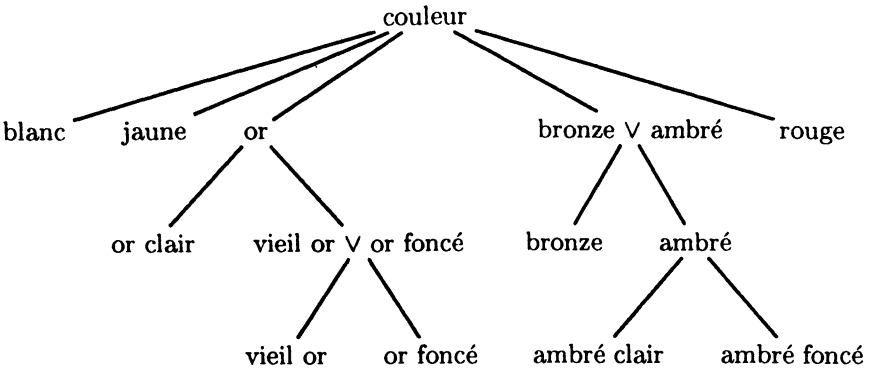


FIG. 3 – *Exemple de langage désignant des couleurs*

3.1.2 Relation de correspondance

Une ***relation de correspondance*** δ est une relation entre deux ensembles d'énoncés. Intuitivement, elle est proche de relations du type "est-un", relation d'héritage, lien catégorie-instance ou général-spécifique.

DÉFINITION 3.1 Une *correspondance* est composée de deux énoncés liés par la relation de correspondance. On note $\delta(x, y)$ la correspondance entre le ***sujet*** x et le ***prédicat*** y.

Dans la théorie définie à partir d'un langage et de la seule relation de correspondance δ nous définissons trois règles d'inférence calculant la fermeture réflexive et transitive δ^* de $\delta \cup \leq_L$ (ensemble de conclusions) :

1. si $\delta(x, y)$ ajouter $\delta^*(x, y)$, $\delta^*(x, x)$ et $\delta^*(y, y)$
2. si $x \leq_L y$ ajouter $\delta^*(x, y)$
3. si $\delta^*(x, y)$ et $\delta^*(y, z)$ ajouter $\delta^*(x, z)$

La dernière étape doit être répétée autant de fois que nécessaire, jusqu'à ce que toutes les possibilités aient été explorées.

Pour expliciter comment un énoncé particulier x est lié aux autres, nous définissons l'*extension* et l'*intension* de x dans δ^*.

DÉFINITION 3.2 L'*extension* d'un énoncé x est l'ensemble E_x des énoncés e tels que $\delta^*(e, x)$. L'*intension* d'un énoncé x est l'ensemble I_x des énoncés e tels que $\delta^*(x, e)$.

Cette définition est d'une grande importance pour notre formalisation. Classiquement l'extension et l'intension font référence à deux différents aspects d'un énoncé : ses *instances* et ses *attributs*. L'extension d'un énoncé est généralement définie comme un ensemble d'*objets* qui sont décrits par cet énoncé ; l'intension d'un énoncé est généralement définie comme un *concept* d'un monde imaginaire qui décrit les énoncés. À quelques petites différences près, l'utilisation de ces deux termes dans la littérature fait toujours référence à une relation entre un énoncé du langage et quelque chose d'extérieur au langage. Dans notre formalisation, ils sont définis à partir d'une relation entre deux énoncés du langage, et font donc référence aux instances et aux attributs d'un énoncé.

Cette définition implique les propriétés suivantes :

1. Chaque correspondance précise l'intension de l'énoncé-sujet et l'extension de l'énoncé-prédicat.
2. Puisque la relation de correspondance est réflexive, tout énoncé a une intension et une extension non vide (chacune contenant au moins l'énoncé lui-même).

Et nous pouvons facilement montrer le résultat suivant: $\delta^*(x,y) \Leftrightarrow E_x \subseteq E_y$ et $I_y \subseteq I_x$.
Preuve

1. Si on a $\delta^*(e,x)$ (e est dans l'extension de x) et $\delta^*(x,y)$, alors $\delta^*(e,y)$ (e est dans l'extension de y). De même, si $\delta^*(x,y)$ et $\delta^*(y,e)$ (e est dans l'intension de x), alors $\delta^*(x,e)$ (e est dans l'intension de x).
2. Soit, $\delta^*(e,x) \rightarrow \delta^*(e,y)$ ($E_x \subseteq E_y$), pour tout énoncé e. Alors pour $e = x$, puisque on a $\delta^*(x,x)$ alors $\delta^*(x,y)$. De même, si, pour tout énoncé e, si $\delta^*(y,e)$ alors $\delta^*(x,e)$ ($I_y \subseteq I_x$), alors, puisque on a $\delta^*(y,y)$, $\delta^*(x,y)$.

En d'autres termes, s'il existe une relation de correspondance entre x et y, alors y "hérite" de l'extension de x et x "hérite" de l'intension de y. Intuitivement cette relation indique que, dans un certain sens, un énoncé peut être utilisé à la place de l'autre. Si le système connaît $\delta^*(x,y)$, alors x peut remplacer y dans toutes les expressions du type $\delta^*(y,e)$ et y peut remplacer x dans toutes les expressions du type $\delta^*(e,x)$. Inversement, si tous les énoncés e qui vérifient $\delta^*(e,x)$ vérifient aussi $\delta^*(e,y)$, ou si tous les énoncés e qui vérifient $\delta^*(y,e)$ vérifient aussi $\delta^*(x,e)$, alors nous avons $\delta^*(x,y)$.

D'après leurs extensions, nous distinguons deux sortes d'énoncés dans un langage: les *objets* et les *concepts*. Plus précisément, l'ensemble des énoncés d'un schéma mental est partitionné en deux sous-ensembles: l'ensemble des objets et l'ensemble des concepts. Un objet correspond à ce que nous pouvons couramment appeler une "description" ou un "objet réel". Il est le résultat d'une observation représentée par un ensemble d'attributs (l'intension de l'énoncé désignateur de cet objet), mais n'est jamais lui-même un élément de description d'un autre énoncé (son extension est réduite à lui-même).

DÉFINITION 3.3 Soit un énoncé x. On dit que x est un *objet mental* ssi $E_x = \{x\}$ (i.e. $\delta^*(e,x) \Leftrightarrow e = x$). On note alors $O_x(y) = \delta^*(x,y)$.
L_Ω est l'ensemble des objets mentaux.

Le concept mental est la notion antagoniste de celle d'objet mental: un énoncé non désignateur d'objet et qui apparaît dans δ^* est un concept mental, élément de description d'un objet existant.

DÉFINITION 3.4 Soit un énoncé x. On dit que x est un *concept mental* ssi $|E_x| > 1$ (i.e. $\exists e \neq x$ tel que $\delta^*(e,x)$).
L_χ est l'ensemble des concepts mentaux.

Cette partition de l'ensemble des énoncés d'un schéma mental nous permet de distinguer deux types de correspondances: les *faits* et les *règles*:

- Un fait est une correspondance $\delta^*(x,y)$ entre un objet et un concept (i.e. $x \in L_\Omega$ et $y \in L_\chi$).
- Une règle est une correspondance $\delta^*(x,y)$ entre deux concepts (i.e. $x,y \in L_\chi$).

Relativement à la propriété d'inclusion des extensions et des intensions énoncée précédemment, nous pouvons remarquer que si $x \in L_\Omega$ alors pour toute correspondance $O_x(y)$, x hérite de l'intension de y, c'est-à-dire que x possède les propriétés dérivées de y, et y (désignant alors un concept) contient x parmi ses instances.

Exemple 3.2 Voici un exemple de correspondance δ^* entre deux ensembles d'énoncés. Elle représente la description de zones géographiques dont le type est désigné par un entier de {0,1,3,6} pour une date donnée. Cet exemple est extrait du problème traité dans [23] décrivant un exemple d'application.
Ici les énoncés décrivant les types de zones géographiques correspondent aux relevés de l'année *1837*. L'énoncé *0?1837* décrit donc une zone de type *0* en *1837*.
Les quatre autres énoncés (*1,2,5,6*) désignent des zones géographiques.
Une correspondance entre deux énoncés est représentée par un 1 dans le tableau ci-dessous.

δ^*	0?1837	1?1837	3?1837	6?1837	3?1837 ∨ 6?1837	1	2	5	6
0?1837	1								
1?1837		1							
3?1837			1		1				
6?1837				1	1				
3?1837 ∨ 6?1837					1				
1				1	1	1			
2				1	1		1		
5			1		1			1	
6			1		1				1

Les énoncés *1,2,5,6* sont des *objets*. Leurs caractéristiques sont données suivant les *concepts 0?1837, 1?1837, 3?1837* et *6?1837*. Les correspondances du type δ^*(X,3?1837 ∨ 6?1837) sont obtenues soit à partir de l'ordre sur le langage (3?1837 $\leq_L$ 3?1837 ∨ 6?1837 et 6?1837 $\leq_L$ 3?1837 ∨ 6?1837), soit en complétant les informations par le calcul de la fermeture transitive de la relation de correspondance δ.

Nous verrons plus loin que δ^* contient également l'ensemble des *exemples* disponibles pour le système, qui est contenu ou dérivé à partir des connaissances de base.

Nous avons défini une théorie composée d'un langage, d'une relation simple entre énoncés et de règles d'inférences qui complètent les connaissances de base. Cependant cette description ne tient pas compte du fait que les connaissances du système sont insuffisantes pour observer des régularités valides. Nous souhaitons associer à la relation de corespondance une valeur de vérité représentant l'incertitude de la relation.

3.2 Représenter l'incertitude

Puisque nous avons supposé que l'apprenti disposait d'une connaissance initiale nulle sur le domaine, et que l'augmentation de la connaissance était incrémentale, nous savons que, en pratique, les résultats présentés précédemment ne sont pas utilisables. Ils ne prennent pas en considération la part d'exemples positifs ou négatifs, ni l'influence de l'ajout d'exemples futurs. Pour réaliser cela, une simple valeur de vérité binaire est insuffisante, nous avons besoin de pouvoir exprimer plus d'information à propos des exemples.

3.2.1 Schéma mental

La relation que nous utilisons dans notre système est de la forme $\Delta(x, y)$ que nous désignons sous le terme de *schéma mental*. Un schéma mental est une généralisation de la relation de correspondance δ. $\Delta(x, y)$ n'est pas simplement vrai ou faux, mais représente la part de vérité liée à la présence d'exemples positifs et négatifs, et rend compte de la part d'incertitude liée à la taille de l'échantillon utilisé.

La relation de correspondance δ représente une situation idéale, avec des valeurs de vérité binaires, et nous interprétons toute relation $\Delta(x, y)$ comme s'il existait un ensemble de correspondances $\delta^*(e_i, e_j)$ correspondant à autant d'exemples disponibles pour décrire x et y.

Levesque [18] a introduit les notions de connaissances implicites et explicites, reprises par Doyle [5] comme manifestes et constructives, pour désigner respectivement les connaissances auxquelles un agent accède directement et celles dérivées, ou obtenues, à partir des premières, comme des conclusions calculables. Ces deux types de connaissances sont aussi désignés par "assertions", "axiomes" ou "connaissances de base" pour les connaissances manifestes, et "théorèmes", connaissances "dérivées" ou "inférées" pour les connaissances constructives. Selon la théorie de Levesque, les connaissances explicites n'ont pas à être consistantes, closes par déduction ou logiquement complètes, de sorte que l'agent puisse émettre certaines incapacités observées chez l'humain. L'ensemble des connaissances d'un agent est vu comme une fonction F des connaissances manifestes M déterminant les connaissances constructives C. Selon les définitions possibles du raisonnement, il est concevable d'obtenir, à partir d'un même ensemble de connaissances manifestes, plusieurs ensembles de connaissances constructives. C est alors choisi dans l'ensemble $F(M)$ ($C \in F(M)$). Cette distinction s'applique dans notre cas à la relation de correspondance et au schéma mental. La première relation contient des informations provenant des descriptions, dont la validité n'est pas mise en doute, alors que la seconde contient des connaissances hypothétiques construites à partir de ces informations. La validité des différents résultats pouvant être obtenus sera confrontée aux nouveaux exemples ajoutés dans la relation de correspondance.

Dans un schéma mental, tout couple d'énoncés est lié à une croyance: la connaissance de "$\Delta(x, y) = c$" signifie que x est dans l'extension de y et que y est dans l'intension de x. La croyance c exprime leur degré d'appartenance.

Nous allons donc devoir définir un *ensemble de croyances* permettant de représenter et d'évaluer une correspondance en terme de vérité et de connaissance. "$\Delta(x, y)$=vrai" est identique à $\delta^*(x, y)$. Cette expression représente elle aussi la situation idéale d'une connaissance parfaitement déterminée. Par contre, un schéma mental nécessite d'exprimer la négation d'une propriété par une relation "$\Delta(x, y)$=faux",

car l'impossibilité d'attribuer ou non une propriété à un énoncé, exprimée par une croyance silencieuse "$\Delta(x, y)$ silence", prend désormais un sens.

Exemple 3.3 Reprenons l'exemple précédent. Les correspondances de δ^* sont désormais représentées par une croyance vraie (V) de Δ. Nous pouvons étendre la description des objets car nous disposons à présent d'une croyance fausse (F) illustrant la non vérification d'un concept par un objet. Les croyances *silence* ne sont pas représentées.

δ^*	0?1837	1?1837	3?1837	6?1837	3?1837 ∨ 6?1837	1	2	5	6
0?1837	V								
1?1837		V							
3?1837			V		V				
6?1837				V	V				
3?1837 ∨ 6?1837					V				
1	F	F	F	V	V	V	F	F	F
2	F	F	F	V	V	F	V	F	F
5	F	F	V	F	V	F	F	V	F
6	F	F	V	F	V	F	F	F	V

3.2.2 Ensemble de croyances

Évaluer une correspondance en termes de vérité et de connaissance nécessite la définition de deux relations d'ordre sur des croyances que nous noterons par $\leq_T$ (*moins vraie*) et $\leq_K$ (*moins connue*). Les bornes de ces deux relations sont respectivement les valeurs *vrai* (V) et *faux* (F), et *silence* (S) et *contradictoire* (C). Le silence exprime l'impossibilité d'attribuer l'absence ou la présence d'une propriété à une observation. Il se distingue de vrai et de faux au sens de la vérité (silence n'est ni vrai ni faux) mais aussi par le fait qu'il représente une connaissance "plus faible". À l'opposé, la contradiction représente le fait qu'une observation et son contraire sont présents dans la base de connaissances. Comme le silence, contradictoire n'est ni vrai, ni faux, mais il représente une information d'une connaissance "plus élevée".
Les quatre symboles *vrai*, *faux*, *silence* et *contradictoire* constituent donc les éléments de base, et les bornes, d'un ensemble de croyances valuant une relation entre deux énoncés dans un schéma mental. Autrement dit, une telle relation est exprimée par une valeur dans un ensemble de croyances contenant $\{V, F, S, C\}$. Notre définition repose sur la notion de bitreillis [9] qui associe à un ensemble les deux relations d'ordre $\leq_T$ et $\leq_K$. Le bitreillis minimal définit un *ensemble de croyances* minimal.

DÉFINITION 3.5 Un ensemble B muni des deux relations d'ordre $\leq_K$ et $\leq_T$ est un *ensemble de croyances* s'il vérifie les propriétés suivantes :

1. $\{F = \inf_T(B), V = \sup_T(B), S = \inf_K(B), C = \sup_K(B)\} \subseteq B$
2. B est un bitreillis.

(i.e. si les deux relations d'ordre sont distinctes, B est un bitreillis non trivial)

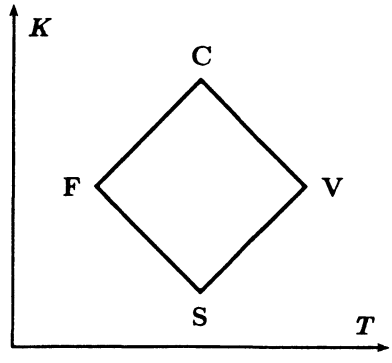


FIG. 4 – *L'ensemble de croyances construit sur le bitreillis non trivial minimal*

Après avoir jeté les bases de la définition d'un ensemble de croyances, nous allons décrire l'ensemble de croyances utilisé dans notre système. Il s'agit de pouvoir représenter des niveaux de connaissance différents afin de distinguer les exemples (énoncés liés par la relation de correspondance δ) des connaissances hypothétiques obtenues par application de mécanismes de raisonnement par le système.

L'ensemble minimal (B={F,V,S,C}) ne permet pas cette distinction. L'ensemble de croyances pour la logique des défauts prioritaires est également une réponse insuffisante à ce problème, car il ne peut représenter la part d'exemples et de contre-exemples associée à une relation entre deux énoncés.

3.2.3 Croyances numériques

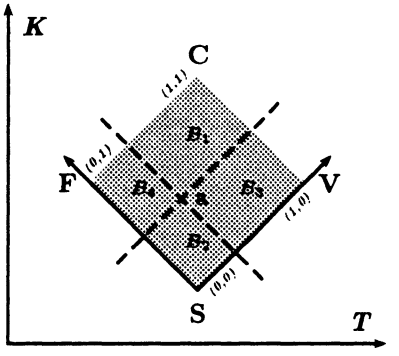
Nous proposons donc de définir un ensemble de croyances, nommé *ensemble de croyances numériques* et noté $\mathcal{B}_{\mathcal{N}}$, comme un sous-ensemble du plan $[0,1]^2$, admettant des niveaux infinis de vérité et de connaissance afin de rendre compte des proportions d'exemples et de contre-exemples à une relation ainsi que de l'incertitude qui lui est associée.

$$\mathcal{B}_{\mathcal{N}} \subseteq [0,1]^2 \tag{1}$$

Une croyance de $\mathcal{B}_{\mathcal{N}}$ est un point dans ce plan dont la position traduit ses valeurs de vérité et de connaissance. Pour une croyance (p,q) de $\mathcal{B}_{\mathcal{N}}$, p représente le "poids vrai" de la croyance, et q son "poids faux". La somme $p+q$ représente un "niveau de connaissance".
Les relations d'ordre $\leq_T$ et $\leq_K$ sont définies, pour deux croyances (p,q) et (p',q') par:

$$(p,q) \leq_T (p',q') \quad \text{ssi } p \leq p' \text{ et } q \geq q' \tag{2}$$
$$(p,q) \leq_K (p',q') \quad \text{ssi } p \leq p' \text{ et } q \leq q' \tag{3}$$

Les bornes de ces relations sont respectivement F=(0,1) et V=(1,0), S=(0,0) et C=(1,1).



$$\forall x \in B_1, \forall y \in B_2, x \geq_K a \geq_K y$$
$$\forall x \in B_3, \forall y \in B_4, x \geq_T a \geq_T y$$

FIG. 5 – *Le bitreillis associé à l'ensemble de croyances numériques*

La figure 5 illustre cet ensemble de croyances et les opérations suivantes, définies à partir des deux relations d'ordre.

- $(p,q) \wedge (p',q') = (\min(\{p,p'\}), \max(\{q,q'\}))$
- $(p,q) \vee (p',q') = (\max(\{p,p'\}), \min(\{q,q'\}))$
- $(p,q) \cdot (p',q') = (\min(\{p,p'\}), \min(\{q,q'\}))$
- $(p,q) + (p',q') = (\max(\{p,p'\}), \max(\{q,q'\}))$
- $\neg(p,q) = (q,p)$

Toute croyance est une borne pour une relation d'ordre et pour chacune des quatre sous-parties du plan qu'elle détermine. Ainsi p est $\inf_K(B_1)$, $\sup_K(B_2)$, $\inf_T(B_3)$ et $\sup_T(B_4)$.

PROPOSITION 3.1 $\mathcal{B}_{\mathcal{N}}$ est un ensemble de croyances.

La preuve est évidente, les opérations permettent de définir un bitreillis qui contient les bornes des deux relations d'ordre.

Évaluer une croyance numérique Les premières fonctions définies ci-après reprennent les notions introduites par Pei Wang [29], analysant quantitativement les croyances. La *fréquence* permet d'évaluer l'incertitude d'une croyance, à un ordre premier, soit une évaluation en terme de vérité. On considère que, pour une croyance $(p,q) \in \mathcal{B}_\mathcal{N}$, p représente le poids vrai de cette croyance, q son poids faux et la somme $p+q$ une évaluation de son niveau de connaissance. Une définition satisfaisante de la fréquence peut être le rapport entre p et la somme des poids $p+q$.

DÉFINITION 3.6 La *fréquence*, f, d'une croyance (p,q) de $\mathcal{B}_\mathcal{N}$ est $\frac{p}{p+q}$. En particulier : $f(\text{faux}) = 0$, $f(\text{vrai}) = 1$ et $f(\text{contradictoire}) = \frac{1}{2}$. Nous fixons $f(\text{silence}) = \frac{1}{2}$.

Cependant, il est intéressant de pouvoir estimer la stabilité relative de ce résultat, dans un environnement où la croissance de la connaissance est incrémentale. Supposons que les informations futures déplacent d'une constante k $(0 < k < 1)$ le point du plan matérialisant une croyance (p,q) de fréquence $\frac{p}{p+q}$. Si ce déplacement s'opère totalement dans le sens d'un accroissement de la vérité, la nouvelle fréquence sera $\frac{p+k}{p+q+k}$. Si, au contraire, les informations apportées vont dans le sens d'une diminution de la vérité de cette croyance, elle sera $\frac{p}{p+q+k}$. La nouvelle fréquence sera donc comprise dans l'intervalle $\left[\frac{p}{p+q+k}, \frac{p+k}{p+q+k}\right]$. La taille de cet intervalle, $\frac{k}{p+q+k}$, est une estimation de l'incertitude du jugement, et son complément à 1, $\frac{p+q}{p+q+k}$, une bonne mesure de sa *confiance* (ou stabilité).

DÉFINITION 3.7 Soit une constante k $(0 < k < 1)$, la *confiance*, c, d'une croyance (p,q) de $\mathcal{B}_\mathcal{N}$ est $\frac{p+q}{p+q+k}$. En particulier : $c(\text{faux}) = c(\text{vrai}) = \frac{1}{1+k} \simeq 1$, $c(\text{contradictoire}) = \frac{2}{2+k} \simeq 1$ et $c(\text{silence}) = 0$.

On montre facilement la propriété suivante : si $b_1 \geq_K b_2$ alors $c(b_1) \geq c(b_2)$

Enfin, il est intéressant, au regard de l'incertitude portée à une croyance, de savoir s'il est possible de considérer une croyance numérique comme représentant une valeur "approximativement vraie" ou "approximativement fausse". Cette notion est particulièrement importante pour notre travail, dans la mesure où les règles d'inférence que nous utilisons (voir section 3.3) s'appliquent sur des croyances dont la *valeur de vérité* doit être connue. Ainsi, la déduction, pour être valide, doit s'appliquer sur un fait qui doit être connu comme vrai (noté $O_A(X) \approx V$) et une règle $(\Delta(X,Y))$ pour déterminer la croyance attribuée au fait conclusion $(O_A(Y))$.

Pour définir cette notion, nous avons repris la définition de la *prévision* e de Wang, qui calcule la valeur future de la fréquence en fonction de la fréquence et de la confiance présentes. e est plus proche que f, d'un facteur c, du point de non préférence $\frac{1}{2}$. Pour nous, la *valeur de vérité* $v = c \times (f - \frac{1}{2}) + \frac{1}{2}$ est donc plus proche que f, d'un facteur c, de l'axe (S,C) du bitreillis, représentant une zone d'incertitude complète, puisqu'à égale distance des valeurs V et F. Nous utiliserons cette valeur, qui tient compte à la fois de la fréquence et de la confiance, pour décider, à l'aide d'un seuil, si une croyance peut être considérée comme vraie ou fausse.

DÉFINITION 3.8 La *valeur de vérité*, v, d'une croyance de $\mathcal{B}_\mathcal{N}$ est $c \times (f - \frac{1}{2}) + \frac{1}{2}$. En particulier : $v(\text{vrai}) \simeq 1$, $v(\text{faux}) \simeq 0$ et $v(\text{silence}) = \frac{1}{2}$, $v(\text{contradictoire}) = \frac{1}{2}$.

Cette valeur de vérité est à rapprocher de la notion d'élément *fondamental* (*grounded*) de Ginsberg [9] (voir section 2.4.4) qui désigne des éléments d'un bitreillis représentant des informations primitives. Nous cherchons à distinguer, parmi les croyances de l'ensemble $\mathcal{B}_\mathcal{N}$, celles qui représentent une information pouvant être interprétée aisément. Ce n'est pas le cas des valeurs proches de l'axe (S,C) du bitreillis, mais c'est le cas des croyances ayant une confiance satisfaisante et une bonne proximité des valeurs V ou F, ou même des deux axes (S,V) et (S,F) du bitreillis. La donnée d'un seuil suffit à désigner les éléments du bitreillis qui représentent des informations interprétables. Ce seuil permet donc de définir la notion de croyance *approximativement* vraie ou fausse utilisée dans la définition des mécanismes de raisonnement pour satisfaire les contraintes de valeur de vérité sur les prémisses des règles d'inférence.

DÉFINITION 3.9 Soit $0 < \sigma_V < \frac{1}{2}$, une croyance (p,q) de $\mathcal{B}_\mathcal{N}$ est

- *approximativement vraie* si $v((p,q)) \geq \sigma_V$, et on note $(p,q) \approx V$
- *approximativement fausse* si $v((p,q)) \leq 1 - \sigma_V$, et on note $(p,q) \approx F$
- *indéterminée* sinon, et on note $(p,q) \approx S$

3.3 Raisonner

Les mécanismes de raisonnement que nous allons décrire ci-après doivent permettre l'accroissement de la connaissance dans un schéma mental. Cette connaissance ainsi générée représente les régularités observées et des compléments d'information (des faits) sur des descriptions incomplètes. Nous utiliserons des règles définies à partir des principales règles d'inférence de la logique. Lorsque deux jugements (relation entre deux énoncés) partagent un même énoncé (le *moyen terme*), ils peuvent être utilisés comme prémisses pour inférer une relation entre deux autres énoncés (non partagés). Nous définissons trois schémas de combinaisons possibles de prémisses et de conclusions, correspondant aux trois règles classiques: déduction, abduction et induction.
Si une définition formelle de la déduction est acceptée, il n'en va pas de même pour l'abduction et l'induction. Nous donnerons notre définition des régles d'inférence que nous désignons par abduction et induction.

Les propositions utilisées comme prémisse ou conclusion d'une règle d'inférence sont décrites dans la table 1. Il existe quatre cas de figure de propositions catégorielles: proposition affirmative universelle (*tous les X ont la propriété Y*), affirmative particulière (*A a la propriété X*), négative universelle (*aucun X n'a la propriété Y*), et négative particulière (*A n'a pas la propriété X*).

Proposition	Interprétation	Formulation logique
$\Delta(X,Y)$=V	*tous les X ont la propriété Y*	$\forall A : X(A) \rightarrow Y(A)$
$\Delta(X,Y)$=F	*aucun X n'a la propriété Y*	$\forall A : X(A) \rightarrow \neg Y(A)$
$O_A(X)$=V	*A a la propriété X*	$\exists A : X(A)$
$O_A(X)$=F	*A n'a pas la propriété X*	$\exists A : \neg X(A)$

TAB. 1 - *Les propositions catégorielles assertées et leur interprétation*

Les résultats de ces règles d'inférence son calculés à partir des applications possibles d'une même règle pour tous les moyens termes trouvés.

3.3.1 Déduction

La déduction est couramment utilisée, explicitement ou implicitement, dans les systèmes d'apprentissage. Elle peut être illustrée par l'exemple suivant

Le Lagavulin est un whisky de la région de l'Islay,
les whiskies de la région de l'Islay ont un goût fumé,
donc le Lagavulin a un goût fumé.

et correspond à l'expression formelle suivante de la logique des propositions $x, x \rightarrow y \vdash y$.

Nous assimilons la déduction à une opération de diffusion de la connaissance dans le schéma mental: les règles de déduction que nous utilisons sont toutes logiquement valides (voir table 2); cette opération n'est donc pas génératrice d'hypothèse, mais propage une hypothèse sur d'autres couples d'énoncés du schéma mental.

(a)	(b)	(c)	(d)
$O_A(X)$=V, $\Delta(X,Y)$=V / $O_A(Y)$=V	$O_A(X)$=V, $\Delta(X,Y)$=F / $O_A(Y)$=F	$O_A(Y)$=V, $\Delta(X,Y)$=F / $O_A(X)$=F	$O_A(Y)$=F, $\Delta(X,Y)$=V / $O_A(X)$=F

TAB. 2 - *Les règles de déduction*

Le contexte de l'application d'une règle de déduction se pose comme suit: disposant d'un ensemble d'hypothèses caractérisant une propriété, déterminer si un objet possède ou non cette propriété. Il s'agit donc d'utiliser des hypothèses formulées sur des relations décrivant une propriété Y (du type $\Delta(X,Y)$) pour déterminer, à partir de ses éléments de description (du type $O_A(X)$), si un objet vérifie la propriété Y, et donc d'appliquer spécifiquement à A des généralités observées sur Y.

3.3.2 Abduction

Nous définissons l'abduction comme une technique de construction de faits hypothétiques. Elle peut être illustrée par l'exemple suivant :

Le Laphroaig a une odeur salée,
les produits de la mer ont une odeur salée,
donc le Laphroaig est un produit de la mer.

et correspond à l'expression formelle suivante de la logique des propositions : $y, x \rightarrow y \vdash x$.

Les résultats de l'application d'une abduction dans un schéma mental sont hypothétiques car les règles d'inférence utilisées (table 3(a) et 3(b)) sont logiquement invalides : elles supposent la symétrie de la relation Δ. La seconde transgresse en plus une règle de base de la formulation d'un syllogisme qui stipule que l'on ne peut rien conclure de deux prémisses négatives (en effet, toutes les conclusions sont alors possibles). Cependant nous nous permettons d'utiliser ces deux règles comme méthode de calcul d'une hypothèse particulière affirmative sur un fait, soutenue par la production d'une *preuve* (dont la construction sera décrite plus loin).

$$\frac{O_A(Y)=\mathrm{V} \qquad \Delta(X,Y)=\mathrm{V}}{O_A(X)=\mathrm{V}} \qquad \frac{O_A(Y)=\mathrm{F} \qquad \Delta(X,Y)=\mathrm{F}}{O_A(X)=\mathrm{V}}$$

(a) (non valide) (b) (non valide)

TAB. 3 – *Les règles d'abduction*

Le contexte de l'application d'une règle d'abduction se pose donc comme suit : disposant d'un ensemble d'hypothèses caractérisant une propriété, construire une hypothèse sur la vérification ou non de cette propriété par un objet.

3.3.3 Induction

L'induction est une technique de production d'hypothèses conjonctives. En logique des propositions, si x est vrai chaque fois que y est vrai, alors x implique y. Cette opération produit une règle hypothétique par généralisation des prémisses. Elle peut être illustrée par l'exemple suivant :

Le Macallan10yo est un whisky de la région des Highlands,
le Macallan10yo a une couleur or foncé,
donc les whiskies de la région des Highlands ont une couleur or foncé.

et correspond à l'expression formelle suivante de la logique des propositions : $x, y \vdash x \rightarrow y$.

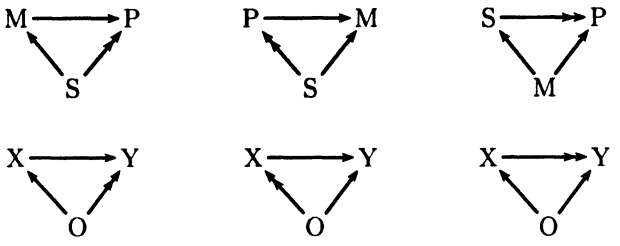
Si l'expression formelle de l'induction est logiquement valide, les règles d'inférence que nous désignerons comme inductives ne le sont pas, par la nature universelle de leurs conclusions. En effet, dans le premier cas (table 4(a)), à partir des deux observations affirmatives particulières $\Delta(Z,X)$ et $\Delta(Z,Y)$, une inférence valide conclurait une description affirmative *particulière*. Dans le deuxième cas (table 4(b)), une inférence valide conclurait une description négative *particulière*. Nous formulons l'hypothèse que ces observations peuvent se généraliser dans une relation utilisée pour compléter la base de connaissances par des faits hypothétiques.
Le contexte de l'application d'une règle d'induction se pose comme suit : disposant d'un ensemble de faits définissant deux propriétés X et Y, généraliser une relation entre X et Y.

La figure 6 illustre le principe général (distinction des éléments des prémisses liés par des flèches simples et de ceux de la conclusion, liés par une double flèche) des trois schémas de règles utilisés pour définir les opérations de raisonnement dans notre système.

$$\frac{\Delta(Z,X)=\mathrm{V} \quad \Delta(Z,Y)=\mathrm{V}}{\Delta(X,Y)=\mathrm{V}} \qquad \frac{\Delta(Z,X)=\mathrm{V} \quad \Delta(Z,Y)=\mathrm{F}}{\Delta(X,Y)=\mathrm{F}}$$

(a) (non valide) (b) (non valide)

TAB. 4 – *Les règles d'induction*



(a) Déduction (b) Abduction (c) Induction

FIG. 6 – *Illustration des mécanismes de raisonnement*

3.3.4 Raisonnement centré

Nous souhaitons pouvoir tolérer la présence d'exceptions à tout résultat d'une règle d'inférence, et incorporer, dans la croyance attribuée à ce résultat, la part d'exemples et de contre-exemples qu'il admet. Or, la présence d'une exception parmi l'ensemble des faits considérés pour une induction par exemple engendre un résultat contradictoire (C). Supposons en effet que parmi l'ensemble de tous les objets vérifiant la propriété X ($\omega = \{i \in L_\Omega \,|O_i(X) = V\}$), il existe un unique objet j ne vérifiant pas la propriété Y ($\exists!\ j \in \omega$ vérifiant $O_j(Y) = F$, et $\forall i \in \omega, i \neq j, O_i(Y) = V$). Le résultat calculé est la somme (borne supérieure au sens de la connaissance d'un ensemble de croyances) de toutes les valeurs $\Delta(Z,X) \wedge \Delta(Z,Y)$ obtenues pour chaque moyen terme Z utilisé. La croyance induite pour la règle $\Delta(X,Y)$ à partir de l'ensemble d'objets ω est donc contradictoire. Afin de corriger ce résultat et de représenter la part d'objets vérifiant ou falsifiant la règle $\Delta(X,Y)$, nous appliquons des règles de calcul majoritaire, dont les résultats sont représentatifs de l'ensemble des données. Ceci est possible pour chacune des opérations de raisonnement en utilisant un opérateur de calcul d'une croyance moyenne dans un ensemble. Cet opérateur est défini pour un ensemble de croyances donné B à partir de la définition d'une distance entre deux croyances de B, et calcule le centre de gravité d'un sous ensemble de B.

3.3.5 Condition d'application des raisonnements

Il convient de préciser les conditions d'application des règles d'inférence que nous venons de définir. En effet, il s'agit d'associer une croyance inférée à une relation uniquement dans le cas où cette opération apporte un gain d'information, ce qui se traduit pour nous par un niveau de connaissance croissant.

> La croyance associée à une correspondance est égale à la valeur obtenue par application d'une règle d'inférence si la croyance du résultat inféré est plus connue que la croyance du résultat courant.

Ce qui donne, pour l'induction par exemple,

> Soit $\Delta(X,Y) = c_0$
>
> de $\Delta(Z,X) = c_1$ et $\Delta(Z,Y) = c_2$ tels que $c_1 \approx V$, conclure $\Delta(X,Y) = c_3$ ssi $c_3 \geq_K c_0$

Cette précision permet en particulier de conserver les croyances associées aux faits contradictoires à une règle induite. En effet, supposons l'existence d'un fait contradictoire (une contradiction) $O_A(X) = V$

et $O_A(Y) = F$ à une règle $\Delta(X, Y) = (0.8, 0.1)$ induite, alors l'application d'une déduction sur $O_A(Y)$ associerait à ce fait la croyance $(0.8, 0.1)$ (proche de F). Or, de toute évidence, nous ne souhaitons pas perdre la connaissance de $O_A(Y)$ issue de la relation de correspondance δ, donc élément de l'ensemble des exemples du domaine d'application. La croyance associée au résultat d'une règle d'inférence étant toujours d'un niveau de connaissance plus faible que les croyances V ou F associées aux exemples, les données sont préservées.

Ce principe assure la monotonie de la connaissance dans un schéma mental.

3.3.6 Clôture déductive

Nous désignons particulièrement l'état d'un schéma mental qui correspond à une saturation de la connaissance. Cet état se nomme la *clôture déductive* et est atteint lorsque tout ce qu'il était possible d'inférer à partir d'un état donné du schéma mental l'a été, par application d'opérations déductives successives sur les éléments de la base de connaissance. En particulier, dans un schéma mental déductivement clos, l'ordre sur les éléments du langage est représenté par les croyances des relations entre ces éléments.

Un schéma mental est déductivement clos lorsque la description de chacun des objets qui le composent sont eux-mêmes déductivement clos. Cette propriété nécessite de pouvoir comparer des objets suivant une relation de spécificité afin de déterminer l'objet le plus général qu'il est possible de définir à partir des données et des hypothèses du schéma mental.

Ordonner les objets et les concepts Les notions de *clôture* et de *clôture déductive* que nous allons définir ci-après nécessitent l'introduction d'une relation d'ordre entre les objets d'un schéma mental. Les objets mentaux sont ordonnés entre eux suivant une relation de spécificité: un objet est moins général qu'un autre s'il est plus connu sur tous ses attributs.

DÉFINITION 3.10 L_Ω est ordonné par la relation $\leq_\Omega$ définie par : pour $a, b \in L_\Omega$, $a \leq_\Omega b$ ssi $\forall x \in L$, $O_a(x) \geq_K O_b(x)$.

La relation $\leq_\Omega$ ordonne les objets en fonction de la connaissance que l'on a de leur description.

Les concepts mentaux sont eux aussi ordonnés, suivant un critère de vérité.

DÉFINITION 3.11 L_χ est ordonné par la relation $\leq_\chi$ définie par : pour $\alpha, \beta \in L_\chi$, $\alpha \leq_\chi \beta$ ssi $\forall x \in L_\Omega$, $K_\alpha(x) \leq_T K_\beta(x)$ $(\Leftrightarrow \forall x \in L_\Omega, O_x(\alpha) \leq_T O_x(\beta))$.

Définition d'une clôture déductive Calculer une clôture d'un objet du schéma mental consiste à compléter la description de cet objet à l'aide des faits et des hypothèses contenus dans le schéma.

DÉFINITION 3.12 Soit un objet A, une *clôture* de A est un objet, noté $\mathrm{cl}(A)$, qui vérifie les propriétés suivantes :

1. $O_{\mathrm{cl}(A)}(e_1) \geq_K \bigwedge_{e_2 \geq_L e_1} O_{\mathrm{cl}(A)}(e_2)$
2. Si la négation est définie sur le langage, alors $\forall a, O_{\mathrm{cl}(A)}(\neg e) = \neg O_{\mathrm{cl}(A)}(e)$
3. Si $\Delta(e_1, e_2) = V$, alors $O_{\mathrm{cl}(A)}(e_1) \leq_T O_{\mathrm{cl}(A)}(e_2)$

A est dit *clos* si $A = \mathrm{cl}(A)$.

Nous nous intéressons à la clôture de A la plus proche de A, au sens de la connaissance, donc *la clôture de A la plus générale* que nous désignerons par $\mathrm{cl}_0(A)$. En effet, $\mathrm{cl}_0(A)$ contient les connaissances implicites contenues dans A et dans le langage. Ceci signifie que $\mathrm{cl}_0(A)$ et A désignent le même objet. Toute autre clôture possible pour A et différente de $\mathrm{cl}_0(A)$ désigne un objet plus spécifique que A.
Soient par exemple la relation $\Delta(a, b) = V$ et la description d'un objet A par $O_A(e_1) = V$ et $O_A(e_2) = S$. $\mathrm{cl}_0(A)$ est défini par $O_{\mathrm{cl}_0(A)}(e_1) = V$ et $O_{\mathrm{cl}_0(A)}(e_2) = S$. La clôture $\mathrm{cl}_\infty(A)$ telle que $O_{\mathrm{cl}_\infty(A)}(e_1) = C$ et

$O_{\mathrm{cl}_\infty(A)}(e_2) = C$ est une clôture de A qui ne représente pas une connaissance semblable à celle contenue dans la description de A et dans le langage.

DÉFINITION 3.13 Soit un objet A, une *clôture déductive* de A est un objet, noté $\mathrm{cl}_0(A)$, qui vérifie la propriété suivante :

$$\forall \mathrm{cl}(A) \neq \mathrm{cl}_0(A), \text{ clôture de } A, \forall e \in L, O_{\mathrm{cl}_0(A)}(e) \leq_K O_{\mathrm{cl}(A)}(e).$$

$\mathrm{cl}_0(A)$ est *la clôture la plus générale* de A.
A est dit *déductivement clos* si $A = \mathrm{cl}_0(A)$.

La clôture déductive vérifie les propriétés de clôture définies par M.L. Ginsberg. En particulier, la propriété $\mathrm{cl}_0(\mathrm{cl}_0(A)) = \mathrm{cl}_0(A)$ est vérifiée par la clôture déductive alors qu'elle ne l'est pas par la clôture simple d'un objet.

DÉFINITION 3.14 Un schéma mental est *déductivement clos* ssi : $\forall a \in L_\Omega$, a est déductivement clos.

Les problèmes d'apprentissage opèrent généralement sur une base de connaissances déductivement close. Les exemples utilisés par le système sont décrits par des connaissances : un objet possède un attribut ou non. Rechercher la clôture déductive d'un tel schéma mental consiste à compléter les descriptions manquantes par des informations contenues dans le langage (le calcul de la relation de correspondance δ^* à partir de δ en est un exemple). En calculant les déductions qui complètent les informations manquantes de chaque nouvel objet, il est aisé de conserver le schéma mental déductivement clos. La clôture déductive est en fait un état du schéma mental que nous cherchons à conserver.
Traiter le cas général d'une base de connaissances dont les données possèderaient a priori une part d'incertitude est bien plus complexe. Il s'agit dans ce cas, non plus de conserver une propriété, mais de calculer un nouvel état du schéma mental.

3.4 Soutenir les hypothèses

La nature des connaissances manipulées dans un schéma mental et les opérations qui les produisent imposent de fournir à ce système des méthodes de validation des règles ou faits qu'il engendre. En particulier, la nature hypothétique des résultats de certaines opérations de raisonnement nécessite un contrôle particulier des conclusions obtenues. Il est réalisé afin d'éviter la redondance d'information, dans le cas de l'induction. Dans le cas de l'abduction, il permet l'argumentation par la production de preuves.

3.4.1 Induction pertinente

Le contrôle des opérations inductives vise à répondre à la question suivante : *"le résultat de l'induction de la règle $\Delta(X,Y)$ apporte-t-il une connaissance supplémentaire?"*. Il s'agit en effet de ne pas représenter une information redondante au regard des connaissances pouvant être obtenues par "transitivité", par la clôture déductive du schéma mental, et de construire ainsi un ensemble minimum de règles, plus économique à réviser si cela s'avère nécessaire.
Supposons la connaissance des deux règles $\forall v, X(v) \rightarrow Y(v)$, et $\forall v, Y(v) \rightarrow Z(v)$, alors le fait $X(A)$ nous permet de conclure $Y(A)$ et $Z(A)$, par application des deux premières règles et de la clôture déductive, sans ajouter la règle $\forall v, X(v) \rightarrow Z(v)$ dans la base de connaissance.

Nous nous assurons donc que chaque résultat induit et représenté sous forme de règle dans un schéma mental ne peut être obtenu par transitivité des règles existantes. La définition 3.15 décrit le principe adopté pour décider de la pertinence d'une conclusion inductive : si la clôture déductive du schéma mental complété de la nouvelle règle induite $\Delta(X,Y)$ est "proche" du schéma mental initial (i.e. les descriptions des objets dans les schémas mentaux déductivement clos sont semblables), alors cette règle *n'apporte pas de connaissance supplémentaire*.

DÉFINITION 3.15 Soient deux schémas mentaux Δ_1 et Δ_2 tels que Δ_2 est égal à Δ_1 complété de la règle induite $\Delta_2(X,Y)$, $\varepsilon > 0$, et une distance d entre croyances. On note $\mathrm{cl}_0^1(A)$ et $\mathrm{cl}_0^2(A)$ la clôture déductive d'un objet A dans les schémas mentaux respectivement Δ_1 et Δ_2. L_Ω désigne l'ensemble des objets de Δ_1 (et donc de Δ_2).
Si $\forall A \in L_\Omega$, $\mathsf{d}(O_{\mathrm{cl}_0^1(A)}(X), O_{\mathrm{cl}_0^2(A)}(X)) < \varepsilon$, et $\mathsf{d}(O_{\mathrm{cl}_0^1(A)}(Y), O_{\mathrm{cl}_0^2(A)}(Y)) < \varepsilon$, alors la règle $\Delta(X,Y)$ *n'apporte pas de connaissance supplémentaire* dans Δ_1.

3.4.2 Abduction prouvée

La nature hypothétique de la connaissance dans un schéma mental et la non validité des opérations d'abduction qui produisent de nouvelles connaissances imposent de fournir au système une méthode de contrôle des résultats obtenus. Mais au-delà, les structures ainsi construites devraient permettre la production d'une explication dans le cadre d'une confrontation avec un usager. Nous nous appuyons sur les *preuves* pour contrôler l'opération abductive.

Une preuve est une explication reconnue et acceptée [4 ; 3]. Le mot preuve fait référence à la fois au processus d'élaboration et à son résultat. Le processus de construction d'une preuve décompose une conjecture originale (un énoncé que l'on cherche à caractériser) en sous-conjectures ou lemmes. Le résultat obtenu, la preuve, est donc un ensemble de lemmes. Tous les exemples d'une conjecture doivent donc vérifier les lemmes qui composent la preuve. Une preuve peut être critiquée en exhibant un exemple de la conjecture qui ne vérifie pas un lemme.

Le principe de la preuve d'une conjecture dans un schéma mental repose sur le mécanisme de contrôle par objections dû à I. Lakatos [17], dont C. Barboux [4] a réalisé une formulation adaptée au contrôle par objections d'une théorie semi-empirique et que E. Mephu Nguifo [19] a appliqué au contrôle et à la révision d'une connaissance apprise. De manière générale, les *objections* sont définies comme des contraintes qui marquent la différence entre des objets quelconques et des exemples de la preuve. Cette différence est due à des incorrections ou à des incomplétudes de l'expression d'un objet par rapport à l'expression d'un exemple de la preuve. Une *objection générale* est une expression de ce qu'il est suffisant de reconnaître sur un objet pour qu'il ne soit pas considéré comme exemple de la conjecture. De cette dernière définition, nous définissons les objections à une conjecture dans un schéma mental comme la négation d'une régularité observée sur les lemmes.

Nous construirons donc la preuve d'une conjecture dans un schéma mental en exhibant un ensemble de lemmes, puis en calculant, à partir de cet ensemble, un ensemble d'objections à cette conjecture. La preuve d'une conjecture dans un schéma mental sera finalement réduite à un sous-ensemble d'objections générales. Prouver un résultat positif consiste alors à exhiber la preuve d'une conjecture et montrer que l'exemple ne vérifie aucune des objections générales retenues ; à l'inverse, la vérification d'une objection générale par un contre-exemple suffit à prouver un résultat négatif.

Ainsi, argumenter que l'objet A vérifie la conjecture C (soit $O_A(C) = V$) consiste à exhiber une conjonction de faits, caractéristique des exemples satisfaisant cette conjecture, les lemmes, et à vérifier que la description de A est en adéquation avec ces lemmes, ou plutôt, que A ne vérifie aucune objection à C, négation d'une régularité observée sur les lemmes de C.

DÉFINITION 3.16 Soient un schéma mental Δ, un énoncé k et un seuil $0 \leq \sigma_L \leq 1$.
L'ensemble des *exemples* de k est $X_k = \{x \in L_\Omega \mid O_x(k) \approx V\}$.
Un *lemme* est un énoncé $l \in L$ tel que : $|\{x \in X_k \,|O_x(l) \approx V\}| \geq \sigma_L$.
$\mathcal{L}_k$ est l'ensemble des lemmes de k. $\mathcal{L}_k$ peut être ordonné de sorte qu'un lemme pour k soit considéré comme "meilleur" qu'un autre s'il est vérifié sur un plus grand nombre d'exemples.

$$l_k^1 \geq_{\mathcal{L}_k} l_k^2 \Leftrightarrow |\{x \in X_k \,|O_x(l_k^1) \approx V\}| \geq |\{x \in X_k \,|O_x(l_k^2) \approx V\}|$$

Un lemme est donc une condition nécessaire pour vérifier une conjecture. À l'opposé, une objection générale, définie comme une régularité observée sur les lemmes, constitue une condition suffisante à la falsification d'une conjecture. Un objet A qui ne vérifie pas un énoncé o défini comme une objection générale à une conjecture k ne vérifie pas k. En effet, puisque o est défini comme une régularité observée sur les lemmes de k (pour une proportion suffisante d'éléments $l \in \mathcal{L}_k$, $\Delta(l, o) \approx V$) alors la non vérification de o par un objet A ($O_A(o) \approx F$) nous permet de conclure à la non vérification des lemmes pour k qui vérifient o. Si A ne vérifie pas les lemmes pour k alors A ne vérifie pas k.

DÉFINITION 3.17 Soient un schéma mental Δ, un énoncé k et un seuil $0 \leq \sigma_O \leq 1$.
Une *objection générale* est un énoncé $o \in L$ tel que : $|\{l \in \mathcal{L}_k \,|\Delta(l, o) \approx V\}| \geq \sigma_O$.
$\mathcal{O}_k$ est l'ensemble des objections générales de k. Comme pour l'ensemble des lemmes de k, $\mathcal{O}_k$ peut être ordonné par une relation $\leq_{\mathcal{O}_k}$ de sorte qu'une objection pour k soit considérée comme "meilleure" qu'une autre si elle est vérifiée sur un plus grand nombre de lemmes.

Prouver qu'un objet vérifie une conjecture consiste à montrer qu'il vérifie chacune de ses objections générales. La preuve de l'événement contraire est réalisée en montrant que cet objet ne vérifie pas un seuil minimal d'objections générales, dépendant du nombre de lemmes exhibés pour la conjecture.
La preuve d'une conjecture est donc finalement un énoncé défini comme la conjonction des objections générales.

DÉFINITION 3.18 Soit un schéma mental Δ, un énoncé k et l'ensemble de ses objections générales $\mathcal{O}_k = \{o_1, \ldots, o_n\}$. Une *preuve* pour k est un énoncé P_k défini par $P_k = o_1 \wedge \cdots \wedge o_n$.

La preuve d'un fait abduit $O_A(k) = V$ se compose donc de la présentation d'une preuve pour k, P_k, et de la vérification par A de cette preuve, soit $O_A(P_k) \approx V$.
Plusieurs preuves peuvent être définies pour une même conjecture, par des sélections différentes de lemmes et objections générales. La validité de chaque preuve peut être vérifiée au cours d'une interaction avec un expert.

Nous avons posé les bases formelles de notre système pour définir le modèle de représentation et les mécanismes de production d'hypothèses et de génération de connaissance dans un schéma mental. Les notions qui seront manipulées par la suite sont celles d'énoncé (élément d'un langage), d'objet (statut donné à un énoncé désignant une description), de concept (notion opposée à celle d'objet), pour manipuler les éléments de la base de connaissances.

4 GESTION DE LA CROISSANCE INCRÉMENTALE DE LA CONNAISSANCE

Notre objet est de formaliser un système d'apprentissage automatique qui enrichit sa base de connaissances, pour un domaine d'application, dans une interaction avec un expert. Nous souhaitons ce modèle formel le plus général possible, non lié à un domaine. Nous avons donc défini un cadre de représentation, par les notions de langage et croyance, sur lequel nous appliquons des opérations de raisonnement qui doivent permettre l'augmentation de la connaissance dans la base, par la recherche de régularités et la construction d'hypothèses sur des éléments inconnus.

L'application répétitive d'un mécanisme d'induction, exhibant les régularités, sur la totalité d'une base de connaissances, et après chaque changement d'état, fait échouer le processus d'apprentissage sur cette base et pour tout domaine d'application pour deux raisons essentielles :

- la généralisation de la totalité d'un domaine, comportant des éléments de nature exceptionnelle, est une ambition vaine. Les nombreux travaux sur la représentation et le traitement des exceptions, qui ont célébré Tweety, l'oiseau non-volant, tendent à le prouver.
- un processus d'apprentissage ignorant le caractère évolutif de la base de connaissances est voué à l'échec, une hypothèse pouvant produire une inconsistance dans une base de connaissances ayant changé d'état.

Il importe donc de contrôler les mécanismes générateurs d'hypothèses afin de contrer les deux résultats excessifs : généralisation silencieuse, si la totalité du domaine tente d'être modélisée, ou contradictoire, si des hypothèses formulées sur un sous-ensemble du domaine sont appliquées sur sa totalité.

Nous présentons une approche formelle, inspirée de la *théorie du noyau central* issue des recherches en psychologie sociale. Le principe général est d'émettre une hypothèse sur le concept à apprendre en exhibant un sous-ensemble de la base de connaissances susceptible d'évoluer de façon monotone au cours de l'apprentissage. Une propriété fondamentale, a priori, de ce sous-ensemble est la cohérence. La propriété recherchée au cours de l'évolution est la stabilité. Nous verrons que la définition du noyau central vérifie ces contraintes.

4.1 Les représentations sociales

Les représentations sociales constituent un aspect particulier de la connaissance, dite de "sens commun". La spécificité de cette connaissance réside dans le caractère social des processus qui la produisent.
Serge Moscovici a initié une approche psychosociologique des représentations sociales, abandonnant une distinction classique entre le sujet et l'objet. Ceci conduit à poser qu'il n'existe pas a priori de réalité objective, mais que toute réalité est représentée, appropriée par un individu ou par un groupe, cette réalité

appropriée et restructurée constituant la réalité même pour le sujet. Or, cette appropriation résulte de l'individu lui-même (de son vécu, de ses expériences propres) et du contexte social et idéologique qui l'environne.

> *"... il n'y a pas de coupure entre l'univers extérieur et l'univers intérieur de l'individu ou du groupe."* [22].

Propriétés Nous pouvons définir les représentations sociales comme un ensemble de connaissances, croyances et opinions partagées par un groupe à l'égard d'un objet donné (nous parlons ici d'objets sociaux : la culture, la nature...), dont nous énumérons les propriétés :

- Les représentations sociales entretiennent avec leur sujet des rapports de *symbolisation* et d'*interprétation* résultant d'une activité mentale : le sujet reconstruit la réalité à laquelle il est confronté (Moscovici parle de *"remodelage mental"*). La réalité devient donc différente d'un groupe à un autre.
- Les informations qui proviennent de l'objet sont sélectionnées, toujours catégorisées, transformées pour les rendre consonnantes avec les valeurs du groupe.
- L'activité de production est socialement marquée.

Une représentation[2] est donc sociale si elle est le résultat d'un ensemble d'interactions spécifiques, et si elle est partagée par les individus d'un même groupe. Elle marque de ce fait la spécificité de celui-ci et contribue à le différencier des autres groupes.

Processus L'activité de production est rendue possible par deux processus d'activité cognitive, l'*objectivation* et l'*ancrage*.
L'*objectivation* est un processus rattaché au fonctionnement de la pensée sociale qui simplifie les informations en provenance d'un objet, concrétise les notions, et résume à partir d'une logique interne au groupe. Cette information est donc décontextualisée et schématisée, on parle de la constitution d'un *noyau figuratif* : les éléments de ce noyau ont un statut d'évidence et sont donc, par ce fait, non négociables.
L'*ancrage* définit l'enracinement social des représentations. La construction d'une représentation s'opère en effet toujours en référence à des croyances, à des savoirs pré-existants, et qui sont dominants dans le groupe.

> *"La pensée constituante s'appuie sur la pensée constituée pour ranger la nouveauté dans des caves anciennes, dans du déjà connu, avec pour effet des transformations de part et d'autre"* [13].

4.1.1 La théorie du noyau central

L'idée de cette théorie [1], reprenant celle du noyau figuratif, est que la représentation est organisée autour d'un noyau central dont la fonction est structurante et qui gère le sens de l'ensemble de la représentation. Ce noyau central sera un sous-ensemble de la représentation, composé d'un ou de quelques éléments dont l'absence la destructurerait ou lui donnerait une signification complètement différente. Notez que ce n'est pas la dimension quantitative qui importe, mais le fait que celui-ci donne la signification de la représentation, soit sa dimension qualitative.

Ce noyau central d'une représentation a deux fonctions et une caractéristique essentielle. La première de ces fonctions est la *génération*, c'est à dire que le noyau central est l'élément par lequel se crée ou se transforme la signification des autres éléments constitutifs de la représentation. La deuxième fonction est l'*organisation* : il détermine la nature des liens qui unissent entre eux les éléments d'une représentation, il est donc unificateur et stabilisateur de la représentation.
Enfin, la caractéristique essentielle est la *stabilité* : le noyau central est l'élément qui résiste le plus au changement. Cette propriété est possible par la présence d'éléments dits *périphériques* dans la représentation. Ces éléments périphériques sont en relation directe avec le noyau central, leur valeur et leur fonction sont déterminées par ce dernier, afin de permettre une modulation individualiste de la représentation

2. Dans cette partie, la notion de *représentation* correspond à celle utilisée en psychologie : il existe une représentation, globale et unique.

tout en maintenant la stabilité du noyau central. De plus, la transformation de ces éléments périphériques constitue un processus de défense de la représentation face à des informations contradictoires, et d'intégration de nouvelles informations sans déstabilisation.

Une représentation se transforme réellement lorsque le noyau central est transformé. Cette transformation peut être brutale, c'est à dire qu'il peut se produire, compte tenu de telle information, de tel domaine pratique, de tel évènement, une attaque directe du noyau central qui aboutit à sa remise en cause immédiate : situation de crise ; mais le mode dominant de l'évolution d'une représentation consiste en une transformation progressive des éléments périphériques, puis celle-ci devenant de plus en plus importante, le noyau central est, à son tour, progressivement transformé.

4.2 Dynamique des représentations sociales

Le changement d'état des représentations n'est ni aléatoire ni accidentel, c'est un aspect fondamental de leur fonctionnement interne : produites par des groupes, elles sont intimement liées à la vie des groupes, et évoluent avec eux. L'état nouveau de la représentation est donc rarement en rupture avec son état ancien.

L'approche théorique de la dynamique des représentations est fondée sur le concept de *cognition prescriptive* :

- Une représentation sociale est constitué d'un ensemble fini et organisé de cognitions.
- Ces cognitions, dans la plupart des cas, désignent l'ensemble des indications qui s'imposent devant une situation particulière. On dit alors qu'elles sont prescriptives.

On retrouve une application de cette idée dans [27] où le terme de *script* désigne un ensemble de scénarios susceptibles d'apparaître dans une situation donnée.

4.3 Explication de la transformation

Les circonstances externes à la représentation, caractéristiques de l'environnement du groupe, doivent être modifiées massivement. Ces modifications résultent d'un évènement important, boulversant l'ordre actuel des choses. On parle alors d'*évènement impliquant*.
Les pratiques sociales liées à l'objet de la représentation sont modifiées : le groupe cherche à s'adapter à la situation nouvelle. Des pratiques nouvelles deviennent de plus en plus fréquentes.
La pondération accrue des schèmes (ou cognitions) nouvellement activés leur donne un rôle de plus en plus important, ils apparaissent donc de plus en plus fréquemment dans le discours. Il y a donc une activation bijective entre les schèmes prescriteurs et les pratiques. Si beaucoup de schèmes sont activés simultanément, il y a complexification du champ de la représentation. Ceci implique une simplification la plus économique possible de la représentation (comportement habituel face à une complexification trop importante), éventuellement par une fusion des schèmes prescripteurs activés en un concept unique, ce dernier devenant le nouveau noyau central.

4.4 Exhiber le noyau central : l'analyse de similitude

Dans la recherche d'une structure significative de la représentation étudiée, Claude Flament expose une méthode d'analyse de données dont les bases mathématiques respectent le caractère généralement qualitatif des données, et surtout certaines idées qu'on peut avoir sur la nature des représentations sociales [6]. À ce propos, Flament émet une conception *minimale* des représentations sociales :

- une représentation est un ensemble de *cognèmes* (éléments significatifs minimums) organisés par de multiples relations ;
- ces relations peuvent être *orientées* (implication, causalité, antagonisme...), mais toutes peuvent se "dégrader" en une relation symétrique traduisant une corrélation ;
- cette relation n'est, en général, pas transitive : si A et B sont corrélés pour certaines raisons, B et C pour d'autres raisons, il se peut fort que A et C ne le soient pas.

La relation symétrique et non transitive mise en jeu dans cette conception minimale est une *relation de similitude*. Elle peut être valuée pour exprimer le fait que deux cognèmes vont ensemble plus ou moins

fortement. La méthode d'analyse de données dite *analyse de similitude*, fondée sur la théorie des graphes, étudie cette relation. Elle se décompose en quatre étapes successives.

Une enquète préliminaire permet d'établir une liste d'items évoquant un objet de représentation déterminé.

Ces items sont présentés à des sujets (sous diverses formes: liste écrite, images ou items inscrits sur des cartes, mais tous les items sont toujours présentés en même temps) qui doivent regrouper ceux qui *vont ensemble*, à leur sens, de façon à faire apparaître les relations de similitude entre les items.

On calcule ensuite *l'indice de similitude* pour chaque paire d'items, valeur numérique calculée à partir de la proportion de leur co-occurence dans les associations réalisées par les sujets. On obtient ainsi un graphe valué mettant en évidence toutes les relations entre les items considérés.

Enfin, la dernière étape cherche à mettre en évidence la structure de la représentation en la présentant sous forme d'un *arbre maximum* (au sens de la somme des valuations) obtenu par l'algorithme de Kruskal [14].

Si, dans un arbre, on supprime une arête, on déconnecte le graphe en exactement deux sous-arbres. Si on part d'un arbre maximum, toute paire composée d'un sommet pris dans un sous-arbre et d'un sommet pris dans l'autre sous-arbre, a une valuation inférieure à la valuation de l'arête supprimée. De façon duale, la similitude entre deux éléments non consécutifs dans l'arbre maximum est inférieure aux valuations des arêtes constituant le plus court chemin, dans l'arbre, entre ces deux sommets.
En supprimant les arêtes dans l'arbre dans l'ordre croissant de leur valuation, en commençant par la plus faible, on obtient une série de classifications de plus en plus fines dite *classification hiérarchique*. Une telle classification permet de choisir une valeur d'arête définissant une décomposition en sous-graphes significative, illustrant les différentes classes contenues dans la représentation.
Enfin un *coefficient de structure* peut être calculé pour estimer la *centralité* de chacun des sommets pour le sous-graphe qui le contient. L'idée consiste à dire que deux sommets x et y se ressemblent en partie parce qu'ils ressemblent tous deux à un autre sommet a. La partie de similitude entre x et y "relayée" par a est au maximum égale à la similitude entre a et x ou entre a et y.

5 NOYAU CENTRAL ET SCHÉMA MENTAL

Nous définissons la notion de noyau dans un schéma mental en nous inspirant des propriétés du noyau central. L'idée est d'extraire de notre base de connaissances un sous-ensemble reconnu comme stable, de désigner celui-ci comme étant le concept à apprendre, et de faire réagir le système de sorte que les valeurs de croyance de ses éléments évoluent monotoniquement.

5.1 Création

Notre critère de stabilité d'un sous-ensemble est lié à sa cohérence. Ainsi, nous émettons l'hypothèse qu'un ensemble d'éléments liés par des valeurs de croyance d'un niveau de *confiance* supérieur à un seuil donné est susceptible d'évoluer monotoniquement.
L'idée que deux énoncés sont fortement associés est assez bien traduite par le fait que la *confiance* accordée à la croyance qui les relie est élevée, et ceci, quelle que soit sa fréquence. En effet, la relation de similitude représente autant une relation de ressemblance que d'antagonisme, le fait essentiel étant que les deux cognèmes mis en jeu soient fortement liés.

5.1.1 L'analyse de similitude

Une première méthode de création d'un noyau dans un schéma mental reprend la méthode d'*analyse de similitude* utilisée en psychologie sociale et présentée précédemment.

Nous pouvons considérer qu'un schéma mental représente une matrice d'indices de similitude entre les énoncés. Plus précisément, une matrice obtenue à partir d'un schéma mental et représentant les valeurs de confiance des croyances associées à tout couple d'énoncés est vue comnme l'expression de la relation de similitude entre les énoncés. À partir d'une telle matrice, la construction de l'arbre maximum s'effectue en appliquant l'algorithme de Kruskal [14] dont le principe est le suivant: on considère l'ensemble des

arêtes par ordre décroissant de la valuation et on retient toutes les arêtes qui ne forment pas un cycle avec celles qui ont précédemment été retenues. Enfin, la suppression des arêtes de valuation inférieure ainsi que le calcul du coefficient de structure des éléments des sous-graphes obtenus nous permet d'évaluer si la classification obtenue est satisfaisante, c'est à dire, si nous pouvons extraire, de l'arbre maximum, un ou plusieurs sous-graphes que nous pourrons considérer comme noyau de notre représentation.

5.1.2 Extraire un concept

Nous pouvons aussi ramener le problème de la découverte d'un ensemble d'éléments fortement connectés à l'extraction dans un tableau binaire, $O \times A$, d'un *concept*, c'est à dire une partition (X, Y) des deux ensembles O et A, représentant un bloc homogène maximum de 1 [19].

I	a	b	c	d	e	f
1		1	1	1	1	
2	1	1	1			
3	1	1				
4				1	1	
5			1	1		
6	1					
7		1	1			

TAB. 5 – *Correspondance binaire*

Le tableau 5 définit un *contexte*, c'est à dire, une relation binaire I entre deux ensembles finis d'objets ($O = \{1,2,3,4,5,6,7\}$) et d'attributs ($A = \{a,b,c,d,e,f\}$). Si l'objet $o \in O$ vérifie l'attribut $a \in A$, alors le couple (o,a) appartient à la relation I (noté oIa).
L'expression en *intension* d'un sous-ensemble d'objets $X \subseteq O$ est définie par $X' = \{a \in A \,|oIa, \forall o \in X\}$.
L'expression en *extension* d'un sous-ensemble d'attributs $Y \subseteq A$ est $Y' = \{o \in O \,|oIa, \forall a \in Y\}$. Ainsi on a ici pour $X = \{1,2,7\}$, $X' = \{b,c\}$ et pour $Y = \{d,e\}$, $Y' = \{1,4\}$.
Un *concept* dans le contexte (O, A, I) est un couple (X,Y), $X \subseteq O$ et $Y \subseteq A$, où $X' = Y$ et $Y' = X$. Dans le tableau ci-dessus, un concept peut être vu comme un rectangle maximal de 1 au sens de l'inclusion. Le couple (X={1,2,7},Y={b,c}) est un rectangle maximal.
Autrement dit, X est l'ensemble des éléments de O qui vérifient tous les éléments de Y (extension), et Y est l'ensemble des attributs communs à tous les éléments de X (intension).

Un schéma mental se ramène à une correspondance binaire en notant par 1 les croyances qui vérifient une valeur de confiance suffisante et une proximité des valeurs V ou F, soit donc les croyances que nous désignons comme *aprroximativement vraies* ($c \approx V$) ou *fausses* ($c \approx F$). Toutes les autres sont désignées par un 0.
Dans un *schéma mental* défini comme un quadruplet $S = (L_1, L_2, B, \Delta)$ où Δ ($\Delta : L_1 \times L_2 \rightarrow B$) définit une relation des deux ensembles d'énoncés L_1 et L_2 vers un ensemble de croyances B, nous définissons comme *noyau de* S le schéma mental (N_1, N_2, B, Δ) tel que :

$$N_1 \subseteq L_1,\ N_2 \subseteq L_2,$$

et

$$\frac{|N_2 \cap N_1'|}{|N_2 \cup N_1'|} \geq \sigma^e \text{ et } \frac{|N_1 \cap N_2'|}{|N_1 \cup N_2'|} \geq \sigma^e,$$

avec :

$$N_1' = \left\{ y \in L_2 \,\middle|\, \frac{|\{x \in N_1 | \Delta(x,y) \approx V \text{ ou } \Delta(x,y) \approx F\}|}{|N_1|} \geq \sigma^m \right\}$$
$$N_2' = \left\{ x \in L_1 \,\middle|\, \frac{|\{y \in N_2 | \Delta(x,y) \approx V \text{ ou } \Delta(x,y) \approx F\}|}{|N_2|} \geq \sigma^m \right\}$$

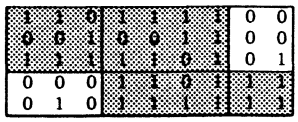
où les valeurs σ^e et σ^m désignent respectivement des seuils d'égalité et de majorité.
Le couple (N_1, N_2) décrit un concept dans le contexte défini par L_1, L_2 et Δ.

Une méthode de construction d'un concept dans un tableau binaire consiste à permuter les lignes et les colonnes du tableau afin d'obtenir des blocs homogènes de 0 ou de 1, définissant des couples de partitions (k,l) des ensembles O et A mis en relation dans le tableau binaire [10]. On associe à chaque couple (k,l) une valeur binaire idéale, 0 ou 1. Le tableau binaire obtenu est appelé le *noyau* [3]. On cherche à minimiser

3. terme employé par Govaert, qui ne désigne pas le *noyau* que nous cherchons à construire ici.

l'écart entre le tableau initial et le tableau idéal, mesurer par le nombre de fois où la valeur associée à un couple (i,j) est différente de la valeur idéale associée au couple de classes auxquelles appartiennent i et j.

Exemple 5.1 Soit le tableau binaire suivant



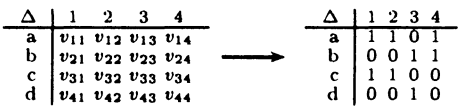
Ce tableau a été écoupé en 2 × 3 partitions (représentées par les parties grisées et blanches du tableau) de façon à former le tableau contenant les valeurs idéales suivantes

1 1 0
0 1 1

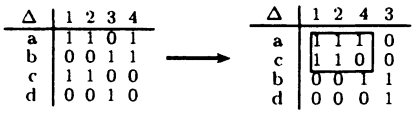
Alors l'écart entre le tableau initial et le tableau idéal est égal à 9. 9 éléments du tableau initial ont en effet une valeur idéale associée différente.

On détermine ainsi l'existence de plusieurs concepts dans le tableau de correspondance binaire pour un nombre de classes fixé. Le choix de l'un ou l'autre de ces concepts s'effectue en ordonnant ceux qui vérifient une taille minimale (il est recherché un noyau le plus grand possible) suivant le rapport $\frac{\text{nombre de trous}}{\text{taille de la classe}}$.
Si le critère de définition d'une croyance approximativement vraie ou fausse est trop fort (le cas extrême étant de ne désigner par 1 que les croyances égales à V ou F), le tableau binaire correspondant contient de nombreux 0, la méthode de Govaert ne permet de ne définir que des concepts de petite taille. Inversement, si ce critère est trop faible, le tableau binaire contient de nombreux 1, il devient alors aisé d'en extraire des concepts qui forment cependant des hypothèses peu fiables.
La construction d'un noyau dans un schéma mental s'opère en trois étapes :

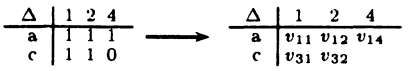
- retenir, dans le schéma mental initial, les éléments dont la valeur de croyance est approximativement vraie ou fausse, quel que soit l'ensemble de croyances utilisé



- le résultat est un tableau binaire duquel il faut extraire un bloc homogène maximum de 1,



- le concept ainsi défini permet de désigner un noyau, schéma mental pourvu des valeurs de croyances initiales.



Ce noyau représente une hypothèse sur le concept à apprendre. Le comportement de l'agent est tel que celle-ci reste consistante autant que possible.

5.2 Gestion

Rappelons la distinction entre connaissances manifestes (ou connaissances de base, désignées par l'ensemble M) et constructives (ensemble C) [5] : l'ensemble des connaissances d'un agent est vu comme une fonction F de M déterminant C. De cette distinction nous définissons un cycle de construction de la connaissance pour un agent (figure 7) : un expert enrichit la base de connaissances de l'agent ; l'ensemble ainsi constitué forme l'ensemble des connaissances manifestes ; l'application des syllogismes, tels qu'ils ont été définis (section 3), permet d'extraire un ensemble de connaissances constructives parmi les différentes solutions proposées ; enfin, certaines de ces dernières, jugées pertinentes, peuvent être intégrées aux connaissances de base. Ceci décrit le principe général de construction de la connaissance dans notre

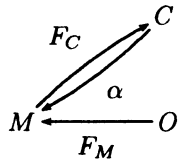


FIG. 7 – *Les connaissances constructives sont obtenues par application des syllogismes (F_C) sur les connaissances de base (M), constituées des connaissances d'un expert. Certaines d'entre elles seront désignées comme manifestes (α).*

système.

Cependant, notre objectif, dans un contexte de croissance incrémentale de la connaissance, est d'opérer des changements minimaux sur la base de connaissances constructives, c'est à dire conserver C tel que $C \in F_C(M)$ sans recalculer $F_C(M)$. Dans ce but, nous décomposons les connaissances manifestes en deux ensembles : les connaissances de base M, et les connaissances du noyau K, avec $K \subset M$. Nous souhaitons pouvoir aisément recalculer les valeurs possibles pour C par des changements incrémentaux du noyau, sans avoir à les recalculer à partir des connaissances de base.

La gestion de la croissance incrémentale de la connaissance dans notre système est réalisée par la base de connaissances d'un expert-oracle, O, un ensemble de connaissances de base, M, un noyau, K, et un ensemble de connaissances constructives, C, selon le processus suivant : M est modifié (change d'état) par l'apport des connaissances de l'oracle ; une application F_C sur K détermine l'ensemble des connaissances constructives C ; enfin, comme précédemment, certains des éléments de C peuvent être intégrés aux connaissances de base par une fonction α.
La procédure de construction du noyau (section 5.1) définit une fonction F_K, déterminant, à partir de M, les différents noyaux possibles. Mais, cherchant à obtenir une stabilité de K, le choix d'un noyau aussi proche que possible de l'état précédent parmi les alternatives de $F_K(M)$ est réalisé par la fonction μ_K. K est choisi par l'agent parmi $\mu_K(K, F_K(M))$. En résumé, nous avons $M \in \alpha(C, F_M(O))$, $K \in \mu_K(K, F_K(M))$, et $C \in F_C(K)$ (figure 8).

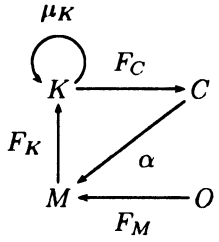


FIG. 8 – *Les connaissances manifestes sont décomposées en connaissances de base (M) et connaissances du noyau (K), maintenu aussi "stable" que possible.*

- F_M est une fonction d'apport de connaissances dans la base de l'apprenti.
- F_K est donc la fonction d'émergence d'un noyau parmi les connaissances de base M définie précédemment.
- F_C est une simple application des mécanismes de raisonnement sur les éléments du noyau pour définir des connaissances constructives.
- Les connaissances constructives sont considérées comme manifestes lorsque leur confiance dépasse un seuil donné. Ceci concerne donc des faits et règles dont la valeur de croyance est indifféremment proche de vrai ou faux, mais dont la confiance est suffisante.

μ_K doit modifier minimalement K en intégrant des informations nouvelles. Nous avons défini K afin que les croyances de cet ensemble soient monotones. La procédure de modification de K suit donc le principe

suivant : les nouveaux éléments insérés dans M et jugés cohérents (non contradictoires) avec le noyau sont intégrés à celui-ci, et les valeurs de croyance des éléments du noyau sont mises à jour (confortées) afin de considérer ce nouvel élément de connaissance ; si au contraire, un nouvel élément est contradictoire avec les connaissances du noyau, il n'y est pas inséré, et la mise à jour ne concerne alors que les connaissances de base.
Formellement, l'opérateur + désignant la borne supérieure au sens de la connaissance (selon la relation d'ordre $\leq_K$) de deux croyances, nous considérons qu'un fait $\Delta(x,y)$ confirme un autre fait $\Delta_K(x,y)$, élément du noyau, si la relation

$$v(\Delta(x,y) + \Delta_K(x,y)) \geq v(\Delta_K(x,y)), \text{ si } \Delta_K(x,y) \approx V$$
$$v(\Delta(x,y) + \Delta_K(x,y)) \leq v(\Delta_K(x,y)), \text{ si } \Delta_K(x,y) \approx F$$

est vérifiée (v est une fonction d'évaluation du niveau de connaissance d'une croyance). Autrement dit, l'information $\Delta(x,y)$ confirme l'information précédente $\Delta_K(x,y)$. Dans ce cas, la valeur de croyance du fait $\Delta_K(x,y)$ est remplacée par la valeur de croyance obtenue par la compilation des connaissances $\Delta(x,y)$ et $\Delta_K(x,y)$, soit

$$\Delta_K(x,y) = \Delta(x,y) + \Delta_K(x,y).$$

Pour deux noyaux K_1 et K_2 tels que $K_1 \subseteq K_2$ on voit simplement que l'ajout d'une information nouvelle dans la base de connaissances conserve cette inclusion. Cette propriété permet de vérifier en particulier que le seuil σ_v qui définit une croyance numérique approximativement vraie ou fausse n'est pas arbitraire : sa valeur influence directement la taille du noyau construit par la méthode d'extraction de concept, mais les résultats obtenus dans les différents noyaux possibles sont liés par inclusion.

Nous désignons comme *masqués* les éléments nouveaux non insérés dans le noyau. Lorsque le seuil définissant la proportion d'éléments masquables autorisée est dépassé, de trop nombreux éléments du schéma mental sont en contradiction avec le noyau, celui-ci est détruit, les hypothèses (obtenues par application des syllogismes) sont recalculées afin d'extraire une nouvelle structure de noyau plus cohérente avec l'ensemble des éléments du schéma mental.

6 CONCLUSION

Nous avons présenté une formalisation des théories semi-empiriques, modèle de représentation de connaissances incomplètes, capable de faire émerger et de maintenir un concept dans une base de données évolutive. Ce formalisme s'inspire de la théorie du noyau central, issue des recherches en psychologie sociale, décrivant l'organisation des connaissances autour d'un noyau stable maintenu par la présence d'éléments périphériques modulables. La propriété de stabilité et le processus de gestion des informations nouvelles, éléments essentiels de cette théorie, ont servi de fondements à l'élaboration d'un mécanisme de construction de la connaissance.

Cette méthode de gestion de la connaissance s'est imposée pour répondre au problème posé par l'ajout d'un nouvel élément susceptible de remettre en cause les hypothèses établies dans une base de connaissances. L'idée, déjà suggérée en d'autres termes par [28] ou [24], consiste donc à isoler un sous-ensemble de la base de connaissances dont les éléments semblent prendre sens ensemble et de faire évoluer monotoniquement les descriptions du concept ainsi exhibé. Les éléments extérieurs à ce concept sont supposés non significatif pour le processus d'apprentissage en cours.

Nous proposons donc une solution au problème de la gestion de la croissance incrémentale de la connaissance dans notre système d'apprentissage automatique à partir d'exemples. Ce modèle définit un comportement d'un tel système face à l'ajout d'information contradictoire avec les hypothèses existantes, en préservant les connaissances de bases et dérivées.

Nous appliquons ce formalisme, dans une interaction avec un expert, afin qu'un agent dispose d'une base de faits et de règles solide pour la résolution de problème.

Nous définissons ainsi un type particulier d'agent, capable de maintenir une théorie en interaction avec des données (des observations) et des experts.

BIBLIOGRAPHIE

[1] Abric J.-C. et Mardellat R., "Les représentations sociales de l'artisan et de l'artisanat", in *Etudes et recherches du G.I.F.R.E.S.H.*, numéro 21, 1976.

[2] Addis T., "Knowledge organisation for abduction", in *Proceedings of the Interdisciplinary Information Technology Conference*, Bradfor University, 1988.

[3] Balacheff N., *Une étude des processus de preuve en mathématique chez des élèves de collège*, thèse de doctorat, Université Joseph Fournier, Institut National Polytechnique, Grenoble, 1988.

[4] Barboux C., *Contrôle par objections d'une théorie incomplète*, thèse de doctorat, Université Montpellier II, mars 1990.

[5] Doyle J., "Constructive belief and rational representation", *Computational intelligence*, 5(1):1–11, 1989.

[6] Flament C. "L'étude des représentations sociales", in *L'analyse de similitude : une technique pour les recherches sur les représentations sociales / (sous la dir. de) Doise W., Palmonari A.*, Textes de base en psychologie, pages 138–156. Delachaux et Niestle, Neuchatel, Paris, 1986.

[7] Ganascia J.-G., "AGAPE : de l'appariement structurel à l'apprentissage", *Intellectica*, 1(2/3):6–27, 1987.

[8] Gennari J. H., Langley P., et Fisher D., "Models of incremental concept formation", *Artificial Intelligence*, 40:11–61, 1989.

[9] Ginsberg M. L., "Multivalued logics : A uniform approach to inference in artificial intelligence", Technical report, Department of Computer Science, Stanford University, 1991.

[10] Govaert G., "Classification simultanée de tableaux binaires", in *Data analysis and informatics III*, Diday E., Jambu M., Lebart L., Pages J., et Tomassone R. éditeurs, pages 223–236, Elsevier Science, 1984.

[11] Hanson S. J. et Bauer M., "Conceptual clustering, categorization, and polymorphy", *Machine Learning*, 3(4):343–372, 1989.

[12] Hempel C. G., "Deductive-nomological vs statistical explanation", in *Minnesota Studies in the Philosophy of Science III*, Feigl H. éditeur, pages 98–169, University of Minnesota Press, 1961.

[13] Jodelet D., *Les représentations sociales.* Sociologie d'aujourd'hui. Presses Universitaires de France, Paris, 1989. Sous la dir. de Denise Jodelet.

[14] Kruskal J. B., "On the shortest spanning subtree of a graph and the travelling salesman problem", in *Proceedings of the American Mathematical Society*, volume 7, pages 48–50, février 1956.

[15] Kyburg H. E., "Believing on the basis of the evidence", Technical Report TR 438, University of Rochester, Department of Computer Science, septembre 1992.

[16] Kyburg H. E., "Uncertain inferences and uncertain conclusions", in *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI-96)*, pages 365–372, Portland, Oregon, 1996.

[17] Lakatos I., *Preuves et Réfutations : Essai sur la logique de la découverte mathématique.* Hermann, 1984.

[18] Levesque H. J., "A logic of implicit and explicit belief", in *American Association for Artificial Intelligence. Fourth National Conference on Artificial Intelligence*, pages 198–202, Austin, Texas, 1984.

[19] Mephu Nguifo E., *Concevoir une abstraction à partir de ressemblances*, thèse de doctorat, Université Montpellier II, mai 1993.

[20] Michalski R. S., "Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts", *Journal of Policy Analysis and Information Systems*, 4(3):219–244, septembre 1980.

[21] Michalski R. S. et Stepp R. E., "Automated construction of classifications : Conceptual clustering versus numerical taxonomy", *IEEE Transactions on pattern analysis and machine intelligence*, PAMI-5(4):396–409, juillet 1983.

[22] Moscovici S., *La psychanalyse, son image et son public*, Bibliothèque de psychanalyse. Presses Universitaires de France, Paris, 1961.

[23] Pierre D., *Formulation et maintenance d'une théorie hypothétique par un agent apprenant*, thèse de doctorat, Université Montpellier II, juin 1997.

[24] Richards B. L. et Mooney R. J., "Automated refinement of first-order horn-clause domnain theories", *Machine Learning*, 19:95 – 133, 1995.

[25] Sallantin J., Barboux C., Quinqueton J., et Aubert J.-P., "Les théories semi-empiriques : éléments de formalisation", *Revue d'Intelligence Artificielle*, 5(1):69–92, 1991.

[26] Sallantin J., Szczeciniarz J.-J., Barboux C., Lagrange M.-S., et Renaud M., "Les théories semi-empiriques : conceptualisation et illustration", *Revue d'Intelligence Artificielle*, 5(1):9–67, 1991.

[27] Schank R. C. et Abelson R. P., *Scripts, plans, goals and understanding : an inquiry into human knowledge structures*, The artificial intelligence series. N.J. Hillsdale : Erlbaum, 1977.

[28] Stepp R. E. et Michalski R. S., "Conceptual clustering of structured objects : A goal-oriented approach", *Artificial Intelligence*, 28:43–69, 1986.

[29] Wang P., "Confidence as higher order uncertainty", Technical Report 93, Center for Research on Concepts and Cognition (CRCC), Indiana University, 1994.