THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : **Informatique**

présentée par

Christophe BROUARD

pour obtention du titre de
DOCTEUR DE L'UNIVERSITÉ PARIS 6

Sujet de la thèse :

Construction et Exploitation de Réseaux Sémantiques Flous pour l'Extraction d'Information Pertinente : le système RELIEFS

Soutenue le 7 Janvier 2000
devant le jury composé de :

Mme Bernadette BOUCHON-MEUNIER	Directrice de thèse
M. Charles TIJUS	Co-Directeur
Mme Anne DOUCET	Examinateur
M. Daniel KAYSER	Rapporteur
M. Jian Yun NIE	Rapporteur

T

Remerciements

Je tiens à remercier ma directrice de thèse, Bernadette Bouchon-Meunier, pour tous ses conseils, commentaires et critiques. Je la remercie pour la liberté qu'elle m'a laissée tout en me donnant des éléments me permettant d'une part, d'amorcer ma recherche et d'autre part, de faire des liens avec différents travaux en logique floue.

Je remercie Charles Tijus, mon co-directeur, pour sa disponibilité, pour son enthousiasme, pour les riches discussions que nous avons eu. Je suis très heureux d'avoir pu partager ses problématiques et celles de son équipe. Cette thèse n'aurait pas eu pour moi la même dimension sans cet échange.

Je tiens à remercier Jian-Yun Nie, pour son accueil chaleureux, et pour l'intérêt immédiat qu'il a manifesté pour mon travail. Quand je l'ai contacté, je ne mesurais pas encore la proximité de nos recherches. Je suis maintenant convaincu de la fertilité de notre future collaboration.

Je remercie Daniel Kayser d'avoir accepté d'être rapporteur. Il fût mon professeur à l'Université Paris Nord et ses travaux sur la représentation des connaissances et sur les sciences cognitives en font pour moi un interlocuteur privilégié. Je suis très heureux et très honoré qu'il fasse partie de mon jury.

Je remercie Anne Doucet d'avoir bien voulu présider ce jury. Son expertise en bases de données permettra très certainement d'ouvrir des perspectives intéressantes.

L'équipe LOFTI a constitué un cadre de travail agréable, un lieu d'échanges et de discussions. Je tiens à remercier tous ses membres en commençant par nos deux « anciens » Herman et Louis et en poursuivant par les membres présents et passés que j'ai eu le plaisir de côtoyer, Andréas, Anne, Christophe (M.), Leïla, Nathalie, Nédra, Maria, Martin, Mohamed et Virginie.

T

Il est souvent difficile de se partager entre deux équipes. La gentillesse, l'ouverture et la bonne humeur de l'ensemble des membres de l'équipe processus cognitif de Paris 8 m'a rendu les choses plus faciles. Je tiens à remercier en particulier ceux avec lesquels j'ai travaillé et dont les compétences ont constitué pour moi une aide précieuse, Elizabeth, Isabel, Jean-Marc et Sébastien, ainsi que ceux avec lesquels j'ai partagé l'aventure des congrès de l'ARC'98 et de CogSci'98, Béatrice, Christine, Denis C., Denis L., Elisabetta, Emanuel, Mojdeh, Séverine.

En parlant de bonne humeur, il est impensable de ne pas mentionner les membres du personnel administratif du LIP6, Andrée, Jacqueline, Ghislaine, Valérie, ainsi que Marie-Claire, notre ingénieur système.

Je remercie Denis Coulombier, Daniel Levy-Bruhl et Anne Perrocheau de m'avoir fait partager avec enthousiasme leur expérience de l'analyse de données épidémiologiques.

Je remercie mes deux compères moniteurs de Créteil, Christophe et Gwenaël pour les nombreuses « poilades » que nous avons partagées. Je remercie également Danielle Beauquier, Alessandra Carbone et Michel Spyratos pour leur initiation au métier d'enseignant.

Enfin, je remercie Caroline pour sa relecture du document, pour ses commentaires, et pour son soutien en général, sans lequel, tout eût été plus difficile.

TABLE DES MATIERES

Chapitre 1

> *Aussi, avoir une mémoire fidèle et procéder d'après le livre sont des points qui constituent pour le vulgaire le summum du bien jouer. Mais c'est dans les cas situés au-delà de la règle que le talent de l'analyste se manifeste ; il fait en silence une foule d'observations et de déductions. Ses partenaires en font peut-être autant ; et la différence d'étendue dans les renseignements ainsi acquis ne gît pas tant dans la validité de la déduction que dans la qualité de l'observation. L'important, le principal est de savoir ce qu'il faut observer. Notre joueur ne se confine pas dans son jeu, et, bien que ce jeu soit l'objet actuel de son attention, il ne rejette pas pour cela les déductions qui naissent d'objets étrangers au jeu.*
>
> Edgar Allan Poe - *Histoires extraordinaires*

1 INTRODUCTION GENERALE

La pertinence : Une notion centrale

Cette thèse traite de la notion de pertinence, qui est d'un avis largement partagé, une notion centrale pour l'Intelligence Artificielle (Subramanian & al., 1997). D'un point de vue cognitif, la pertinence est aussi décrite comme un concept clef (Anderson, 1983a) et certains en font même un principe qui régit nos interactions avec l'environnement ou encore la façon dont nous communiquons (Sperber & Wilson, 1989). Plus généralement, on comprend que toute résolution de problème présuppose la sélection des informations à prendre en considération, avant même leur utilisation dans un raisonnement quelconque, car la prise en compte de l'ensemble des informations rendrait le raisonnement totalement inefficace (contrainte de temps et d'espace).

Actuellement, le foisonnement de données informatiques et le développement d'Internet en particulier ont encore renforcé l'intérêt pratique de l'étude de cette notion. L'information est là, mais encore faut-il pouvoir l'extraire des gigantesques bases de données et la question récurrente est : quelle information est pertinente ?

L'utilisateur de système informatique est démuni s'il ne dispose pas d'outils lui permettant de faire face à un important volume de données. Le développement de champs de recherche tels que la Fouille de Données (*Data-Mining*) et la Recherche d'Information (*Information Retrieval*) témoignent de l'intérêt actuel pour cette problématique. Dans le cas de la Fouille de Données (Holsheimer & Siebes, 1994), qui touche différentes disciplines comme l'Analyse de Données, les Bases de Données, et l'Intelligence Artificielle, on tente de fournir à l'utilisateur une vue synthétique des données, en extrayant de celles-ci, les régularités les plus fortes (par exemple, sous forme de règles « SI *condition* ALORS *conclusion* »). Dans le cas de la Recherche d'Information (Salton, 1983 ; Schaüble, 1998), on s'applique à sélectionner les documents adaptés à un besoin d'information exprimé par l'utilisateur au moyen d'une requête (on peut penser par exemple, à la recherche par mots-clefs).

Ce qui distingue l'homme de la machine

Bien que les objectifs de l'Intelligence Artificielle ne se limitent pas à l'imitation de la pensée humaine (Ford & Hayes, 1998), s'il est une capacité où l'homme apparaît comme un modèle qui pourrait profiter à la machine, il semble bien s'agir de son habileté à discerner ce qui est pertinent de ce qui ne l'est pas.

L'exemple le plus fréquemment cité pour illustrer cette idée est celui des jeux et en particulier celui de Deep Blue, système informatique dédié au jeu d'échecs qui examine environ deux millions de positions par seconde et qui est en mesure de battre le champion du monde d'échecs depuis 1997. On peut en effet s'étonner qu'une telle quantité de calcul soit nécessaire pour rivaliser avec un être humain qui n'étudie véritablement qu'une poignée de solutions (des études montrent qu'il en étudie tout au plus une centaine). Ainsi, le champion du monde d'échecs de 1921 à 1927, Raoul Capablanca, à la question portant sur le nombre de coups envisagés dans l'analyse d'une position d'échecs répondit : « Un seul, mais le bon ». Aussi, on a coutume d'opposer la recherche « bêtement » calculatoire des machines et le discernement dont fait preuve le champion que l'on tient pour être le signe d'une véritable intelligence. Sans renvoyer à la notion mal définie d'intelligence, on peut simplement constater la différence évidente dans les stratégies utilisées par l'homme et la machine (de force équivalente) et la supériorité de l'homme à sélectionner les solutions pertinentes en élaguant largement l'espace de recherche compensant ainsi son manque de rapidité.

Dans le cadre de l'Intelligence Artificielle, on s'autorise à introduire cette intelligence humaine, faite d'heuristiques, d'approximations, d'imperfections qui permettent de contourner le problème du calcul exhaustif d'une solution. Notre travail se situe dans ce cadre et en particulier dans celui de la logique floue spécifiquement dédiée au traitement de données imprécises et/ou incertaines. A l'heure actuelle, il n'existe pas de mécanismes de sélection de l'information suffisamment généraux pour être appliqués à différents types de problèmes et on peut considérer que la difficulté pour l'Intelligence Artificielle à concevoir des systèmes ne se limitant pas à des domaines très circonscrits (où l'information pertinente est en quelque sorte pré-sélectionnée) est liée pour une grande part à ce manque.

Une notion difficile à formaliser

La pertinence est une notion humaine qui ne se laisse pas cerner facilement. Ce constat a conduit certains à la juger trop subjective ou trop ambiguë pour être définie de façon formelle. Elle est d'autre part très contextuelle, le moindre changement dans la situation pouvant introduire un jugement de pertinence très différent. Il est aussi très difficile d'énumérer de façon exhaustive l'ensemble des critères pouvant jouer un rôle dans le jugement de pertinence. La notion de pertinence est en fait une notion très souvent évoquée et même utilisée dans différentes argumentations, bien qu'elle n'ait pourtant jamais été définie de façon formelle dans toute sa généralité sans tomber dans la paraphrase. Ainsi, bien que la définition proposée par Sperber et Wilson (1989) soit certainement la plus générale et constitue une proposition concrète et bien argumentée, elle reste empirique et ne donne pas les bases suffisantes à la définition d'algorithmes permettant de sélectionner l'information pertinente. Notre sujet porte sur la recherche d'une formalisation de la pertinence qui puisse être implémentée par un programme informatique.

Articulation du document

Pour cerner ce que recouvre la notion de pertinence, la démarche n'a pas été de prendre un problème particulier et de faire une micro-analyse de la notion de pertinence sur ce problème. La méthode suivie consiste au contraire, à tenter d'extraire une synthèse à partir des développements faits autour de cette notion dans différents domaines de recherche.

Dans le deuxième chapitre, nous analysons donc la notion de pertinence sous ces différentes facettes à travers différentes approches, pour finalement proposer une vue synthétique de cette notion.

Cette vue nous conduit dans le troisième chapitre, à considérer un certain nombre de systèmes de représentation des connaissances développés en Intelligence Artificielle qui soient en mesure d'extraire une structure rendant compte de l'ensemble des régularités de l'environnement et qui permettent de les exploiter dans une situation donnée pour sélectionner l'information pertinente. Parmi ces méthodes, nous nous intéressons en particulier à celles relatives aux réseaux sémantiques vers lesquelles s'oriente notre recherche.

Le quatrième chapitre concerne la présentation du système RELIEFS, système de construction et d'exploitation de réseaux sémantiques flous que nous proposons.

Dans le cinquième chapitre, nous utilisons ce système de trois façons différentes, dans trois domaines différents. Nous l'appliquons ainsi, à l'interrogation d'une base de données relative à la surveillance médicale de population (données épidémiologiques) puis, à l'analyse de données sensorielles et enfin à l'interprétation de commandes verbales.

Dans le sixième et dernier chapitre, nous concluons en résumant le travail effectué et en donnant nos perspectives.

Chapitre 2

Il faut être deux pour inventer. L'un forme des combinaisons, l'autre choisit, reconnaît ce qu'il désire, et ce qui lui importe dans l'ensemble des produits du premier.

Ce que l'on appelle le « génie » est bien moins l'acte de celui-ci (celui qui combine) que la promptitude du second à comprendre la valeur de ce qui vient de se produire et à saisir ce produit.

Paul Valéry – *De la simulation*

2 UNE APPROCHE PLURIDISCIPLINAIRE DE LA NOTION DE PERTINENCE

On trouve dans ce chapitre une étude de la notion de pertinence. L'exposé commence par une brève étymologie et une série de questions spontanées sur ce que peut recouvrir cette notion. Il se poursuit par une revue des différentes approches de la pertinence comprenant un point de vue psychologique, un exposé de ce qu'on entend par « pertinence » en Intelligence Artificielle (IA), en Recherche d'Information (RI), en Logique et en Linguistique. On se place ici à un certain niveau de généralité s'intéressant plus à la façon dont la notion est définie qu'au détail des solutions apportées aux différents problèmes posés. Un exposé plus technique des différents systèmes de représentation dédiés à l'extraction d'information pertinente sera donné dans le chapitre suivant. Enfin, une synthèse extrait les points communs entre ces différentes approches, met en avant une approche qui apporte des éléments de réponses aux questions initialement posées et finalement dégage les grands principes d'un système d'extraction d'informations pertinentes.

2.1 Etymologie

Selon le dictionnaire étymologique (Baumgartner & Ménard, 1996), on attribue à l’adjectif pertinent et au nom pertinence, le sens et l’origine suivants :

Pertinent, emprunté au latin pertinens « qui concerne, relatif à, qui vise à », employé en France dès le moyen age (XIV$^{\text{ème}}$ siècle) au sens de « qui a rapport à, approprié à ».

Pertinence, emprunté au latin impertinens « sans rapport avec », d’abord en ce sens en français, puis au sens de inconvenant (XV$^{\text{ème}}$ siècle) et insolent (XVII$^{\text{ème}}$ siècle).

L’étymologie des mots présentés comme synonymes dans le même dictionnaire est :

Relatif (XIII$^{\text{ème}}$ siècle), emprunté au latin relativus « qui est en rapport avec », employé en français au sens de « qui présente une relation avec quelque chose », « qui n’est ni absolu, ni indépendant », d’où « incomplet imparfait ».

Concerner (XIV$^{\text{ème}}$ siècle), emprunté au latin concernere « concerner, considérer », dérivé de cernere « distinguer, discerner ».

Approprié (XII$^{\text{ème}}$ siècle) renvoie à propre, emprunté au latin proprius « qui n’appartient qu’à soit, spécial, caractéristique ».

2.2 Quelques questions

Nous commençons par une série de questions volontairement spontanées et allusives. Ces questions constituent une grille de lecture de la suite du chapitre, les différentes approches de la pertinence venant apporter des éléments de réponses. Ces éléments seront utilisés par la suite pour formuler une synthèse qui tentera de répondre à ces différentes interrogations.

Qu'est-ce que la pertinence ? Proposer une définition, un modèle… ?

Nous allons dans ce chapitre, tenter de cerner ce que recouvre cette notion et nous poser la question de sa conceptualisation et de sa modélisation.

Ce qui ressemble ou ce qui est différent ?

Considérons deux situations opposées :

1- Dans le cas d'une tâche de surveillance (épidémies, pollution, activités sismiques…), on s'intéressera en particulier à des données anormales, des changements par rapport à des situations habituelles.

2- Dans d'autres cas, comme dans le cas de la recherche documentaire par exemple, c'est au contraire ce qui est similaire à ce qu'on cherche qui est digne d'intérêt.

Comment mesurer une distance ou une similarité ?

Il est toujours délicat de justifier le choix d'une mesure de distance ou de similarité, même si l'on peut dégager de la tâche à réaliser des indications sur les propriétés de la mesure de comparaison qui convient et que l'on peut déduire de ces indications une famille spécifique de mesures (Rifqi, 1996).

Plus globalement, quels critères pour l'évaluation ?

L'objectif, c'est-à-dire l'utilisation que l'on veut faire des informations, entre indubitablement dans l'évaluation de leur pertinence. Par exemple, la position du soleil dans le ciel peut être une information pertinente si l'objectif est de s'orienter. Néanmoins, d'autres critères comme, les connaissances préalables (encore faut-il connaître le lien entre cette position et l'heure de la journée), ou encore le contexte courant dans lequel se situe l'évaluation (cette information est inutile si l'on dispose d'une boussole) semblent jouer un rôle important.

Qui juge de la pertinence ?

Faut-il définir une mesure de pertinence objective, c'est-à-dire directement calculable par le système à partir de ses connaissances, ou tenter de modéliser la pertinence d'une information pour un utilisateur donné avec toute la subjectivité que cela comporte ?

On peut penser que l'utilisateur n'a pas nécessairement toutes les connaissances lui permettant de distinguer ce qui est pertinent et donc pencher en faveur d'une mesure objective surtout si l'objectif est de lui permettre d'acquérir de nouvelles connaissances. Cependant, encore faut-il que l'utilisateur puisse formuler ses requêtes dans un langage qu'il maîtrise et qu'il puisse aussi intégrer les informations extraites par le système. Il semble donc important d'inclure dans le système un modèle de l'utilisateur. Plus fondamentalement, la question porte sur l'objectivité des connaissances. Existe-t-il des connaissances qui ne soient pas empruntes d'une certaine subjectivité ?

Sélection ou synthèse d'information ?

On pense naturellement lorsque l'on parle d'extraction d'information pertinente à la sélection d'information, c'est-à-dire à un tri permettant de retenir parmi un grand nombre d'informations celles qui satisfont un critère donné. Mais le résumé d'un texte ou la découverte d'une loi physique ne constituent-t-ils pas une extraction d'information pertinente au sens d'une information utile et synthétique ?

Unicité du sens ?

On voit se dégager différents sens de la notion de pertinence (différent/identique, synthèse/sélection). Peut-on en présenter une modélisation unique ? Ou doit-on définir différents sens, en recherchant pour chacun, un modèle particulier ?

2.3 Différentes approches

L'objectif de cette partie n'est pas de faire un catalogue exhaustif de la notion de pertinence, mais plutôt d'examiner ce à quoi elle peut correspondre à travers différents domaines dans lesquels elle a donné lieu à des recherches approfondies, pour néanmoins, en extraire une vue synthétique (on notera que même si le lien n'est pas toujours complètement explicité dans un premier temps, l'ensemble des idées exposé est repris dans la synthèse ou dans la suite du document).

Partant de la remarque que notre système cognitif semble être un système d'extraction d'information pertinente performant et qu'il constitue une source d'information intéressante pour notre problème, nous examinerons dans un premier temps des travaux de recherche en Psychologie Cognitive portant sur la mémoire et l'attention. L'organisation des connaissances en mémoire et leur mode d'accès sont en effet directement liés au problème de la sélection et du résumé. De même, l'attention peut être vue comme un mécanisme de sélection de l'information permettant de mobiliser des ressources de traitement pour les informations focalisées au détriment d'autres informations (conversation dans un environnement bruyant par exemple).

Dans un second temps, nous examinerons la pertinence du point de vue de sa modélisation en nous tournant vers les sciences de l'information. L'Intelligence Artificielle et la Recherche d'Information sont deux champs de recherche visant à concevoir des systèmes informatiques dans lesquels la notion de pertinence joue un rôle particulièrement important.

En Intelligence Artificielle, l'un des objectifs fondamentaux est de tendre vers des systèmes intelligents, ne consistant pas en une algorithmique rigide mais capables de traiter un grand nombre de problèmes en faisant face à la diversité des situations rencontrées. Plus la gamme de problèmes que le système est en mesure de résoudre est étendue, plus la quantité de connaissances intégrées au système est considérable et un pré-requis à tout traitement est : "Quelles informations vais-je utiliser pour résoudre le problème X sachant que je suis dans une situation Y ?"

En Recherche d'Information, c'est encore plus explicite, le problème est précisément de sélectionner le document pertinent vis-à-vis d'une requête. La notion de pertinence est donc centrale, et c'est d'ailleurs dans ce champ de recherche que cette notion a fait l'objet du plus grand nombre d'investigations.

Puis, nous appuyant sur des exemples pratiques définis dans ces deux domaines, nous aborderons la pertinence du point de vue plus formel de la logique qui sous-tend certains de ces systèmes.

Enfin, comme nous le verrons cette notion est intimement liée à celle de communication, nous décrirons donc brièvement les travaux de Sperber et Wilson (1989) réalisés en linguistique sur ce thème.

2.3.1 Pertinence et cognition

La pertinence est une notion humaine et de ce point de vue, comprendre le système cognitif et identifier d'une façon plus précise ce à quoi peut correspondre cette notion est primordiale. On peut aussi considérer les mécanismes cognitifs qui régissent l'extraction d'information pertinente comme des modèles dont il est possible de s'inspirer pour la conception de systèmes artificiels. En

Psychologie Cognitive, les travaux sur la mémoire sémantique et ceux sur l'attention ont un lien direct avec la notion de pertinence.

2.3.1.1 La mémoire sémantique

La mémoire est un vaste champ de recherche en Psychologie Cognitive. On distingue aujourd'hui différents types de mémoires, par exemple, la mémoire à court terme et la mémoire à long terme sur la base de la durée du stockage, la mémoire procédurale et la mémoire déclarative sur la base de la nature de l'information stockée. Notre objectif est simplement ici de pointer sur des travaux généraux liés à l'organisation des connaissances, et leur mode d'accès.

L'un des premiers modèles de la mémoire sémantique est dû à Ross Quillian (1968) qui propose de représenter l'organisation des connaissances en mémoire par un réseau de concepts et qui définit une typologie des différents liens entre concepts. Dans ce réseau, la définition de la notion de concept est très générale, elle peut s'appliquer par exemple à des objets, des actions, des propriétés. Deux concepts peuvent être reliés par de nombreux liens et, à partir de ce réseau, en considérant que plus il y a de liens entre deux concepts, plus il sont liés sémantiquement, Quillian définit la notion de relation (*relatedness*) entre concepts (figure 2.1).

L'idée générale repose sur le fait que le sens d'un concept émerge de sa relation avec les autres concepts du réseau. Pour reprendre l'exemple donné par Damasio et Damasio (1992), "une tasse de café évoque à la fois des représentations visuelles ou tactiles de sa forme, de sa couleur, de sa texture, de sa température, l'odeur et le goût du café, ainsi que la trajectoire de la main et du bras pour porter la tasse aux lèvres".

Pour rendre compte de la façon dont cette mémoire est exploitée, il définit aussi un ensemble de mécanismes de propagation de l'activation d'un nœud du réseau vers les autres nœuds du réseau (Quillian , 1968 ; Collins & Loftus, 1975).

Cette théorie rend compte de nombreux résultats expérimentaux concernant l'accès des connaissances en mémoire. Elle rend compte par exemple, des résultats des expériences d'amorçage qui étudient l'effet de la présentation d'un objet (l'amorce) sur le traitement d'un objet associé (la cible). Ainsi, le traitement du mot « chien » (dans une tache de décision lexicale qui consiste à dire si le mot présenté existe dans la langue française) est facilité (on observe des temps de réponse plus courts) s'il est précédé du mot chat (voir Neely, 1991, pour une revue globale du phénomène d'amorçage). Dans ce cas, la facilitation peut s'expliquer par une pré-activation de la cible par l'amorce.

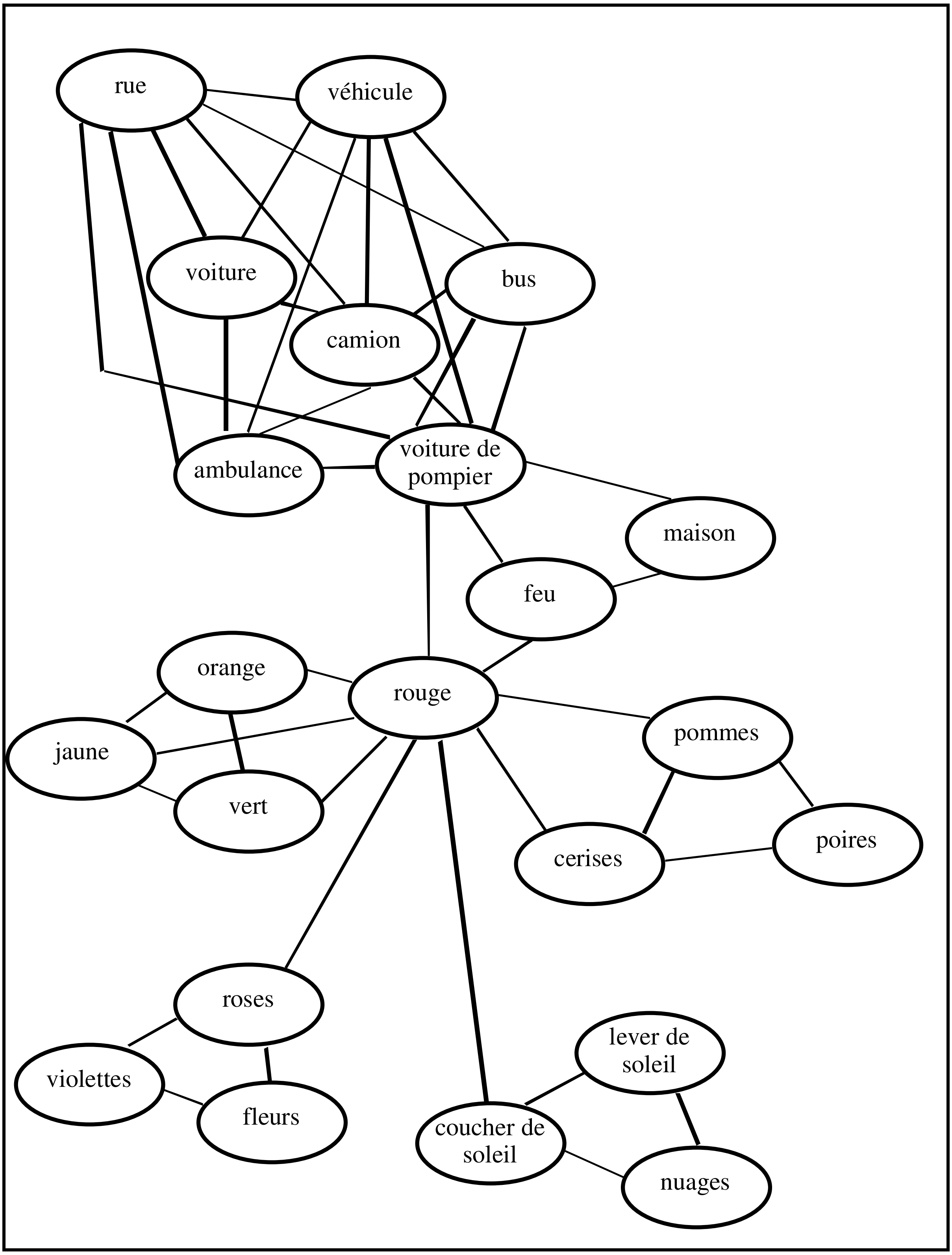


FIG. 2.1 Une représentation schématique des relations entre concepts d'un fragment de la mémoire humaine (plus le lien est court, plus la relation est forte). Extrait de Collins et Loftus (1975).

Cette théorie est aussi en accord avec un résultat célèbre portant sur les temps de recherche en mémoire (Collins & Quillian, 1969) qui montre que le temps de réponse à une question du type "A est-il un B ?" dépend de la longueur du chemin à parcourir dans la hiérarchie pour relier le concept A au concept B, comme on peut le prévoir si on considère que la réponse repose sur une propagation de l'activation entre les nœuds du réseau représentant respectivement A et B. Enfin on peut aussi citer, les travaux de Kintsch (1988) qui reprennent cette théorie dans le cadre de l'interprétation du langage.

Indépendamment de la plausibilité psychologique de cette théorie, il est intéressant de noter que ce mode de fonctionnement permet une mise à disposition automatique de la connaissance liée par exemple aux objets perçus dans le contexte courant, ou bien à l'objectif, donc à la connaissance susceptible d'être pertinente. C'est pour cette raison que Quillian considère que le mécanisme de propagation d'activation est bien adapté à l'Intelligence Artificielle. Il évite ou réduit le processus coûteux de la recherche de connaissance particulièrement problématique lorsque les bases de données sont importantes (Anderson, 1983a, p. 88)

Anderson (1983a ; 1983b) reprend ces principes dans son modèle général de la cognition (appelé ACT). Il insiste sur l'importance du processus de sélection d'information : *"One of the key factors in human intelligence is the ability to identify and to utilize the knowledge that is relevant to a particular problem"* . Il en rend compte par la propagation d'activation dans un réseau sémantique représentant les connaissances déclaratives (savoir). Les connaissances procédurales (le savoir-faire) sont modélisées par un ensemble de règles de production qui sont déclenchées suivant le degré d'activation des prémisses. A chaque instant, trois sources d'activation sont considérées : le but, l'environnement perçu, le résultat des règles de production.

Ce modèle rend ainsi compte, en quelque sorte, du « fil de la pensée ». Par exemple, si l'objectif est la survie à l'attaque d'un prédateur et l'environnement est celui d'une forêt, le nœud représentant la forêt et celui représentant la survie sont activés. L'activation se propage respectivement au nœud représentant l'arbre et celui représentant la fuite (mais pas seulement, les nœuds "bois" et "se cacher" peuvent aussi être activés). Ainsi, la règle qui dans le cas de la présence d'un arbre et dans l'objectif d'une fuite conclut de grimper à l'arbre a de forte chance d'être déclenchée puisque la prémisse est fortement activée.

Le réseau sémantique dans lequel s'effectue la propagation est de nature associative. Ainsi, les éléments représentés dans les nœuds du réseau (*chunks*) sont reliés soit manuellement, soit suivant le principe suivant (Wallach & Lebière, 1998) :

« la force d'association S_{ij} entre les éléments (*chunks*) C_i et C_j est une mesure de la fréquence avec laquelle C_j est utile quand C_i est un élément du but ».

En fait, la nature associative de la mémoire est une hypothèse largement répandue et ancienne puisque déjà Aristote considérait qu'une condition suffisante pour que deux événements soient liés en mémoire était qu'ils se déroulent simultanément. De ce point de vue, les travaux sur l'apprentissage associatif (Pavlov, 1927 ; Le Ny, 1972) constituent une source intéressante, de même que ceux de Hebb (1949) qui a montré la réalité de ce principe au niveau neurophysiologique. Ainsi, il remarque que "Quand un axone de la cellule A est assez proche pour exciter une cellule B et quand, de façon répétée et persistante, il participe à son activation, un certain processus de croissance ou un changement métabolique s'installe dans une cellule ou dans les deux tel que l'efficacité de A, en tant qu'elle est une des cellules qui active B, est augmentée".

Enfin, parce que l'accès à la connaissance est conditionnée par son organisation, les travaux portant plus spécifiquement sur la catégorisation, c'est-à-dire la faculté de construire des catégories pour regrouper différents objets et construire des hiérarchies de catégories sont aussi des références intéressantes. Les questions portent par exemple, sur la représentation des catégories (conjonction de propriétés ou prototypes) et sur la façon dont elles sont construites. L'intérêt des catégories réside dans l'économie qu'elles permettent (Rosch, 1976). Non seulement, elle minimise la quantité de connaissance à stocker en opérant des regroupements (plutôt que de mémoriser chaque observation individuellement), mais elle permettent aussi, à partir d'un petit nombre de propriétés perçues pour un objet quelconque, de lui attribuer toutes les autres propriétés attachées au concept auquel se rattache l'objet. De façon générale, l'organisation des connaissances facilite les traitements sur ces connaissances. Il a par exemple été démontré expérimentalement que les sujets rappelaient un plus grand nombre de mots lorsqu'ils étaient regroupés en catégories sémantiques (Baddeley, 1992, pp. 198-199). Ces regroupements représentent un avantage pour mémoriser un plus grand nombre d'informations sachant que la capacité de la mémoire à court terme est limitée à 7 unités plus ou moins deux (Miller, 1956). Saracevic (1975, p. 329) considère d'ailleurs la limitation de la mémoire à court terme comme étant de première importance vis-à-vis de la notion de pertinence.

2.3.1.2 L'attention

L'attention peut être étudiée de deux façons : soit en tant que processus réglant (quels changement interviennent dans le traitement d'une information suivant le degré d'attention que je lui accorde ?) ou processus réglé (qu'est ce qui explique qu'elle puisse subitement changer de cible pour porter sur un nouvel objet ?) (Camus, 1996).

Dans le premier cas (processus réglant), de nombreuses expériences convergent sur la facilitation du traitement d'information attendue (figure 2.2) et l'altération du traitement de l'information inattendue. Il est question, selon les auteurs, d'une profondeur de traitement accrue pour les informations attendues ou encore d'une sensibilité plus grande.

Toujours dans cette perspective de processus réglant, l'attention fait aussi l'objet d'études en neurophysiologie. Certaines expériences ont par exemple montré une amplification du niveau d'activité nerveuse manifestée dans certains sites cérébraux participant à l'élaboration de la représentation des stimuli engagés dans un processus d'attention (par rapport à une situation où ces stimuli sont présentés mais ne sont pas pertinents pour le système cognitif). Laberge (1995) insiste sur cette idée d'amplification pour envisager l'attention en terme de canal sélectif et d'une distribution préparatoire d'activité.

FIG. 2.2 Illustration de la facilitation du traitement liée à une information attendue. La détection de la présence d'un dalmatien dans ce dessin nécessite une recherche active. Elle est facilitée si la présence du dalmatien est révélée préalablement.

Dans le second cas (processus réglé), on peut s'intéresser aux conditions dans lesquelles certains objets « sautent aux yeux » sans qu'il faille les chercher de façon délibérée. C'est ce qu'on appelle le phénomène de "pop-out" qui a été étudié par Treisman (1992). Ces recherches ont notamment permis de constater que les objets se différenciant des autres objets sur une propriété simple sautaient aux yeux (un rond rouge parmi des ronds bleus par exemple) alors que lorsqu'il fallait considérer une conjonction (un rond rouge parmi des ronds bleus, des carrés rouges et des carrés bleus) ce n'était plus le cas. Il a également été montré qu'il y avait une asymétrie dans les temps de recherche d'objets se différenciant par l'absence d'un trait par rapport à des objets se différentiant par la présence d'un trait (figure 2.3), le second type d'objet étant retrouvé plus rapidement. La comparaison de différentes caractéristiques a aussi révéler que certaines propriétés étaient plus saillantes, c'est-à-dire permettaient de retrouver plus rapidement l'objet qui la portait (figure 2.4).

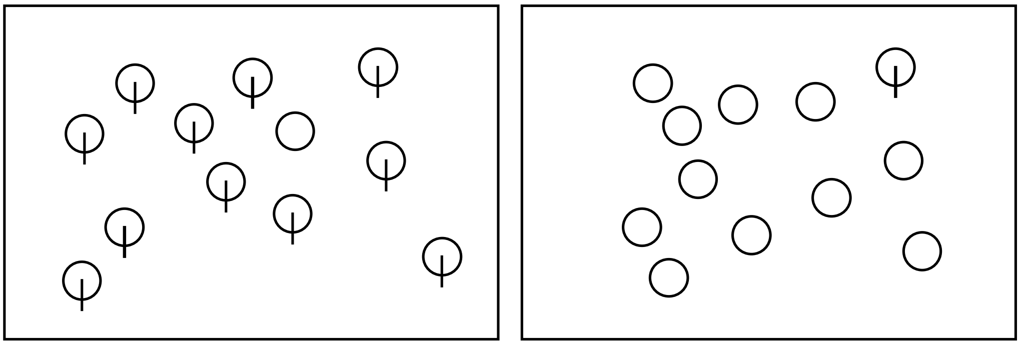


FIG. 2.3 Le temps de recherche est plus court pour l'objet qui se distingue par la présence d'un trait (à droite) que pour l'objet qui se distingue par l'absence d'un trait (à gauche).

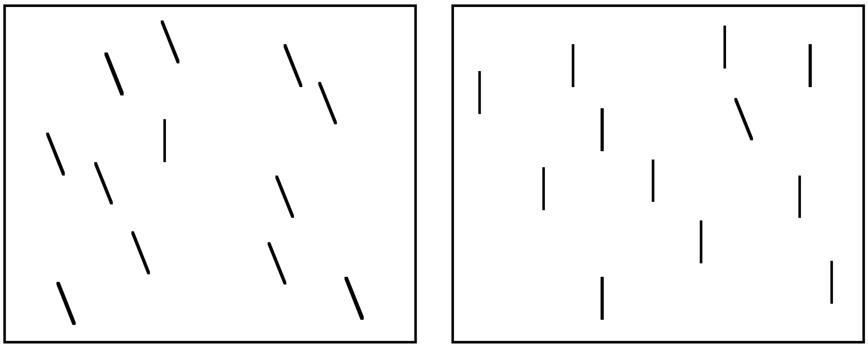


FIG. 2.4 Le temps de recherche est plus court pour un trait oblique (à droite) que pour un trait vertical (à gauche).

Ce double aspect (processus réglant/processus réglé) met en avant un principe intéressant pour la conception d'un système d'extraction d'information pertinente. Ce système doit en effet permettre de sélectionner l'information en fonction des objectifs (donc de jouer le rôle de filtre) et il doit aussi permettre de réagir à d'éventuels événements inattendus. Cela renvoie à l'idée d'une interaction entre un processus « descendant » c'est-à-dire guidé par des connaissances et un processus « ascendant » (guidé par les caractéristiques des stimuli). Nous reviendrons sur cette interaction dans la suite.

Par ailleurs, de nombreux modèles psychologiques de l'attention ont été proposés. Ainsi, l'attention est envisagée comme un filtre, comme un effort, comme un canal sélectif et une distribution d'activité préparatoire ou encore comme un projecteur permettant de regrouper des conjonctions de traits (Treisman, 1992). Ces différents modèles sont autant des sources d'inspiration pour la conception d'un système d'extraction d'information pertinente. Nous nous en tiendrons cependant ici aux principes généraux.

2.3.2 Pertinence et sciences de l'information

2.3.2.1 En Intelligence Artificielle

En IA, le problème de la pertinence se pose fréquemment et sous de multiples formes. Au point que dans l'introduction du numéro spécial de « Artificial Intelligence » consacré à la pertinence, Subramanian et al. (1997) parlent d'omniprésence. Que ce soit le choix des attributs susceptibles d'avoir un impact dans une décision, l'extraction de règles pour la décision, la sélection de la règle à appliquer, ou encore le choix d'un coup à jouer dans une partie d'échecs, différentes solutions ont été proposées. Nous les examinons brièvement dans cette partie dans l'objectif de montrer comment ces différents problèmes ont été abordés.

2.3.2.1.1 La sélection des attributs pertinents

Avant de décrire le problème de la sélection des attributs pertinents, nous décrivons celui plus général dans lequel il s'insère, c'est-à-dire celui de l'apprentissage inductif. Le principe d'induction consiste à généraliser une loi à partir d'un ensemble de cas particuliers. En IA, ce principe a été utilisé pour donner à des systèmes informatiques la faculté d'apprendre. Dans ce cadre, les lois peuvent, par exemple, prendre la forme de règles du type « SI condition ALORS conclusion » où les conditions et les conclusions des règles sont des combinaisons (conjonctions ou disjonctions) de couple attribut-valeur (du type COULEUR = bleue) que l'on nommera aussi propriétés. La construction des règles s'appuie sur un ensemble d'exemples (que l'on nommera aussi observations) consistant en une conjonction d'attribut-valeur. Un exemple peut être « FORME = carré ET COULEUR = bleue ». On se place aussi souvent dans le cadre d'un problème de classification dans lequel la conclusion des règles que l'on cherche à extraire porte sur un attribut particulier qui désigne la classe de l'exemple.

Le problème de la sélection des attributs consiste ici à décider parmi un ensemble d'attributs, lesquels sont pertinents et doivent donc être retenus pour être éventuellement combinés en vue de construire des règles concluant sur la classe de l'exemple. Ce problème est véritablement crucial lorsque l'on n'a pas d'idée a priori sur les attributs à prendre en compte, ce qui est le cas lorsque l'on se place par exemple, dans une problématique de découverte. Par rapport à cette problématique, différentes définitions de la pertinence ont été proposées. Blum et Langley (1997) dans leur synthèse sur ce type de problème commence par proposer la définition suivante :

Définition 1 (Pertinence vis-à-vis d'une classe) :

Un attribut X_i est pertinent vis-à-vis d'une classe C s'il existe une paire d'exemples A et B tels que A et B diffèrent uniquement sur l'attribut X_i et $C(A) \neq C(B)$ (la classe associée à A est différente de celle associée à B).

Selon cette définition, un attribut est donc pertinent si, pour un exemple donné, le fait de changer sa valeur influe sur la classe attribuée à l'exemple. Blum et Langley (1997) signalent les différents désavantages que cette définition comporte. Tout d'abord l'apprentissage est réalisé sur un échantillon, il est donc difficile d'évaluer correctement ce genre d'assertion qui nécessite l'existence d'un grand nombre d'exemples (puisque des paires ne doivent varier que sur un seul attribut). De plus, si certains attributs sont présents plusieurs fois dans la description d'un exemple, il devient impossible pour deux exemples de différer uniquement sur l'attribut redondant. Pour contourner ce problème John et al. (1994) proposent une notion de pertinence plus faible mais plus concrète, que Blum et Langley (1997) ramènent à la notion de pertinence vis-à-vis d'un échantillon (l'ensemble d'apprentissage). Ils distinguent une pertinence forte et une pertinence faible.

Définition 2 (pertinence forte vis-à-vis d'un échantillon) :

Un attribut X_i est fortement pertinent vis-à-vis d'un échantillon, s'il existe des exemples A et B dans l'échantillon qui diffèrent seulement sur X_i et qui appartiennent à des classes différentes.

On peut réécrire cette définition comme une dépendance conditionnelle de la classe par rapport à X_i. Soit, si on considère $S_i = \{X_1, \ldots, X_{i-1}, X_{i+1}, X_n\}$, l'ensemble des attributs excepté X_i : $P(C=c / X_i=x_i, S_i=s_i) \neq P(C=c/ S_i=s_i)$.

La seule différence par rapport à la première définition est que A et B sont dans l'échantillon, soit P(Xi=xi, Si=si) > 0.

Définition 3 (pertinence faible vis-à-vis d'un échantillon) :

Un attribut X_i est faiblement pertinent vis-à-vis d'un échantillon S, s'il est possible de supprimer un sous-ensemble d'attributs tel que X_i devienne fortement pertinent.

Ces deux définitions peuvent être utilisées pour caractériser le type d'attributs que sont en mesure d'extraire certaines méthodes de sélection d'attributs. Certaines permettant plus spécifiquement d'extraire des attributs fortement pertinents et d'autres se limitant aux attributs faiblement pertinents.

Blum et Langley (1997) proposent une synthèse des différentes méthodes de sélection des attributs en les présentant toutes comme des méthodes heuristiques d'exploration de l'espace des sous-ensembles d'attributs, variant selon 4 critères, soit :

- L'ensemble à considérer initialement (on peut par exemple considérer l'ensemble vide et ajouter petit à petit de nouveaux attributs ou au contraire considérer l'ensemble de tous les attributs et en enlever au fur et à mesure).
- L'organisation de la recherche (méthode simple qui, par exemple, à chaque étape, sélectionne le meilleur attribut à ajouter ou à enlever suivant un critère particulier, ou des méthodes plus sophistiquées basées sur une exploration en largeur d'abord ou encore une exploration ordonnée selon le nombre d'attributs dans le sous-ensemble),
- La stratégie utilisée pour évaluer le sous-ensemble (soit une évaluation globale basée sur le pouvoir discriminant du sous-ensemble, soit un test de la précision d'un algorithme d'apprentissage réalisé sur le sous-ensemble d'attributs à évaluer)
- Le critère d'arrêt de la recherche (par exemple, la recherche s'arrête lorsqu'il n'y a plus de gain de précision, ou lorsque parcours est terminé, ou encore lorsque l'ensemble des attributs classe parfaitement les exemples).

Ces méthodes peuvent se ranger dans trois grands types d'approches : 1) l'approche « intégrée» (*embedded approach*) lorsque la sélection est intégrée à la méthode d'induction (par exemple, dans le cas des arbres de décision, on choisit à chaque étape, l'attribut le plus discriminant vis-à-vis de la classe), 2) l'approche « filtre » (*filter approach*) qui cette fois-ci consiste à considérer le processus de sélection d'attribut et le processus d'apprentissage séparément, et 3) l'approche « encapsulée » (*wrapper approach*) qui consiste à choisir une méthode d'induction particulière, et à guider la recherche d'un ensemble d'attributs par des essais successifs des différents sous-ensembles d'attributs avec la méthode d'induction choisie. On peut citer parmi les algorithmes de sélection d'attributs les plus connus : RELIEF (Kira & Rendell, 1992), FOCUS (Almuallim & Diettrerich, 1991) (relevant de l' approche « filtre ») et le système OBLIVION (Langley & Sage, 1994) (relevant de l' approche « encapsulée »).

RELIEF associe à chaque attribut une valeur mesurant sa « pertinence » et choisit ensuite les k meilleurs. Cette valeur est attribuée en prélevant les exemples de l'ensemble d'apprentissage par un tirage aléatoire et en mettant à jour à chaque tirage la pertinence des attributs en se basant sur la différence d'un nouvel exemple tiré avec d'une part, l'exemple (déjà tiré) le plus proche et ayant une classe différente et d'autre part, le plus proche des exemples de la même classe. Ainsi, si la différence de l'exemple avec l'exemple de même classe est plus faible que la différence avec l'exemple de classe différente, la pertinence de l'attribut est augmentée. Finalement la pertinence calculée rend compte de la qualité de l'attribut à prédire la classe.

FOCUS recherche le plus petit ensemble d'attributs en mesure de déterminer la classe de tous les exemples de l'ensemble d'apprentissage en considérant d'abord des attributs simples puis des combinaisons de plus en plus importantes.

OBLIVION considère d'abord l'ensemble de tous les attributs et élimine à chaque étape celui qui entraîne (par sa suppression) les meilleures performances sur l'algorithme des plus proches voisins qui attribue à chaque nouvel exemple la classe des plus proches voisins.

Ces méthodes ont permis d'améliorer très significativement des résultats obtenus sans sélection. Cependant, elles comportent un certain nombre de défauts. Par exemple, RELIEF ne constitue pas une aide pour les attributs redondants. Si deux attributs sont très corrélés mais tous les deux liés à la classe étudiée, ils sont tous les deux sélectionnés. De même, en cherchant l'ensemble minimal, FOCUS peut par exemple n'extraire pour un diagnostic médical que le numéro de sécurité sociale du patient. En effet, compte tenu du fait qu'il n'y a qu'un numéro par patient, le numéro suffit à déterminer le diagnostic. Globalement, on peut retenir ici que ces méthodes consistent en un parcours partiel de l'ensemble des sous-ensembles d'attributs et que, de ce fait, il est toujours possible de les mettre en défaut. En particulier, on peut considérer que les méthodes qui évaluent la pertinence des attributs indépendamment les uns des autres ne tiennent pas compte des interactions entre ces attributs. Il est possible, par exemple, que deux attributs pris séparément n'apportent aucune information sur la classe d'un exemple, mais que combinés ils discriminent parfaitement les exemples sur les différentes classes. On peut donc penser que la seule méthode valide consisterait à explorer tous les sous-ensembles d'attributs soit 2^n sous ensembles (où n est le nombre d'attributs).

Une alternative à la sélection d'attributs est celle qui consiste à attribuer des poids et donc à considérer ainsi des degrés de pertinence intermédiaires. C'est en fait celle qui est mise en œuvre dans les réseaux de neurones (figure 2.5). La mise à jour des poids est effectuée sur la base des erreurs commises, l'exemple le plus simple étant certainement la règle de mise à jour du perceptron. Ainsi à chaque prise en compte d'un nouvel exemple, le poids qui représente la participation de l'attribut à la décision est recalculé. La mise à jour consiste en cas d'erreur, à retirer ou à ajouter une valeur suivant respectivement la participation ou l'absence de participation de l'attribut à la décision. On notera que certaines études tendent à montrer que la mise à jour par ajout ou retrait d'une valeur, donc une mise à jour additive est moins à même d'écarter de la décision des attributs non pertinents qu'une mise à jour multiplicative. Littlestone (1988) propose par exemple l'algorithme Winnow :

```
Initialiser tous les poids $w_1, \ldots, w_n$ des différents attributs à 1.
POUR tous les exemples faire
DEBUT
Soit un exemple $(x_1, \ldots, x_n)$. La sortie est 0 si $w_1x_1 + \ldots + w_nx_n < n$ et 1 sinon.
SI la sortie est incorrecte, c'est-à-dire si la valeur de référence donnée pour l'apprentissage est différente FAIRE:
    (a) SI la valeur prédite est 0 et que la valeur de référence donnée pour l'apprentissage est 1 ALORS pour chaque $x_i$ égal à 1, doubler la valeur de $w_i$.
    (b) SI la valeur prédite est 1 et que la valeur de référence donnée pour l'apprentissage est 0 ALORS pour chaque $x_i$ égal à 1, diviser par deux la valeur de $w_i$.
FIN POUR
```

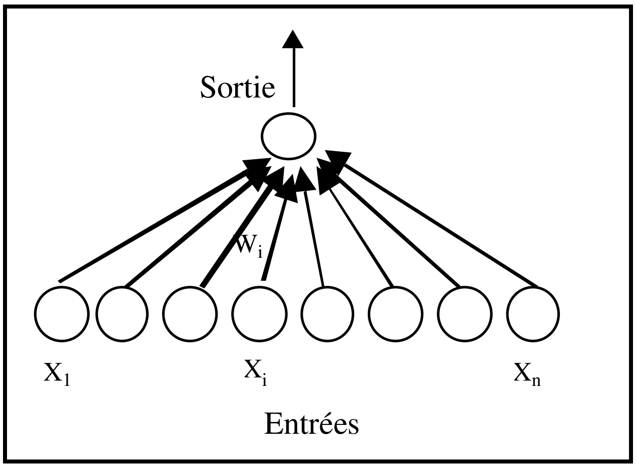


FIG. 2.5 Architecture d'un neurone qui comporte n entrées, dont l'importance respective est représentée par les différents poids w_i.

La sélection des exemples pertinents a aussi donné lieu à des recherches. De la même façon que pour la sélection des attributs pertinents, il existe les méthodes intégrée, filtre, et encapsulée. Parmi les méthodes « intégrées » on peut par exemple, citer les méthodes d'apprentissage telles que celle du perceptron ou de Winnow qui ne prennent finalement en compte que les exemples qui vont à l'encontre des prédictions réalisées à partir des exemples précédents. Concernant les méthodes « encapsulées » on pourra par exemple se reporter à Quinlan (1983). Dans l'ensemble de ces méthodes, les exemples pertinents sont, soit ceux qui vont à l'encontre des règles extraites à partir des exemples précédents, soit si on procède de façon globale, les exemples qui donnent les meilleures performances en apprentissage.

2.3.2.1.2 L'extraction de règles

En combinant des attributs a priori pertinents, il est possible de construire des règles de décision. On peut, soit les recueillir auprès d'un expert, soit les construire automatiquement selon le principe d'induction. Dans les deux cas, le problème est d'extraire des connaissances utiles à la résolution d'une tâche donnée, il semble donc bien s'agir d'un problème d'extraction d'information pertinente. L'information est ici pertinente au sens des prédictions qu'elle permet et au sens de la synthèse qu'elle constitue.

On notera que, dans le cadre de l'apprentissage automatique, l'extraction de règles peut être présentée comme un problème de sélection consistant en une recherche dans l'espace des règles possibles (Holsheimer & Siebes, 1994), basée sur une fonction d'évaluation des règles appelée fonction de qualité qui de ce point de vue peut être considérée comme une évaluation de pertinence que nous nous proposons donc d'exposer.

Considérons un ensemble d'exemples S sur lequel se base l'apprentissage, une description D (par exemple une conjonction de couples attribut-valeur du type « FORME=carré & COULEUR=vert »), l'ensemble des exemples couverts par la description D noté $\sigma_D(S)$ et une classe C (par exemple la classe des figures vertes). La fonction de qualité d'une règle prend en compte deux facteurs différents : sa validité et sa correction. La validité correspond au degré avec lequel la règle s'applique correctement (sa prédiction est vérifiée) à des exemples non vus, c'est-à-dire n'appartenant pas à S. Pour ce critère, une solution consiste à se baser sur la simplicité de la règle (nombre de conditions intervenant dans la prémisse) en considérant comme Hosheimer et Siebes (1994) l'indiquent que, plus la règle est simple, plus il est probable qu'elle décrive une relation existant réellement dans la base de données. On peut en effet penser qu'une règle trop complexe ne serait validée que par un petit nombre d'observations et pourrait n'être valable que pour l'échantillon utilisé pour l'apprentissage. Le second critère a trait à l'ensemble d'apprentissage (S). Elle est fonction de la précision de la règle (la proportion d'exemples bien classés) et de sa couverture (c'est-à-dire le nombre d'exemples appartenant à la classe qu'elle permet de classer). Si la précision vaut 1, D correspond à une condition nécessaire et on dit que la règle est complète. Si la couverture vaut 1, D correspond à une condition suffisante et on dit que la règle est déterministe. Si les deux valent 1, la règle est correcte et le critère de correction f_c vaut 1.

$$\text{précision} = \frac{|\sigma_D(S) \cap C|}{|\sigma_D(S)|}$$

$$\text{couverture} = \frac{|\sigma_D(S) \cap C|}{|C|}$$

Pour rendre compte de la correction d'une règle, Piatesky-Shapiro (1991) propose par exemple, que la mesure f_c ait les propriétés suivantes :

- $f_c = 0$ si D et C sont statistiquement indépendants. Soit : $\frac{|\sigma_D(S) \cap C|}{|\sigma_D(S)|} = \frac{|C|}{|S|}$

- f_c est monotone croissante avec $|\sigma_D(S) \cap C|$

- f_c est monotone décroissante avec $|\sigma_D(S)|$ (ou $|C|$)

Une fonction simple qui satisfait ces propriétés est : $|\sigma_D(S) \cap C| - \frac{|\sigma_D(S)||C|}{|S|}$

Elle consiste donc en une différence entre le nombre d'exemples bien classés par la règle et le nombre d'exemples classés correctement si D et C étaient indépendants.

L'extraction de règles est souvent réalisée dans l'objectif de créer des règles de décision sur un attribut particulier appelé label ou classe. En médecine par exemple, on s'intéressera ainsi plus spécifiquement aux règles dont la conclusion porte sur le caractère bénin ou malin de l'observation. Cependant, on peut aussi concevoir de ne pas s'intéresser à un attribut particulier et plutôt considérer l'ensemble des attributs comme des classes potentielles. Dans ce cas, même si le nombre de règles à générer est potentiellement important et la complexité de l'algorithme pour la génération, irrémédiablement plus grand, le même type d'évaluation des règles peut être utilisé et il est tout aussi possible de parler de pertinence (tout attribut pouvant être considéré comme un objectif). L'intérêt de ne pas simplement considérer les règles pour la décision mais plutôt d'extraire la connaissance des dépendances entre les différentes variables peut être lié à la recherche d'une compréhension plus profonde des données. Il peut aussi être lié à l'absence d'objectif particulier a priori. Dans ce dernier cas, on s'intéresse aux régularités présentes dans les données, se situant plutôt dans une problématique de découverte.

Un nouveau champ de recherche s'est constitué autour du thème de la Fouille de Données dont la problématique est précisément celle de la découverte. « Vous avez de l'or dans vos données et vous ne le savez pas », dans cette formule qui promeut la Fouille de Données, l'or en question correspond à des régularités exprimables par exemple, sous forme de règles « SI *condition* ALORS *conclusion* ».

La spécificité de la Fouille de Données par rapport au cadre classique de l'apprentissage en Intelligence Artificielle, est de porter sur le traitement de très grandes bases de données. Plus globalement, on considère l'absence de pré-traitements manuels qui permettraient par exemple de sélectionner les attributs pertinents et de supprimer d'éventuelles données bruitées. De ce fait, la tâche de sélection occupe un rôle plus important et la notion de pertinence est d'autant plus présente.

2.3.2.1.3 La sélection des règles à appliquer

Les systèmes experts ont pour objectif de résoudre des problèmes dans un domaine spécifique habituellement résolu par un expert humain. Le système MYCIN (Shortliffe, 1976) a par exemple été utilisé pour le diagnostic de maladies du sang. Ces systèmes sont composés de deux parties : la base de connaissances qui comprend un certain nombre de faits et de règles (SI *condition* alors *conclusion*) acquises auprès d'un expert, et le moteur d'inférences chargé d'exploiter ces connaissances. Le nombre de règles est souvent important, aussi est-il indispensable, compte tenu des capacités de traitement limitées, d'utiliser uniquement dans le raisonnement, les règles appropriées à la tâche. Ainsi, le moteur d'inférence comporte trois étapes :

1. Sélection des règles candidates
2. Choix d'une règle
3. Application de la règle choisie

Pour cette sélection, le moteur d'inférence peut utiliser la donnée du problème qui consiste en un ensemble de faits (par exemple, les symptômes d'un patient pour MYCIN) et éventuellement, une question qui précise la recherche de l'utilisateur (maladie x ?)

Ainsi, seules les règles dont les prémisses qui correspondent aux faits décrivant le problème sont utilisées. Ces règles sont appliquées et leurs conclusions deviennent alors de nouveaux faits qui permettent d'appliquer de nouvelles règles (on parle de chaînage avant) et ainsi de suite jusqu'à l'obtention d'un résultat. Le moteur d'inférence peut aussi fonctionner dans le sens inverse pour prendre en compte l'objectif de l'utilisateur. Ainsi, lorsque la conclusion d'une règle correspond à l'objectif, les faits correspondant à la prémisse de cette règle peuvent alors être appariés avec la conclusion d'autres règles et ainsi de suite jusqu'à ce que l'on retrouve les faits décrivant le problème (on parle de chaînage arrière). Ces mécanismes de chaînage constituent une première proposition pour régler le problème de la sélection des règles. Ils s'appuient sur le contexte courant (les faits représentant les données du problème) et l'objectif (le but entré par l'utilisateur).

Le résultat est un ensemble de règles candidates et la seconde étape consiste à choisir une règle parmi celles-ci. Si le but n'est pas d'être exhaustif dans la recherche, mais, de trouver une solution le plus vite possible, le choix de la règle est crucial. Ici le moteur d'inférence peut adopter différentes stratégies. Il peut, par exemple, choisir la règle la plus spécifique (celle qui comporte le plus de conditions) ou encore la plus récemment applicable. Une autre solution réside dans l'introduction de connaissances sur les connaissances, c'est-à-dire de métaconnaissances (Pitrat, 1990). Ainsi une méta-règle spécifiera, par exemple, de préférer plutôt un certain type de règles par rapport à tel autre type dans une situation donnée.

2.3.2.1.4 Un parcours sélectif de l'espace des états d'un problème

Certains problèmes peuvent être représentés par un arbre (encore appelé espace des états) dans lequel les différents nœuds représentent des états du problème et les arcs représentent les actions qui permettent de passer d'un état à l'autre. On trouve de nombreux problèmes de ce type dans le domaine des jeux, comme par exemple, les jeux d'échiquier. La situation initiale (définie par exemple par la disposition initiale des pièces sur l'échiquier) correspond à la racine de l'arbre. Les différentes actions (les différents coups possibles à partir de cette disposition initiale) mènent à d'autres situations (la nouvelle disposition des pièces sur l'échiquier suite au coup joué) à partir desquelles d'autres actions sont possibles. La résolution du problème consiste à choisir la séquence d'action qui mène à la solution ou à la meilleure solution lorsque la solution n'est pas unique. Pour des problèmes dans lesquels l'espace des états est réduit, une exploration totale est possible et la solution est donnée par des algorithmes du type min-max (figure 2.6) qui permettent d'évaluer les différentes situations suivantes possibles et de choisir celles dont la valeur est la plus forte. L'algorithme alpha-béta permet de n'explorer qu'une partie de l'espace, utilisant les précédentes explorations pour éliminer certaines branches de l'arbre duquel on ne peut espérer aucune amélioration. Cependant, lorsque l'espace des états est important, même ce type d'exploration n'est plus envisageable. Il est possible de limiter l'exploration en réduisant la visibilité à un certain nombre d'actions (on limite donc la profondeur de l'arbre dans ce cas). Cette solution n'est pas totalement satisfaisante car elle introduit une limitation identique pour toutes les branches de l'arbre alors que certaines peuvent paraître pourtant plus prometteuses que d'autres. On en vient donc à la solution qui distinguera la « force brutale » de la machine (qui repose sur la rapidité de l'évaluation de toutes les possibilités) de l'expertise humaine, qui consiste à introduire des connaissances heuristiques (auxquelles il n'est pas possible d'attacher des certitudes absolues) guidant l'exploration de l'espace de recherche (Pitrat, 1977). Ceci nous ramène donc à un système à base de règles et aux trois problèmes évoqués précédemment. Nous considérons donc que ces méthodes d'exploration de l'espace de recherche n'apportent qu'une solution partielle.

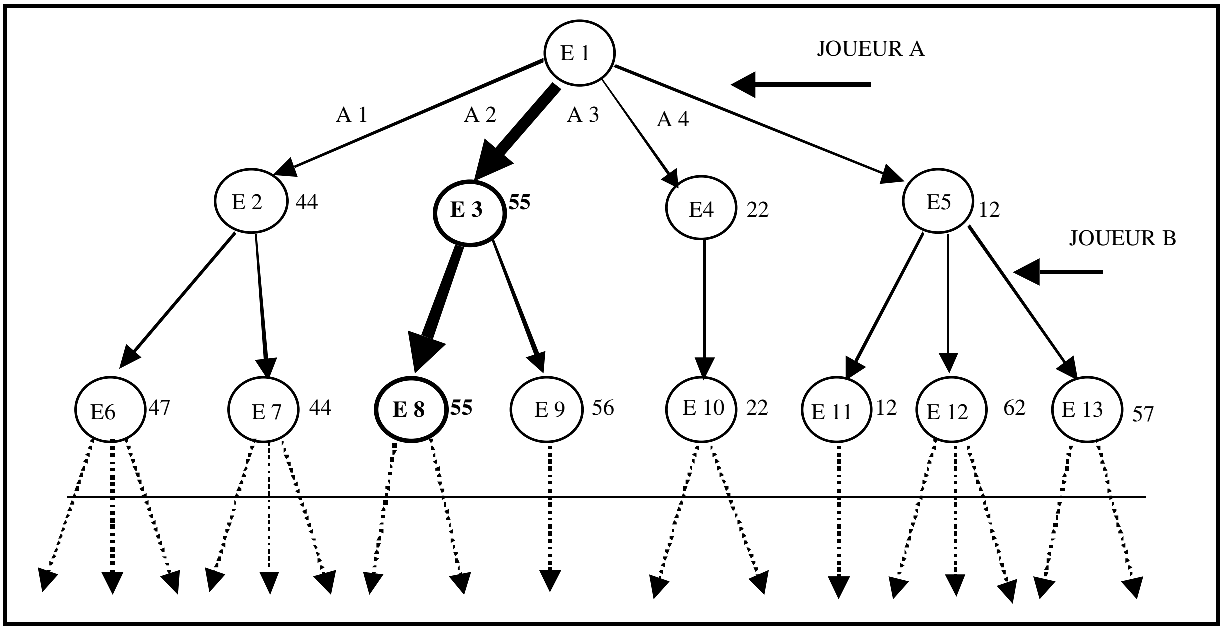


FIG. 2.6 Le joueur A est initialement dans le situation E1, quatre actions sont possibles. C'est l'action A3 menant à l'état E3 qui maximise la fonction d'évaluation des états. L'évaluation de E3 est elle-même faite en considérant cette fois-ci le point de vue du joueur B, elle correspond donc à considérer le minimum. L'évaluation des différentes solutions consiste donc à remonter l'arbre, des feuilles vers la racine, en alternant les fonctions min et max dans l'évaluation.

2.3.2.1.5 Conclusions

En Intelligence Artificielle, on peut constater que la pertinence concerne aussi bien des problèmes de sélection que des problèmes de synthèse, et les problèmes de sélection abordés sont eux-mêmes assez différents. Aucun consensus sur cette notion n'a été dégagé, si ce n'est peut-être l'idée, pour reprendre les termes de Subramanian et al. (1997), qu'un objet (une action, un exemple, un attribut, une inférence) n'est pas pertinent pour une tâche donnée et dans un contexte donné, si la réponse à la tâche ne change pas lorsque l'on change l'objet.

2.3.2.2 En Recherche d'Information

La pertinence est une notion tout aussi centrale pour la Recherche d'Information et contrairement à l'Intelligence Artificielle où les problèmes de pertinence peuvent paraître un peu disparates, le problème est ici bien défini. Il s'agit de concevoir des systèmes qui permettent de stocker un ensemble de documents (textes, images, documents multimédia) et de retrouver parmi cet ensemble, ceux qui respectent des critères formulés dans une requête. Nous présentons ici le problème de pertinence qui se pose pour la Recherche d'Information et la façon dont il est abordé d'un point de vue pratique et théorique.

2.3.2.2.1 Problématique

Un système de recherche d'information est un système informatique qui permet à un utilisateur de formuler une requête exprimant une recherche particulière et qui ensuite, à partir de cette requête, extrait automatiquement d'une base de documents, un ensemble de documents censés être liés à la recherche de l'utilisateur (ces documents pouvant apparaître dans un ordre reflétant la plus ou moins grande adéquation du document avec la requête). Il s'agit donc d'un système de sélection de l'information qui évalue une correspondance entre les différents documents de la base et la requête de l'utilisateur.

Schématiquement, un système de recherche d'information peut être décrit à partir de trois éléments :
- Un langage de requête, qui permet à l'utilisateur du système d'exprimer (dans une certaine mesure) son besoin d'information. Pour des documents textuels, il s'agira, par exemple, d'un ensemble de mots combinés éventuellement par des connecteurs logiques.

- Un langage de description des documents. Pour des documents textuels, il s'agira, par exemple, d'un ensemble de mots-clefs.

- Et enfin, une fonction de correspondance qui évalue la correspondance entre la requête (représentation de la recherche de l'utilisateur) et la description d'un document. Cette fonction peut par exemple, simplement évaluer la présence des mots de la requête parmi les mots-clefs représentant le texte.

C'est la fonction de correspondance qui évalue en quelque sorte la pertinence d'un document vis-à-vis d'une requête, bien qu'elle soit déjà déterminée par le choix des langages de requête et de description des documents puisqu'elle est construite sur la base des critères retenus pour la représentation des requêtes ou des documents. Le plus traditionnellement, elle correspond à l'évaluation du degré de satisfaction des critères de recherche exprimée dans la requête par le document en opérant un simple appariement de la représentation du document avec celle de la requête (identification par exemple des mots de la requête avec les mots-clefs représentant le document). Cependant, d'autres types de relations peuvent jouer un rôle dans l'évaluation de la pertinence. Les critères d'évaluation et le choix de la relation considérée pour comparer la requête et le document sont donc les deux éléments qui interviennent dans l'évaluation de la pertinence.

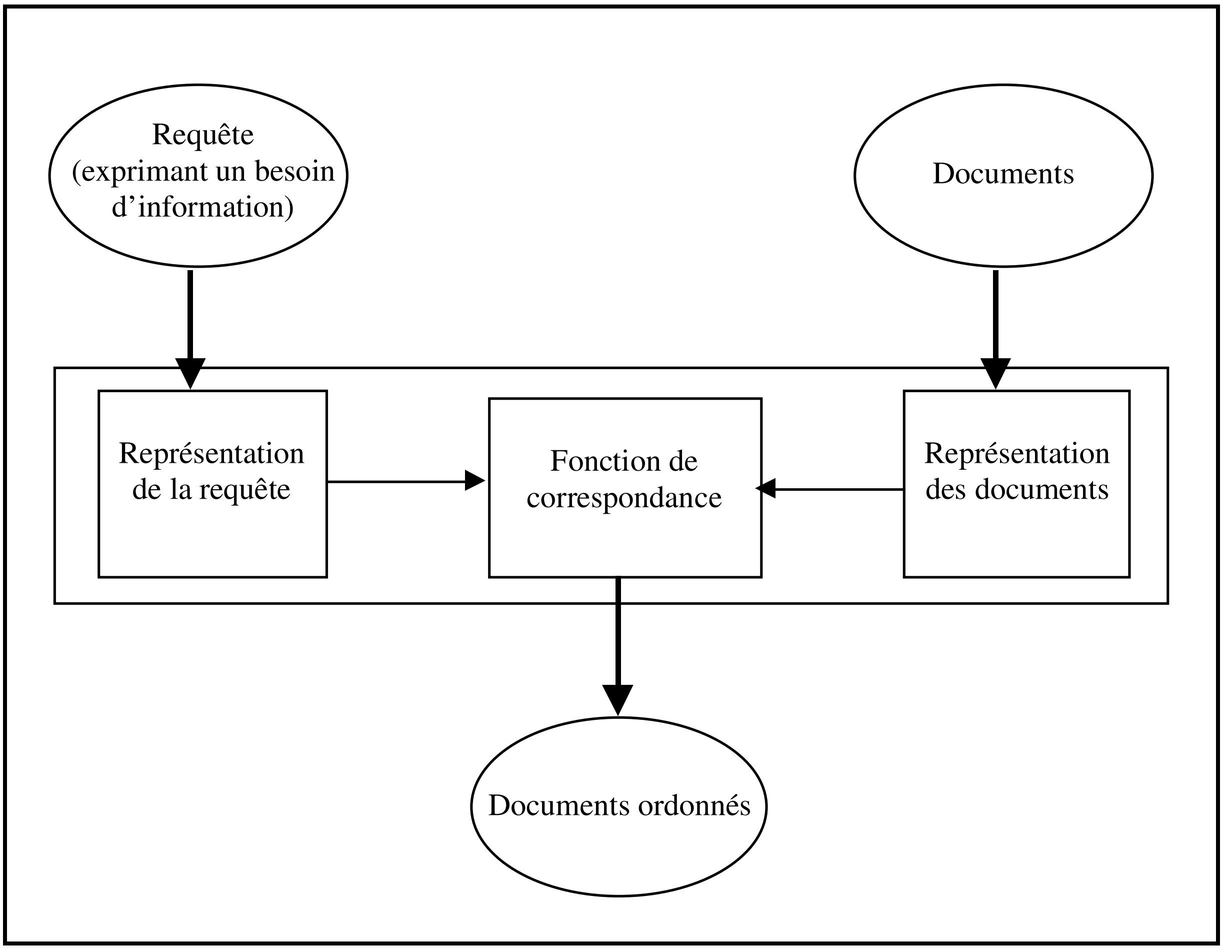


FIG. 2.7 Représentation de l'architecture d'un système de recherche d'information (encadré) et de ses entrées et sorties.

2.3.2.2.2 Des éléments pour l'évaluation de la pertinence

Les premières recherches ont plus spécifiquement porté sur le critère thématique considérant une recherche de documents portant sur un sujet donné (par exemple, la recherche de documents portant sur les oiseaux migrateurs). De plus, elles vérifiaient simplement la présence des termes de la requête parmi les termes représentant le document considérant donc une simple relation d'appariement entre le thème de la requête et les thèmes abordés par le document (par exemple la migration des cigognes). Par la suite, le rôle important joué par de nombreux autres types de relations et d'autres critères dans l'évaluation de la pertinence ont été mis en évidence.

Ainsi, Green et Bean (1995) identifient pour le critère thématique, différents types de relations à partir de l'analyse de la relation liant les thèmes d'un guide thématique d'une œuvre (en l'occurrence une œuvre religieuse « Church of Jesus Christ of Latter-day saints ») aux thèmes traités par des passages de textes référés dans le guide (parmi les différents thèmes, on trouve entre autres : « le paradis », « la reconnaissance de Dieu », « la persévérance »). Green et Bean (1995) observent ainsi que seulement le tiers des relations sont de type « appariement » . Il s'agit par exemple, de références au thème (passage décrivant le fait que Dieu créé un jardin au paradis dans lequel il place l'homme), ou de définitions ou encore de caractérisations du thème (passage décrivant la profondeur de la sagesse et de la connaissance de Dieu). Ils constatent qu'un dixième des relations sont de type hiérarchique. Il s'agit par exemple, de sous-classe ou de membre de la classe désignée par le thème (passage décrivant une forme de persévérance). Enfin, ils remarquent qu'une bonne moitié des relations sont de type structurel. Il s'agit d'un des composants décrivant le thème (ce dernier étant considéré comme une structure). Par exemple, si le thème est « la reconnaissance de Dieu », le passage décrivant envers qui Dieu est reconnaissant renseigne sur le composant « receveur » du thème.

Indépendamment de la remise en question de la relation d'appariement pour le critère thématique, on trouve aussi une remise en question du critère thématique lui-même. En effet, bien qu'il soit reconnu comme un élément indispensable à l'évaluation de la pertinence, de nombreux autres critères jouant un rôle non négligeable dans l'évaluation de la pertinence ont été mis en évidence. Pour citer une étude récente, Barry (1994) distingue, à partir de l'interrogation d'un ensemble d'utilisateurs sur les raisons de leur jugement de pertinence, 23 catégories qu'elle regroupe en 7 grandes classes :

- Contenu du document (profondeur du sujet, précision, validité objective, efficacité, clarté, récence)
- Expériences précédentes de l'utilisateur (passé, expérience, capacité à comprendre, nouveauté du contenu, nouveauté de la source, nouveauté du document)
- Croyances et préférences de l'utilisateur
(précision, validité subjective, affectivité)
- Autres informations et autres sources
(consensus, vérification externe, disponibilité)
- Sources de documents

(qualité de la source, réputation, visibilité)

- Document comme entité physique

(facilité pour se le procurer, coût)

- Contexte de l'utilisateur

(contraintes de temps, relation avec l'auteur)

Lorsque ces critères restent des attributs des documents (par exemple, la taille du document, la notoriété de l'auteur), il est envisageable de les inclure dans les langages de requête et d'indexation et d'étendre la fonction de correspondance de façon à les prendre en compte. Cependant, les recherches récentes mettent en évidence des critères liés à l'utilisateur (par exemple, ses connaissances, ses préférences etc.) qui ne peuvent pas aussi aisément être pris en compte du fait justement des différences inter-individuelles, de la multitude de ces critères, de la difficulté à y accéder et de leur aspect dynamique. En effet, les connaissances, les croyances des utilisateurs peuvent changer constamment (Bruce, 1994). On peut penser en particulier à l'influence de la réponse du système.

2.3.2.2.3 Evolution de la notion de pertinence en Recherche d'Information

Force est de constater que le débat théorique opposant la « pertinence système » et la « pertinence utilisateur » est le débat récurrent en Recherche d'Information (autour de la notion de pertinence). La « pertinence système » renvoie à une mécanique formelle évaluant la pertinence du contenu d'un document vis-à-vis d'une requête, elle est associée aux critères thématiques. A l'opposé, « la pertinence utilisateur » considère les critères thématiques comme des critères parmi l'ensemble des autres critères dont les critères utilisateurs, et plutôt que de considérer une requête comme origine de la recherche, il renvoie à la notion beaucoup moins formelle de « besoin d'information ». La philosophie de la conception du système n'est pas la même suivant qu'on adopte le point de vue « système » ou le point de vue « utilisateur ». Dans le premier cas, le système est conçu comme une mécanique reposant sur des principes logiques ou statistiques formels. Dans le second cas, le système peut-être considéré comme une tentative de modélisation des processus psychologiques de l'utilisateur impliqués dans la recherche d'information pertinente. Comme le fait remarquer Saracevic (1975), il est difficile de trancher en faveur de l'un ou l'autre de ces points de vue qui se complètent. Green (1995) considère pour sa part que l'utilisateur est le véritable juge de ce qui est pertinent mais que d'un autre côté, il n'est pas le meilleur, remarquant qu'il ne dispose pas nécessairement des connaissances qui lui permettront d'évaluer la pertinence du document et qu'il est même quelquefois dans l'incapacité d'exprimer son besoin d'information par une description de contenu des documents potentiellement pertinents.

Ce débat a pour origine les premières recherches sur l'évaluation des systèmes (Saracevic, 1975, p. 328) qui ont posé les questions suivantes : Quelle méthodologie d'évaluation choisir ? Qui juge de la pertinence ? Le système ou l'utilisateur ? Initialement on considérait qu'un document pertinent non retrouvé indiquait une défaillance du système mettant en cause la représentation des documents ou encore la façon dont ces représentations étaient utilisées (Saracevic, 1975, p. 327). Mais dès lors que l'on fait intervenir l'utilisateur comme juge et comme référence pour l'évaluation du système, comment ne pas noter, comme le souligne Barry (1994, p. 152), le décalage important qui existe alors entre ce que font des systèmes ne prenant pas en compte les critères de l'utilisateur évoqués ci-dessus et leur mode d'évaluation. Face à cette problématique, deux réactions ont été observées. La première consiste à juger que la pertinence est une notion humaine trop ambiguë et multiforme (Cooper, 1971) et qu'elle ne peut-être utilisée pour l'évaluation. Une autre consiste à tenter d'intégrer au système les facteurs psychologiques qui interviennent dans l'évaluation de la pertinence.

Devant la difficulté à rendre compte de la notion de pertinence, certains ont proposé de scinder la notion. Ainsi, Cooper (1971) fait une distinction entre la notion de pertinence logique (qui rend compte de la relation thématique par la logique du premier ordre) et celle d'utilité qui fait intervenir des considérations plus empiriques. Dans le même ordre d'idée, Wilson (1973) distingue pertinence psychologique et pertinence logique, et Foskett (1972) distingue pertinence privée (qui renvoie donc aux critères utilisateurs et qu'il appelle « pertinence » en anglais) et pertinence publique (qu'il appelle « relevance » en anglais).

A l'opposé, les définitions de la pertinence proposées peuvent être vues comme des tentatives pour une vision unifiée de la notion. Et bien qu'on leur reproche d'être des paraphrases qui renvoient à des notions qui ne sont pas mieux expliquées (Saracevic, 1975 p. 328 ; Cooper, 1971, p. 20), elles constituent néanmoins une information intéressante sur la façon dont on peut approcher cette notion. Cuadra et Katter (1967) proposent par exemple la définition suivante :

« Relevance is the correspondence in context between an information requirement statement and an article ; i.e., the extent to which the article covers material that is appropriate to the requirement statement ».

2.3.2.2.4 Solutions, tendances actuelles

Comme en témoigne le récent numéro spécial de « Journal of the American Society for Information Science » consacré à la pertinence en Recherche d'Information (Froehlich, 1994), les préoccupations actuelles se portent plutôt vers la pertinence orientée utilisateur et par conséquent sur l'étude psychologique de l'utilisateur. Ainsi, n'ayant toujours pas pu aboutir à une définition consensuelle de la pertinence, les recherches tentent de faire l'inventaire des critères qui peuvent jouer un rôle dans l'évaluation de la pertinence par l'utilisateur (Barry, 1994 ; Howard, 1994). D'autre part, comme nous l'évoquions précédemment, partant des considérations que certains critères utilisateur ne peuvent pas être intégrés facilement à la fonction de correspondance, et que les connaissances et donc le besoin d'information de l'utilisateur peut changer sous l'influence des réponses du système, une solution consiste à baser la recherche d'information sur une interaction constante visant à utiliser autant que possible l'aide que peut apporter l'utilisateur. Des recherches se sont donc portées sur l'étude de cette interaction (Bruce, 1994), rejoignant ici l'opinion défendue par (Saracevic, 1975, p. 321) « In the most fundamental sense, relevance has to do with effectiveness of communication ». D'un point de vue pratique, le système proposé par Denos (1997) constitue un bon exemple de ce type de recherche. Ce système tente de réduire la distance qui sépare l'utilisateur du système, en prenant appui sur l'interaction. La recherche est conçue comme une suite d'itérations du type : formulation de requête par l'utilisateur, réponse du système, puis reformulation de l'utilisateur sur la base de cette réponse etc. Le langage utilisé pour la formulation des requêtes intègre autant que possible les critères susceptibles d'entrer en ligne de compte pour l'utilisateur et les documents retrouvés sont regroupés par classes de pertinence. L'idée est de permettre une meilleure lisibilité des résultats ainsi qu'une meilleure compréhension du fonctionnement du système par l'utilisateur conditionnant une meilleure reformulation de la requête à la prochaine itération et par conséquent une interaction plus efficace.

2.3.2.3 Une logique de la pertinence

Un certain nombre de travaux ont tenté de s'abstraire des différents problèmes particuliers que nous avons évoqués pour proposer une logique rendant compte de la pertinence. Il en est question ici.

2.3.2.3.1 La pertinence comme une relation d'implication

La pertinence d'une information se définit toujours relativement à une autre information ou ensemble d'informations. La pertinence se définit donc comme une relation et, par conséquent, la compréhension de ce que recouvre cette notion renvoie au problème de l'identification de la nature cette relation.

Bien que cette question reste encore largement ouverte, comme en témoignent les recherches menées par Green (Green, 1995 ; Green & Bean, 1995), on peut tout de même noter que les propositions convergent vers la notion de relation d'implication. Ainsi, selon ce point de vue, une information A est pertinente relativement à une information B si A intervient dans une inférence qui permet d'arriver à l'information B. Cooper (1971), par exemple insiste explicitement, sur le lien étroit entre pertinence et conséquence logique : « One might, on first reflection , suppose that logical consequence could be only distantly related to relevance, but it will be argued shortly that when problem is posed in terms of declarative sentences, logical consequence and relevance are very intimately connected indeed » (Cooper, 1971, p. 22). Plus précisément, se plaçant dans la problématique de recherche d'information, il considère une information comme logiquement pertinente (le terme « logiquement pertinent » est utilisé pour bien séparer cette notion d'autres critères utilisés par des utilisateurs de systèmes de recherche d'information qui n'ont selon lui aucun rapport) vis-à-vis d'un besoin d'information (représentée ici par une question), si elle fait partie d'un ensemble minimal de prémisses qui permet de conclure sur la question posée.

Toujours dans le cadre de la recherche d'information, van Rijsbergen (1986) définit aussi la relation de pertinence comme une implication de la représentation logique du document (donc de l'objet pertinent) vers celle de la requête (l'objet vis-à-vis duquel la pertinence se définit).

Pour illustrer cette idée, considérons un ensemble de documents $\{D_1, D_2, D_3\}$, chaque document étant représenté dans une logique booléenne par une conjonction de termes (leurs mots-clefs par exemple) :

$$D_1 = \{A\}$$

$$D_2 = \{B, C\}$$

$$D_3 = \{A, B, C\}$$

Soit la requête $Q = A\&B$ qui indique une recherche portant sur A et B.

D_1 n'est pas assez spécifique, il n'est donc pas pertinent. D_2 ne respecte pas les critères exprimés dans la requête. D_3 satisfait les critères exprimés dans la requête. Il est pertinent et $D_3 \rightarrow Q$ (contrairement à D_1 et D_2).

On notera cependant que si $D_4 = \{A, B\}$, on a « $D_4 \rightarrow Q$ », et D_4 est encore plus en adéquation avec la requête. Ainsi, l'implication ne permet pas ici de prendre en compte cette différence. Pour le faire, il faut considérer « $Q \rightarrow D$ » et donc comme le suggère Nie (1988) considérer une fonction agrégeant $P(D \rightarrow Q)$ et $P(Q \rightarrow D)$ soit $F(P(D \rightarrow Q),P(Q \rightarrow D))$. On peut cependant considérer que ce dernier sens de l'implication permettrait simplement de départager différents documents et que l'implication $D \rightarrow Q$ joue un rôle plus important ($Q \rightarrow D_1$ et pourtant D_1 est manifestement moins pertinent que D_3). Nie (1988) considère que $Q \rightarrow D$ joue un rôle particulièrement important vis-à-vis des novices et que $D \rightarrow Q$ s'applique mieux aux experts.

van Rijsbergen (1986) considère des degrés de pertinence et propose une mesure de la pertinence d'un document D vis-à-vis d'une requête Q basée sur la mesure de l'incertitude de l'implication « $D \rightarrow Q$ », soit $P(D \rightarrow Q)$. Pour cette évaluation, il propose d'utiliser le principe d'incertitude suivant :

« Etant donné deux assertions x et y, la mesure d'incertitude de « $x \rightarrow y$ » relative à un certain ensemble de connaissances est déterminée par l'information minimale à ajouter à cet ensemble pour que « $x \rightarrow y$ » soit vrai. »

Cette définition constitue un principe général qui peut être formalisé dans la sémantique des mondes possibles (Nie & Brisebois, 1996). Elle se base sur des travaux portant sur la logique conditionnelle que nous évoquerons dans le paragraphe suivant.

En IA, les définitions d'attributs pertinents reposent aussi sur une relation d'implication. La définition d'un attribut pertinent est en effet basée sur la probabilité de la classe sachant la valeur de l'attribut. De plus, Galles et Pearl (1997) dans un cadre probabiliste et Darwiche (1997) dans un cadre logique propose de formaliser la « non pertinence » par une indépendance conditionnelle.

2.3.2.3.2 L'inadéquation de l'implication matérielle

En Recherche d'information, l'implication matérielle qui s'écrit $\neg D \vee Q$ (tableau 2.1) est écartée sur la base de l'argument suivant : Considérons un document D décrit par un ensemble de termes peu cohérents (voire même contradictoires) alors $D \rightarrow Q$ pour n'importe quelle requête (donc même si D n'est pas pertinent) puisque l'évaluation de $\neg D$ est grande. Reprenant cet argument, van Rijsbergen (1986), dans son évaluation de $D \rightarrow Q$ qu'il note $P(D \rightarrow Q)$, insiste sur le fait que cette évaluation de l'incertitude n'est ni une implication matérielle, ni la probabilité d'une implication matérielle P(¬D∨Q), mais une probabilité conditionnelle P(Q/D) (correspondant au rapport du nombre de termes communs à la requête et au document sur le nombre de termes représentant le document) que l'on peut approcher dans un cadre logique par le principe d'incertitude énoncé dans le paragraphe précédent.

En fait, ce qui est en cause une nouvelle fois ici, concerne le fait bien connu que la représentation de la relation d'entraînement « SI A ALORS B » par ¬A∨B n'est pas totalement satisfaisante. En effet, en particulier lorsque A est faux, ¬A∨B est vérifiée pour n'importe quel B. Or, intuitivement, la règle « SI A ALORS B » se comprend comme « SI A est vérifié ALORS B est vérifié » ne considérant pas le cas dans lequel A est faux. De ce point de vue, un nouveau connecteur qui ne statue pas sur la relation d'entraînement lorsque A n'est pas vérifié, notons le B/A, peut sembler plus adéquate.

A	B	¬A∨B	B/A
1	1	1	1
1	0	0	0
0	**1**	**1**	**?**
0	**0**	**1**	**?**

TAB. 2.1 Comparaison de la table de vérité de l'implication matérielle (en logique classique) et de la valeur de B/A.

Si l'on considère un monde fermé dans lequel il est possible de statuer sur la valeur de vérité (vrai ou faux) de toute proposition (le cadre par exemple de la preuve de théorème), l'implication matérielle peut être justifiée par la preuve suivante :

Partant du constat que « SI A ET B ALORS B » est une tautologie, donc que A∧B→B est vraie quelle que soit la valeur de vérité de A et celle de B et en établissant la table de vérité de A∧B→B, on obtient les deux dernières lignes de la table de l'implication matérielle (représentées en gras). Ainsi, lorsque A∧B est faux, quelle que soit la valeur de vérité de B, l'implication est vérifiée, d'où l'adéquation de l'implication matérielle.

A	B	$A \wedge B$	$\neg(A \wedge B) \vee B$
1	1	1	1
1	**0**	**0**	**1**
0	**1**	**0**	**1**
0	0	0	1

TAB. 2.2 La valeur de vérité d'une implication dont la prémisse est fausse est nécessairement vraie si l'on considère « $A \wedge B \rightarrow B$ » comme une tautologie.

Cette preuve ne tient plus si on se situe dans un monde ouvert dans lequel la valeur de vérité d'une proposition peut être indéterminée, la notion de tautologie n'ayant alors plus le même sens. Dans ce cas, on peut définir une proposition logique comme une tautologie par le fait qu'elle soit vraie lorsqu'elle peut être valuée, et indéterminée dans les autres cas. Ainsi « SI A ET B ALORS B » qui représente le fait que quand A ET B est vrai, B est vrai, est une tautologie car dans tous les cas où elle peut être évaluée, elle est évaluée comme vraie (seule la première ligne du tableau 2.2 nous intéresse).

2.3.2.3.3 Vers une logique de la pertinence

Un certain nombre de logiques ont proposé des solutions à cette inadéquation. Il s'agit des logiques conditionnelles (pour une revue de ces différentes logiques on pourra se reporter à Dubois & Prade, 1991) ainsi que des logiques de la pertinence (Anderson & Belnap, 1975). On notera que c'est vers ce type de logique que se dirige la Recherche d'Information (van Rijsbergen, 1989 ; Nie & Brisebois, 1996 ; Meghini & Straccia, 1996).

Ainsi le principe d'incertitude proposé par van Rijsbergen (1986) repose sur des travaux issus de la logique conditionnelle et l'implémentation du principe de Ramsay par la logique modale floue proposée par Nie et Brisebois (1996) s'identifie bien à la proposition faite par Stalnaker (1968), qui propose de vérifier la vérité de l'assertion « SI A ALORS B », quand A est faux, en se plaçant dans la situation la plus proche (le monde le plus proche) telle que A est vraie et en vérifiant alors la vérité de B dans ce cas. Les recherches concernant la logique conditionnelle ont aussi porté sur la définition d'un nouveau connecteur représentant « A sachant B », indépendamment de toute mesure. Certaines recherches ont aussi fait le lien avec la logique des défauts où « A sachant B » est interprété par « SI A est vrai ALORS à moins que B puisse être prouvé faux, B est vrai ».

La logique de la pertinence proposée par Anderson et Belnap (1975) propose la notion d'entraînement qui consiste en une dépendance plus forte entre la prémisse A et la conclusion B d'une règle «SI A ALORS B» que la simple implication matérielle «$\neg A \vee B$».

Ils proposent de considérer « SI A ALORS B » comme le fait que B soit prouvable à partir de A. Dans le formalisme de la déduction naturelle, ils définissent ainsi un système de preuve pour tour à tour faire disparaître des axiomes vérifiés par l'implication matérielle qu'ils trouvent inacceptables. Ainsi, l'axiome $A \rightarrow (B \rightarrow A)$ disparaît (il ne peut plus être démontré dans le système) à partir de la distinction entre vérité contingente et vérité nécessaire (si B est nécessairement vraie, elle ne peut entraîner un fait contingent A), et $B \rightarrow (A \rightarrow A)$ disparaît en prenant en compte l'utilité des propositions dans les différentes preuves (B n'est pas utilisée dans la preuve de $(A \rightarrow A)$). Dans le système de preuve résultant, le fait que l'on puisse opérationnellement déduire B à partir de A (condition nécessaire et suffisante) permet de conclure à la relation d'entraînement « SI A ALORS B ». Une condition nécessaire mais pas suffisante à la relation d'entraînement est que les propositions A et B partagent des variables (contenu sémantique non disjoint).

Ce type de logique a été utilisé en Recherche d'Information. Ainsi Meghini et Straccia (1996) reprennent la logique à quatre valeurs (vrai, faux, incohérent, indéterminé) introduit par Dunn (1976) similaire à la logique proposée par Anderson et Belnap (1975) mais se limitant au cadre propositionnel.

On trouve aussi des logiques ou des formalismes qui portent plus spécifiquement sur la prise en compte de la situation (Wilson, 1973 ; Devlin, 1991). La logique de Wilson (1973), permet d'exprimer des points de vue subjectifs (préférences, centres d'intérêt). Ainsi une information est pertinente si elle apporte une réponse ou fournit une aide à la réponse (c'est-à-dire peut être utilisée dans un raisonnement logique conduisant à la réponse) à un ensemble de questions comportant un intérêt pour le système (c'est cet ensemble d'intérêts qui constitue la situation).

Pour résumer, si l'on considère une logique de la pertinence qui rende compte des différents aspects évoqués, cette logique doit être construite autour d'une relation d'implication (plus contraignante que l'implication matérielle), et autoriser une subjectivité considérant donc non plus la notion de vérité mais plutôt la notion de croyance. Certaines croyances pouvant être révisées suite à la prise en compte de nouvelles informations, l'idée est donc plutôt de considérer un monde ouvert laissant place à l'indétermination et la préférence va donc aux logiques non monotones (Kayser, 1997,

p. 170). Il existe aussi un consensus général autour de l’idée que la pertinence est une affaire de degrés. On considèrera donc en particulier des logiques multi-valuées.

A la vue de cette liste d’aspects souhaités, on peut penser que le cadre probabiliste est suffisant. L'approche logique n'est cependant pas à écarter (Nie & Brisebois, 1996) du fait de la facilité avec laquelle elle permet d'introduire des connaissances du domaine, et la souplesse avec laquelle on peut manipuler des degrés de vérités (tout en gardant un cadre théorique) relativement à l’approche probabiliste qui fait intervenir des hypothèses d'indépendance et des règles de calcul strictes. On peut aussi ajouter que toutes les incertitudes ou imprécisions ne sont pas nécessairement liées à la notion de fréquence.

2.3.3 Pertinence et communication

C’est sur la notion de pertinence que Sperber et Wilson (1989) fondent leur théorie de la communication, prolongeant ainsi l'idée de Grice (1975). Ils la définissent selon deux considérations. Plus l'effet cognitif de l'information est grand (c'est-à-dire plus l'information entraîne de nouvelles croyances ou au contraire, plus elle permet de supprimer d'anciennes croyances), plus l'information est pertinente. Par ailleurs, plus l'effort nécessaire au traitement de cette information est important, moins l'information est pertinente. L'exemple qu'ils donnent pour illustrer leur définition (Sperber & al., 1995) s'appuie sur trois assertions différentes, adressées à une personne désireuse de se rendre le plus tôt possible à Manchester :

1) Le prochain train pour Manchester est à 5:30 de l'après-midi,
2) Le prochain train pour Manchester est parfois après 4:00 de l'après-midi,
3) Le départ du prochain train pour Manchester est prévu 7500 secondes après 3:25 de l'après-midi.

Ainsi, l'information 1 est plus pertinente que l'information 2 selon le premier critère et plus pertinente que l’information 3 selon le second critère. Sur la base de cette définition de la pertinence d’une information, ils définissent deux principes.

Le premier principe est cognitif. Il spécifie que la cognition humaine est guidée par la recherche de la pertinence maximale (plaçant là la notion de pertinence au centre de la cognition).

Ainsi, lorsque des effets cognitifs sont attendus d'une information particulière l'attention est dirigée vers cette information. Au contraire dans le cas où il n'y a pas d'attente particulière, ce sont les considérations d'effort qui jouent le premier rôle.

Le second principe de pertinence s'applique à la communication. Il spécifie que tout énoncé communique au destinataire la présomption de sa pertinence optimale. Ce principe selon le point de vue des deux auteurs suffit à expliquer la compréhension d’un énoncé, c'est-à-dire la façon dont l'implicite est interprété. Ainsi, par exemple, s’il nous est demandé de poster une lettre sans autre précision, on choisira une tarification "classique" plutôt qu'une tarification rapide ou lente car, si ce devait être le cas, on peut supposer que cette information qui est pertinente dans le sens où elle remet en cause nos croyances (si on considère que le tarif normal est la croyance par défaut) aurait été communiquée.

2.4 Synthèse

Nous tentons maintenant de dégager de ce parcours, une vue synthétique de la pertinence, répondant à nos différentes interrogations initiales, et dressant les grands principes d'un système pour l'extraction d'information pertinente. Afin d’avoir des éléments de réponse, nous commençons par lister un ensemble de points communs entre les différentes approches et nous mettons ensuite l’accent sur un courant de pensée qui apporte un certain nombre d’explications.

2.4.1 Quels points communs entre toutes ces approches ?

Malgré les apparentes différences entre toutes les approches décrites, il est néanmoins possible d’extraire des constantes.

2.4.1.1 Les critères d’évaluation

Les critères de pertinence désignés dans les différents domaines se recouvrent largement. Ces critères peuvent être regroupé en trois grandes catégories.

Il s'agit toujours soit :

- de **l'objectif du système**

L'attribut à prédire (IA), la requête ou le besoin d'information (RI), le but (première source d'activation) dans le modèle ACT, l'objet focalisé dans le cadre de l'étude de l'attention (recherche active).

- de **l'identité du système**

(auquel nous associons un caractère plus durable que l'objectif)

Les connaissances de l'utilisateur, ses préférences (RI), les règles présentes dans le système (IA), les croyances de l'interlocuteur (en linguistique), les inférences internes dans le modèle ACT (troisième source d'activation).

- du **contexte courant**

(Il s'agit des données externes présentent dans l'environnement)

Les données du problème, par exemple, la disposition des pièces sur l'échiquier (IA), les documents disponibles (RI), l'environnement perçu (seconde source d'activation) dans le modèle ACT ou encore l'amorce dans le cadre des expériences concernant l'accès en mémoire.

Nous rassemblons ces trois critères généraux sous le terme de **situation**.

2.4.1.2 L'évaluation d'une implication

Nous avons souligné dans la partie concernant la logique de la pertinence, qu'il existe un large consensus autour de l'idée d'implication. Ainsi, un objet O est pertinent vis-à-vis d'une situation S si « O→S » indiquant donc une relation de spécificité de O vis-à-vis de S. Certains considèrent aussi que l'implication inverse « S→O » joue un rôle, bien que secondaire.

Nous montrons ici comment il est possible de représenter la pertinence telle qu'elle est définie dans les différentes approches à partir de l'implication « O→S » et nous considérons le rôle éventuel de « S→O ».

Dans le cadre de la sélection d'attributs :

S est un attribut à prédire (OBJECTIF).

O est un attribut.

La pertinence de O repose sur l'évaluation de « O→S ». En effet, l'attribut est dit pertinent si la connaissance de sa valeur influe sur celle de l'attribut à prédire.

Dans le cadre de la sélection d'une règle à appliquer :

S est un but (OBJECTIF) ou les données du problème (CONTEXTE COURANT).

O est la règle dont on considère respectivement la conclusion ou la prémisse.

La sélection des règles dans les systèmes experts repose sur un appariement des objectifs ou des données du problème avec respectivement les prémisses ou les conclusions des règles. Aussi, cet appariement ne consiste pas en une simple relation d'égalité mais bien en une relation d'implication. Ainsi, on privilégiera plutôt les règles dont les prémisses s'appliquent le plus spécifiquement aux données du problème, ou encore les règles qui concluent spécifiquement sur l'objectif. On notera que la relation de spécificité « O→S » se décrit ici sur des parties de O.

Dans le cadre de l'extraction de règles :

S est l'ensemble des observations (CONTEXTE COURANT) sur lesquelles porte la règle et éventuellement des buts particuliers (OBJECTIF).

O est la règle extraite dont on considère les observations auxquelles elle s'applique.

La règle explique en quelque sorte les observations. La mesure de correction (reposant sur la précision et la couverture) renvoie à la proportion d'observations à laquelle la règle s'applique correctement. En effet, la couverture correspond à la proportion d'observations à laquelle la règle s'applique et la précision correspond à la proportion de ces dernières observations à laquelle la règle s'applique sans erreur. On se ramène donc à la mesure de l'inclusion de l'ensemble des observations dans celui des observations à laquelle la règle s'applique sans erreur ($S \subset O$) (qui n'est pas limité à l'ensemble des observations données pour l'apprentissage) ce qui a le sens d'une implication (« O→S »).

Dans ce même cadre, on peut aussi considérer une autre configuration pour laquelle la notion de pertinence a un sens. Considérons le cas où il s'agit de déterminer de nouvelles informations pertinentes en les inférant à partir des règles présentes.

S est l'ensemble des conditions de la prémisse de la règle qui peuvent renvoyer aux données du problème ou même à l'objectif (CONTEXTE COURANT, OBJECTIF).

O est la conclusion inférée.

On remarquera que la mesure de correction d'une règle qui repose sur la précision et la couverture renvoie à une double implication entre la conclusion et la prémisse. En effet, si l'on considère la règle « SI D ALORS C », la précision correspond à l'évaluation de « $D \rightarrow C$ » ($P(D \cap C)/P(D)$) et la couverture correspond à l'évaluation de « $C \rightarrow D$ » ($P(D \cap C)/P(C)$). La pertinence d'une information inférée renvoie donc à une double implication entre cette information et les conditions exprimées en prémisse représentant la situation.

Si lors de l'extraction, on considère les deux sens de l'implication, on a coutume, lors de l'application des règles de ne considérer qu'un seul sens. Pourtant, on peut imaginer que l'implication inverse permette de sélectionner les informations les plus spécifiques limitant ainsi la complexité du raisonnement. Ceci revient à intégrer raisonnement déductif et raisonnement abductif.

Dans le cadre de la Recherche d'Information

S est une requête (OBJECTIF).

O est un document.

On peut se ramener à l'évaluation de « $O \rightarrow S$ » (van Rijsbergen, 1986) et « $S \rightarrow O$ » peut jouer un rôle secondaire (voir l'exemple donné en 2.3.2.3).

De plus, de façon similaire à l'estimation de la correction des règles, les mesures les plus utilisées pour l'évaluation des performances d'un système de Recherche d'Information sont la mesure de précision qui correspond à : $\frac{\text{nombre documents trouvés et pertinents}}{\text{nombre de documents trouvés}}$, et la mesure de rappel qui correspond à : $\frac{\text{nombre documents trouvés et pertinents}}{\text{nombre de documents pertinents}}$. Cette double implication estime la pertinence globale des réponses du système, la mesure de précision qui correspond à « O→S » était d'ailleurs initialement appelée pertinence (Saracevic, 1975, p. 327).

Dans le cadre du modèle ACT

S est une situation comportant les trois éléments, représentés comme trois sources d'activation (OBJECTIF, IDENTITE DU SYSTEME, CONTEXTE COURANT).

O est la règle à appliquer.

Le principe est ici identique à celui de la sélection de règles à appliquer sur la base de leurs prémisses dans les systèmes experts. Les propositions dans les prémisses de règles de production peuvent rendre compte du but, de l'environnement perçu, de faits qui peuvent être produits de façon interne.

Autres cadres

Jusqu'à maintenant la pertinence se définit par l'adéquation de l'objet pertinent vis-à-vis d'un autre objet comme la requête, l'attribut à prédire ou une situation plus globale. La pertinence des exemples en apprentissage (ceux qui remettent en cause les prédictions faites à partir des exemples précédents) et la pertinence définie par Sperber et Wilson (1989) (plus une information remet en cause les croyances, plus elle est pertinente) expriment une toute autre idée, celle d'une nouveauté, d'un changement. On peut néanmoins en rendre compte de la même façon en considérant la notion d'état courant désignant précisément l'état des connaissances du système résultant de la prise en compte de l'exemple.

Sélection d'un exemple pertinent

S est l'ensemble des règles caractérisant les connaissances du système et prenant en compte le nouvel exemple.

O le nouvel exemple.

Plus l’exemple remet en cause les règles extraites à partir des exemples précédents, plus l’exemple est spécifique à l’état courant, (« O→S »). Si l’exemple ne remet en cause aucune règle, il n’est pas spécifiquement associé à l’état courant mais aussi aux autres états.

Selon la théorie de Sperber et Wilson

S est l’ensemble des croyances de l’individu révisées par l’information communiquée.

O est l’information communiquée.

Selon le critère d’effet, une information O est pertinente si l’effet est fort, c’est-à-dire si l’information communiquée conduit à une révision des croyances. Le principe est ici identique à celui d’exemples pertinents pour l’apprentissage. Ce qui est pertinent est ce qui est spécifique à l’état courant (« O→S »). Le critère d’effort renvoie à la facilité avec laquelle l’information est intégrée. Elle l’est d’autant plus qu’elle est en accord avec les différentes croyances et qu’elle peut être inférée à partir de ces dernières soit « S→O ».

2.4.1.3 Un lien privilégié avec la notion d’interaction

La recherche d’information peut gagner à être conçue comme une recherche interactive. Que l’interaction se déroule avec une machine ou avec un être humain, on peut en effet penser qu’il existe initialement deux systèmes dont les connaissances sont différentes et que la difficulté à répondre de façon satisfaisante à une requête (c’est-à-dire qui satisfasse celui qui énonce la requête) est liée à la distance qui sépare les deux interlocuteurs. On peut aussi penser que l’interaction consiste à rapprocher les connaissances des deux systèmes au moins sur le problème exprimé dans la requête. Le second principe de pertinence donné par Sperber et Wilson (1989) est de ce point de vue, une forme de métaconnaissance qui peut accélérer le processus de rapprochement. En supposant que l’information communiquée est celle qui a été jugée pertinente par l’interlocuteur, il est possible de déceler l’implicite (comme dans le cas de l’exemple de la lettre à poster dont le tarif n’est pas spécifiquement indiqué) et d’inférer (même lorsque l’implicite n’a pas été convenablement traité) les connaissances de l’interlocuteur en considérant ce qui pour lui allait de soi. Pour Saracevic (1975), la pertinence d’un point de vue général est une mesure d’efficacité d’un échange entre une source et une destination dans le cadre d’une communication. Il semble que pertinence et communication soient intimement liées, d’une part, parce que l’interaction permet un gain dans la pertinence des réponses, et d’autre part, parce qu'un principe reposant sur la notion de pertinence est en mesure d'expliquer la communication. Nous reviendrons sur ce point par la suite.

2.4.1.4 Construction et exploitation d'une structure

Dans tous les domaines abordés, on retrouve l'idée de construction et d'exploitation d'un réseau rendant compte des dépendances conditionnelles. En Intelligence Artificielle, il est aussi bien question de pertinence au niveau de la construction des règles (apprentissage) qu'au niveau de leurs utilisations (choix des règles à appliquer). En Recherche d'Information, les premiers modèles consistaient à rechercher les termes présents dans la requête parmi les termes représentant le document, mais maintenant la plupart des systèmes de recherche d'information utilisent un thésaurus qui rend compte des connaissances sur les liens entre les termes. Ainsi la connaissance de la synonymie de deux mots permet de retrouver un document comportant l'un des deux mots synonymes à partir d'une requête contenant l'autre mot. De même, le modèle de la mémoire sémantique basé sur l'association et la propagation d'activation et le compromis effet/effort évoqué par Sperber et Wilson (1989) rendent compte de cette idée.

On peut donc concevoir un processus d'extraction qui se décompose en un processus de synthèse qui construit des associations reflétant les dépendances entre les informations (résumant ainsi les régularités de l'environnement) et un processus de sélection qui exploite cette structure pour, dans une situation donnée, extraire les informations liées à la situation.

Les deux processus sont bien distincts, mais ils sont intimement liés, car la sélection dépend de la structure construite et les modifications de la structure dépendent de l'information sélectionnée.

2.4.2 La cognition située : une approche explicative

La pertinence est une notion que l'on ne peut définir que relativement à une situation. La notion de pertinence apparaît ainsi intimement liée à la relation qui peut exister entre cognition et situation. C'est cette relation que développent les partisans du courant de pensée action située (Norman, 1993 ; Clancey, 1997). Ainsi, il est illusoire de chercher à comprendre le système cognitif sans considérer le milieu dans lequel il évolue. Selon Varela (1989), autant d'un point de vue phylogénétique (à l'échelle de l'espèce) que d'un point de vue ontogénétique (à l'échelle de l'individu), le système est intimement lié à son histoire. Du fait que certaines caractéristiques sont relativement stables à l'échelle de l'individu (notre perception de la couleur par exemple), on peut être amené à penser, que la couleur est une donnée du monde externe et qu'elle préexiste à notre perception. Pourtant, d'autres espèces ont élaboré des mondes chromatiques différents (certains oiseaux semblent tetrachromats alors que nous sommes trichromats) et plus généralement des mondes différents (Nagel, 1974). L'information provenant de l'environnement est donc traitée en fonction de l'identité du système.

Dans l'autre sens, l'environnement n'est pas non plus à négliger. Par exemple, Rosch (1976) souligne le fait que notre environnement comporte des régularités fortes, que certaines co-occurrences de traits sont beaucoup plus fréquentes que d'autres. Ainsi "ailes et bec" est plus fréquent que la co-occurrence "pelage et bec". La relation entre le système et son environnement doit donc être vue comme un couplage (figure 2.8).

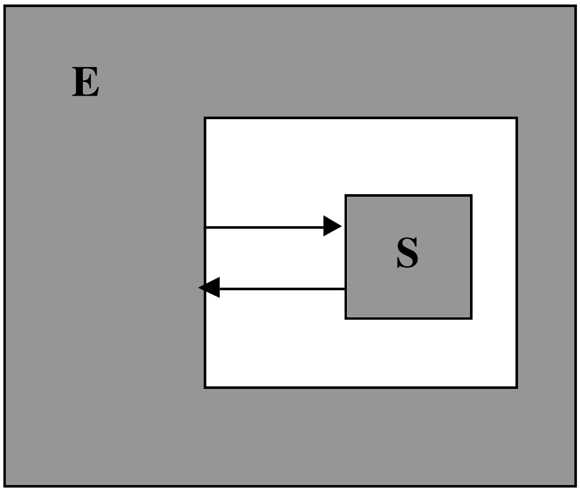


FIG 2.8 Le couplage du système (S) avec son environnement (E). Extrait de Clancey (1997).

Comme nous l'avons vu précédemment, Sperber et Wilson (1989) dégagent un principe général de la cognition basé sur la notion de pertinence. Ils précisent aussi que dans l'évaluation de la pertinence, le critère du minimum d'effort peut jouer un plus grand rôle que le critère du maximum d'effet, ce dernier n'étant pas toujours facilement estimable. Dans ces conditions, ils répondent à l'interrogation de la validité d'un système basé uniquement sur un critère d'effort, par une argumentation liée à la perspective auto-située. Ainsi, ce qui est facile à saisir pour le système cognitif correspond aux effets cognitifs les plus importants puisque le système est adapté à son environnement.

Dans cette perspective, l'état du système cognitif détermine la pertinence de l'information. L'extraction ne correspond pas à un processus actif d'évaluation basé sur une suite d'inférences recherchant d'éventuels liens de la nouvelle information avec la tâche en cours, mais plutôt à une superposition de l'information avec l'état du système. Si cette information s'intègre bien dans la structure préexistante, elle est retenue. Dans le cas contraire, elle correspond à un effort d'intégration trop important et par conséquent le système cognitif n'en tient pas compte.

En résumé, cette approche est explicative, en particulier parce qu'elle met en avant l'aspect subjectif de la pertinence. On notera que cette approche renforce et explicite le lien qui existe entre pertinence et interaction que nous évoquions précédemment.

2.4.3 Les grands principes d'un système d'extraction d'informations pertinentes

Nous proposons des réponses aux questions posées initialement afin d'extraire les principes généraux d'un système d'extraction d'information pertinente.

Ce qui ressemble ou ce qui est différent ?

Dans l'introduction, nous avons distingué deux cas, dans lesquels la notion de pertinence semblait renvoyer à des mesures opposées. Ainsi, dans le cadre de la recherche documentaire, si la requête porte sur les oiseaux migrateurs, un document portant sur les cigognes semble bien s'appliquer et on peut penser modéliser la pertinence en définissant une mesure de similarité du thème de la requête et de celui du document. Les études sur la pertinence se ramèneraient donc à comparer différentes mesures de similarités comme le font Jones et Furnas (1987).

Dans le cadre de la surveillance épidémiologique, supposons l'interrogation d'une base de données au sujet d'une maladie particulière, par exemple la varicelle. Une réponse indiquant une période particulière, juin 89 par exemple, s'applique bien si effectivement une épidémie s'est déclarée à ce moment là. On peut penser, dans ce cas, modéliser la pertinence en définissant une mesure du changement pour mesurer les recrudescences de cas. Les études sur la pertinence se ramèneraient cette fois-ci à la sélection de mesures de distances appropriées.

La pertinence renverrait donc à des notions différentes et ne pourrait être modélisée de façon unique. Pourtant en examinant le problème de plus près, il semble que la mesure de l'implication puisse rendre compte à la fois de ce qui est différent et de ce qui est identique.

En effet, si on sélectionne les objets O_i pour lesquels l'évaluation de l'implication « $O_i \rightarrow S$ » (qu'on note $P(O_i \rightarrow S)$) est la plus grande, on peut à la fois obtenir les documents qui s'appliquent le mieux à une requête et les informations concernant d'éventuels changements par rapport aux situations habituelles dans le cas par exemple de la surveillance épidémiologique. Dans les deux cas, l'objet pertinent O est tel que $P(O \rightarrow S)$ est maximal. On a en effet P(Thème = Cigogne→Thème = Oiseau migrateur)> P(Autres thèmes→Thème = Oiseau Migrateur) et P(Mois = Juin 89→Maladie = Varicelle)>P(Autres mois→Varicelle).

Dans les deux cas, il faudra néanmoins tenir compte de l'implication inverse, même si elle joue un rôle secondaire de façon à ne pas extraire de documents trop spécifiques, ou des périodes comportant un nombre de cas trop réduit.

Comment mesurer une distance ou une similarité ?

Bien que la question ait perdu un peu de son intérêt puisque la notion de pertinence ne paraît plus aussi directement liée aux mesures de similarités ou de distance, nous verrons qu'elles interviennent néanmoins dans l'évaluation des implications. On peut noter que l'approche située nous enseigne qu'il est illusoire de rechercher une « bonne mesure », le degré de résolution du système pour chaque variable dépend de son identité. On peut, pour illustrer cette idée, penser par exemple à l'expertise d'un sommelier qui est, plus qu'un néophyte, en mesure de distinguer de faibles différences entre deux vins. Il est donc important de définir des mécanismes qui permettent de construire des distances de façon adaptée à l'utilisateur.

Plus globalement, Quels critères pour l'évaluation ?

La pertinence d'un objet peut se définir suivant différents critères. On a distingué trois grandes catégories de critères correspondant aux différentes composantes d'une situation. Tous les critères peuvent ne pas être pris en compte systématiquement.

Qui juge de la pertinence ?

Selon l'approche située, il est difficile de parler d'objectivité, on considère plutôt différents points de vue. Bien que des régularités existent dans l'environnement, celles-ci sont toujours perçues à travers le prisme déformant de l'identité du système et de ses objectifs.

La distinction entre « pertinence objective » (celle du système) et la « pertinence subjective » (celle de l'utilisateur) n'a donc plus la même importance, la pertinence relevant généralement d'un point de vue. Dans ces conditions la recherche interactive d'information qui privilégiera la compréhension mutuelle de l'identité de l'utilisateur par le système et celle du système par l'utilisateur est une solution.

Un autre point très différent concerne la quantité de connaissances du système liée à la quantité d'observations traitées. On peut en effet imaginer que parmi deux systèmes identiques, c'est-à-dire des systèmes percevant le même type d'informations (sensibles aux mêmes variables et mêmes propriétés et ayant initialement les mêmes connaissances), l'un dispose de plus d'informations et soit donc à même de fournir une réponse plus complète. La différence est cette fois-ci de nature quantitative et non qualitative. On peut donc penser que l'utilisateur d'un système d'extraction d'informations pertinentes est le véritable juge de la pertinence d'une information mais qu'un système modélisant la qualité de l'information pour l'utilisateur (ce qui correspond à définir des distances, des variables, des propriétés) soit en mesure de fournir une réponse plus complète.

Pour conclure sur cette question, l'approche située nuance la différence entre critères subjectifs et critères objectifs, tous les critères étant plus ou moins subjectifs. La question se trouve déplacée vers la différence d'identité des systèmes communiquant et leur interaction.

Sélection ou synthèse d'information ?

On a distingué deux processus distincts qui contribuent à l'extraction d'information pertinente. L'un consiste à établir des liens entre les informations pour établir une structure représentant les régularités de l'environnement. L'autre exploite cette structure pour, dans une situation donnée retrouver l'information pertinente. La question est : la règle extraite est-elle une information pertinente au même titre qu'une information sélectionnée dans une situation donnée ? On répondra affirmativement si l'on se fie au fait que l'on puisse voir le processus d'extraction de règle comme un processus de sélection et que l'on puisse exprimer la pertinence de la même façon que dans le cas de la sélection.

Unicité du sens ? Qu'est-ce que la pertinence ? Proposer une définition, un modèle…?

A la lumière de ce parcours, il semble que la mesure de la spécificité d'un objet O_i vis-à-vis d'une situation S, qui correspond à l'évaluation de « $O_i \rightarrow S$ » rende bien compte de la notion de pertinence. Même si on peut aussi y ajouter le rôle secondaire joué par l'implication inverse, « $S \rightarrow O_i$ ». Si on se place dans le cas où différentes informations sont en concurrence, l'information la plus pertinente est donc l'information la plus spécifique. Ce constat n'est pas sans évoquer la notion de résonance. En effet, un objet est dit en résonance avec un autre si sa réponse produit une amplification de l'objet source. Or ici, l'information est d'autant plus pertinente, qu'elle est spécifique et donc qu'elle répond de manière forte à la situation (figure 2.9).

Nous proposons donc les grands principes d'un système d'extraction d'information pertinente suivants:

- Le système doit construire une structure qui rende compte des régularités de l'environnement. Ces régularités ont la forme de dépendances conditionnelles. Par ce processus, le système extrait la structure du monde qui l'entoure s'adaptant ainsi à son environnement.
- Le système doit pouvoir exploiter cette structure pour, dans une situation donnée, sélectionner l'information adaptée. La situation comporte trois éléments : l'objectif du système, son identité (connaissances, préférences) et le contexte courant (ce qui est perçu dans son environnement). La mesure de la pertinence de l'information repose sur une double mesure de dépendance conditionnelle entre l'information et la situation, privilégiant la spécificité de l'information vis-à-vis de la situation.

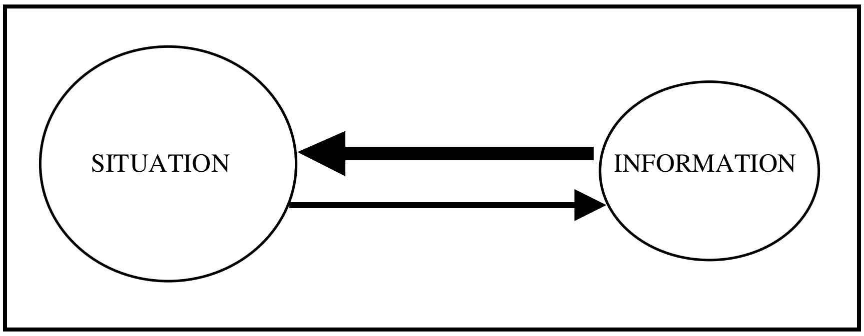


FIG. 2.9 La pertinence d'une information vis-à-vis d'une situation repose principalement sur l'évaluation de la spécificité de l'information vis-à-vis de cette situation.

Nous mettons donc en avant l'idée de couplage dans notre approche de la pertinence. On le retrouve en effet au niveau de l'interaction du système avec son environnement et au niveau de la caractérisation de l'information pertinente vis-à-vis d'une situation. Cette notion de couplage renvoie à la citation de Paul Valéry donnée en épigraphe dans laquelle la nécessité du nombre deux est la première idée importante comme le note Varela (1989) auquel la citation et l’utilisation qui en est faite a été empruntée. Le lien du couplage entre un système et son environnement et celui présent au niveau de la caractérisation de l’information pertinente peut être fait si l’on considère que la recherche de pertinence maximale (mesurée par la double implication mettant en relation la situation et l'information) est le principe majeur guidant le système et que ce principe induit l’interaction du système avec son environnement.

Avant de proposer une solution plus précise au problème d’extraction d’information pertinente, nous allons parcourir différents systèmes de représentation qui pourraient remplir ce rôle.

Chapitre 3

> *Rather than relying on reasoning to intervene between perception and action, we believe activity mostly derives from very simple sorts of machinery interacting with the immediate situation. This machinery exploits regularities in its interaction with the world to engage in complex, apparently planful activity whithout requiring explicit model of the world…*
>
> *Before and beneath any activity of plan following, life is a continual improvisation, a matter of deciding what to do now based on how the world is now.*
>
> Agre et Chapman, *extrait de Clancey (1997)*

3 SYSTEMES DE REPRESENTATION DES CONNAISSANCES POUR L'EXTRACTION D'INFORMATION PERTINENTE

Ce chapitre correspond à un état de l’art sur les systèmes de représentation des connaissances adaptés à l’extraction d’information pertinente telle que nous l’avons définie dans le chapitre précédent. La connaissance représentée dans ces systèmes porte sur les dépendances entre les propriétés des différentes observations. Adoptant le point de vue exprimé par Agre et Chapman (Clancey, 1997), nous nous en tenons à cette simple connaissance, ne considérant donc pas des modèles plus explicites du monde.

Nous traitons ici de systèmes en mesure d’extraire des régularités et de les utiliser par la suite dans une tâche de sélection. Nous nous intéressons en particulier à différents types de systèmes de représentation développés en Intelligence Artificielle qui ont la particularité de ne pas considérer une décision sur un attribut déterminé dans leur phase d’apprentissage. Nous souhaitons en effet permettre, par la suite, à la situation dont un des éléments constitutifs est justement l’objectif, de déterminer l’information pertinente. Il s’agit donc de systèmes dans lesquels la structure rendant compte des régularités s’auto-organise à mesure de la prise en compte d’observations provenant de l’environnement.

Nous donnerons, dans un premier temps, un descriptif assez bref de systèmes de représentation que nous avons classés en trois grandes familles : les réseaux neuronaux, les réseaux bayésiens et les systèmes à base de règles. Nous donnerons, dans un second temps, une présentation plus détaillée des réseaux sémantiques.

3.1 Un tour d'horizon des différentes méthodes utilisées en Intelligence Artificielle et en Recherche d'Information

Cette partie se veut descriptive. L’objectif est de donner un pointeur sur des systèmes existants respectant les principes cités ci-dessus. On emprunte ces systèmes à de grandes familles dont la frontière n'est pas clairement délimitée. On respectera tout de même un ordre en partant des méthodes les plus numériques pour arriver aux méthodes les plus symboliques, terminant par les systèmes à base de règles qui fournissent une introduction aux réseaux sémantiques sur lesquels nous nous attarderons ensuite.

Nous partons du simple modèle vectoriel utilisé en Recherche d’Information qui fait des hypothèses d’indépendance à partir desquelles, nous motivons la recherche de représentations plus complexes qui rendent compte des dépendances entre les informations fournies en entrée du système.

3.1.1 Le modèle de l’espace vectoriel

Le modèle de l’espace vectoriel est un modèle d’école en Recherche d’Information. Il consiste à représenter les requêtes et les documents par un vecteur dans un espace prédéterminé de caractéristiques. Dans le cas de la recherche documentaire, ces caractéristiques φ_i correspondent généralement à un ensemble de termes (mots-clefs, par exemple) et le vecteur représente le poids de chacun des termes. Si on considère m caractéristiques, on peut par exemple, écrire :

$$\vec{d_j} = (a_{1,j},\dots,a_{i,j},\dots,a_{m,j})$$

$$\vec{q} = (b_1,\dots,b_i,\dots,b_m)$$

où $\vec{d_j}$ est le vecteur représentant le document, $a_{i,j}$ est le poids de la caractéristique φ_i pour le document d_j, $\vec{q}$ est le vecteur représentant la requête et b_i est le poids attribué à la caractéristique φ_i dans la requête. Dans le cas de la recherche documentaire, ce poids est par exemple fixé à 0 si le terme correspondant n'apparaît pas dans la requête.

Ces vecteurs peuvent être calculés automatiquement. Ainsi, pour chaque caractéristique φ_i, on compte son nombre d'occurrences dans le document d_j soit $ff(\varphi_i,d_j)$ et son nombre d'occurrences dans la requête q soit $ff(\varphi_i,q)$. De plus, on calcule pour chaque caractéristique φ_i un terme noté $idf(\varphi_i)$ (pour inverse document frequency) qui rend compte de la rareté de la caractéristique dans l'ensemble des documents.

$$idf(\varphi_i) = \log\left(\frac{1+n}{1+df(\varphi_i)}\right)$$

où n est le nombre de documents et $df(\varphi_i)$ est le nombre de documents ayant la caractéristique φ_i. On remarquera que si la caractéristique φ_i est partagée par tous les documents, son poids $idf(\varphi_i)$ est nul. Ce terme permet d'accorder plus d'importance aux caractéristiques spécifiques au document et moins d'importance aux caractéristiques très générales, partagées par un grand nombre de documents.

On définit alors $a_{i,j}$ et b_i de la façon suivante :

$$a_{i,j} = ff(\varphi_i,d_j)idf(\varphi_i)$$

$$b_i = ff(\varphi_i,q)idf(\varphi_i)$$

La fonction qui mesure l'adéquation d'un document à la requête est la suivante :

$$\rho_{\cos}(q,d_j) = \frac{\sum_{i=1}^{m} a_{i,j}b_i}{\sqrt{\sum_{i=1}^{m} b_i^2}\sqrt{\sum_{i=1}^{m} a_{i,j}^2}}$$

Cette expression mesure à quel point le produit scalaire des deux vecteurs est grand comparé à la longueur du document et la longueur la requête. Ce modèle donne de relativement bons résultats, cependant, il fait de fortes hypothèses d'indépendance en considérant des axes orthogonaux. Or, ces hypothèses ne sont généralement pas vérifiées. Par exemple, dans le cas de la recherche documentaire où chaque dimension correspond à un terme, certains termes ont un sens très proche, d'autres au contraire ont un sens très éloigné. Le modèle de l'espace vectoriel doit donc être enrichi pour permettre de prendre en compte les dépendances entre les différentes caractéristiques. Les systèmes de représentation présentés dans la suite offrent cette possibilité.

3.1.2 Réseaux neuronaux

Un réseau de neurone artificiel correspond à un ensemble d'unités simples (les neurones) reliées par des connexions auxquelles sont associées des poids (les poids synaptiques). La connaissance est, dans ce système de représentation, stockée au niveau des poids synaptiques et l'apprentissage consiste en une mise à jour de ces poids à chaque nouvelle observation.

Il ne s'agit pas ici, d'une étude exhaustive de l'ensemble des réseaux de neurones qui supportent l'idée d'auto-organisation mais néanmoins, d'un ensemble de systèmes représentatifs. Les systèmes présentés ont tous la particularité de considérer un apprentissage associatif de type hebbien qui consiste à renforcer la connexion entre deux neurones dont l'activité est simultanée (Hebb, 1949). Plus généralement on notera que les deux premiers systèmes ont la particularité d'être d'inspiration neurophysiologique. La forme la plus générale d'apprentissage hebbien est décrite par la règle:

$$w_{ij(t+1)} = w_{ij(t)} + \eta x_{i(t)} x_{j(t)}$$

où $x_{i(t)}$ et $x_{j(t)}$ représentent respectivement l'activité des unités i et j à l'instant t, $w_{ij(t)}$ représente le poids synaptique liant les deux unités à l'instant t et η est une constante d'apprentissage qui contrôle la plasticité de la synapse. Le poids synaptique liant les unités i et j est augmenté si une activité simultanée des deux unités se produit.

3.1.2.1 Les cartes auto-organisatrices (*Self-Organising Maps*)

Ce type de réseau de neurones a été introduit par Kohonen (1982). Le principe est celui d'une carte, c'est-à-dire, d'un ensemble de neurones répartis de façon à quadriller une zone rectangulaire (figure 3.1), qui s'adapte au cours de son apprentissage aux données placées en entrée. Cette adaptation consiste :

- A reproduire l'espace des entrées en associant chaque partie de l'espace des entrées à une zone particulière de la carte, et plus précisément chaque entrée à un neurone préférentiel.

- A reproduire au niveau de la carte la répartition des stimuli de l'espace des entrées. Ainsi, les neurones répondant respectivement de façon préférentielle à deux stimuli différents sont éloignés si les stimuli sont éloignés, et voisins si les stimuli sont voisins (figure 3.2).

- A reproduire les différences de densités de l'espace des entrées. Ainsi, si une partie de l'espace des entrées est particulièrement dense, la zone de la carte la représentant est particulièrement importante, ce qui correspond à une adaptation de la granularité du système à son environnement.

- A isoler des dimensions pertinentes dans les données.

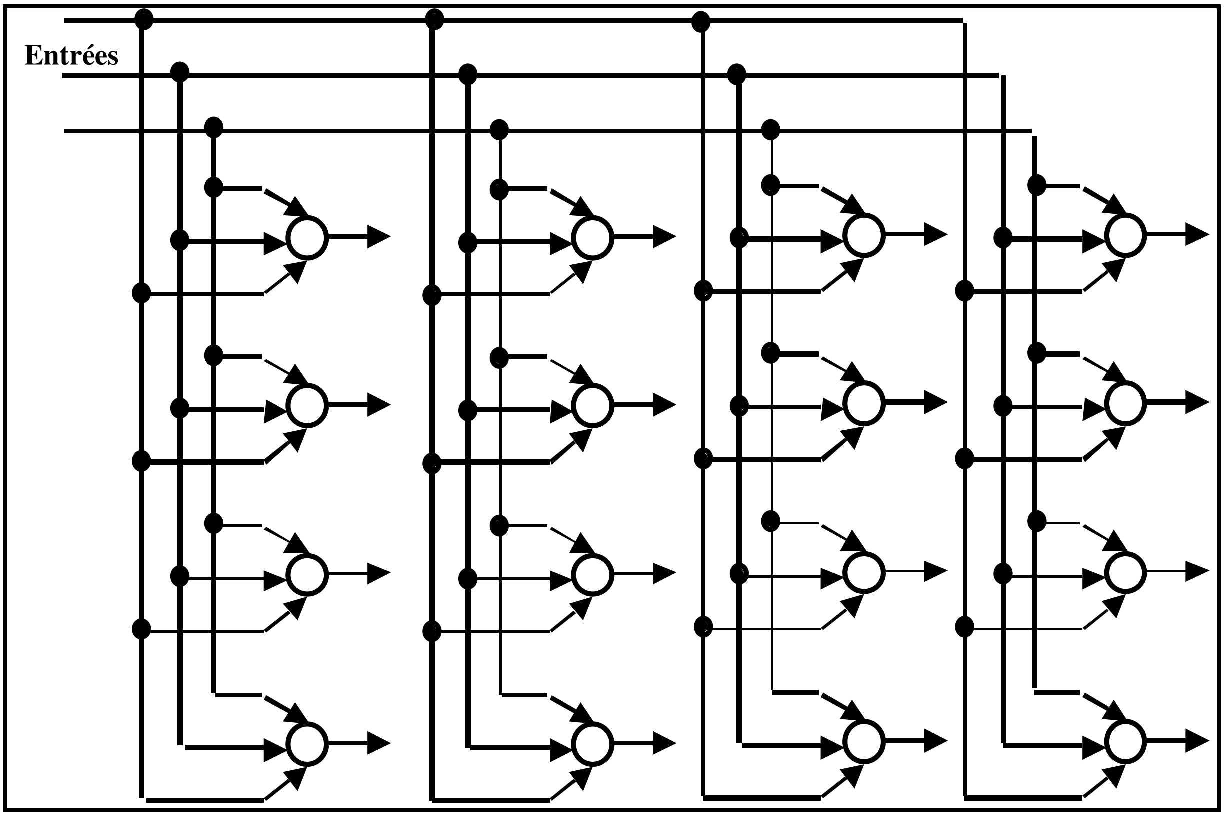


FIG. 3.1 Architecture d'une carte auto-organisatrice. On remarquera que les neurones de la carte ne sont pas connectés entre eux mais qu'ils sont tous connectés aux entrées (ici trois entrées sont représentées).

Les principes de construction de la carte sont ceux de compétition et de voisinage. L'apprentissage va consister à calculer, pour chaque neurone, ses connexions avec les neurones d'entrée sur la base des différentes observations en partant d'un ensemble de poids synaptiques tirés aléatoirement. L'algorithme d'apprentissage se décompose de la façon suivante :

1- Pour chaque nouvelle entrée, on détermine le neurone dont l'activité est maximale (principe du « winner take all »).

2- On propage l'activité du neurone vainqueur à son voisinage (défini sur la base de la distance sur la carte).

3- Enfin, on met à jour les connexions des neurones activés selon un principe associatif renforçant le lien des neurones activés aux neurones représentant l'entrée courante. Et ainsi de suite jusqu'à ce qu'il n'y ait plus de changement notable de la carte.

La convergence de l'algorithme est facilitée par la réduction progressive au cours de l'apprentissage, du voisinage et du paramètre réglant le degré avec lequel il est possible de modifier une connexion sur la base d'une observation. La règle d'apprentissage qui définit la mise à jour des connexions du neurone j aux neurones d'entrée en considérant que le neurone i est le neurone vainqueur est la suivante :

$$w_{j(n+1)} = w_{j(n)} + \eta_{(n)} h_{j,i(x)(n)} (x - w_{j(n)})$$

On retrouve la forme générale de la règle de Hebb $w_{j(n+1)} = w_{j(n)} + \eta_{(n)} h_{j,i(x)(n)} x$ où la fonction déterminant le voisinage $h_{j,i(x)(n)}$ représente l'activation du neurone j sachant que i est le neurone vainqueur. Le terme $\eta_{(n)} h_{j,i(x)(n)} w_{j(n)}$ est un terme d'oubli qui permet de faire varier la connexion dans les deux directions et d'éviter une croissance illimitée du poids synaptique.

Les cartes de Kohonen sont utilisées en IA pour extraire des concepts dans cadre d'un apprentissage non supervisé. Elles sont aussi utilisées en Recherche d'Information où elles permettent par exemple de rendre compte de la distance entre différents termes ou différents documents (la proximité des neurones représentant les différents termes ou documents rendant compte de leur plus ou moins grande similarité).

En considérant par exemple, un ensemble d’entrées correspondant à différents animaux décrits par leur mode de déplacement, leur taille et différentes caractéristiques (nombre de pattes, plumes, pelage,…), on obtient après apprentissage la carte suivante :

chien	.	.	renard	.	.	chat	.	.	aigle
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	hibou
.	.	.	.	.	.	tigre	.	.	.
loup	.	.	.	.	.	.	.	.	faucon
.	.	.	lion	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	colombe
cheval	.	.	.	.	.	.	poule	.	.
.	.	.	.	vache	.	.	.	.	oie
zèbre	.	.	.	.	.	.	canard	.	.

FIG. 3.2 Carte créée à partir d’un apprentissage basé sur les caractéristiques de différents animaux. Le nom de l’animal est placé au niveau du neurone répondant le plus à sa description à l’issue de l’apprentissage. On peut constater que la répartition obtenue rend relativement bien compte des regroupements que l'on pourrait faire entre ces différents animaux. Extrait de Haykin (1999).

3.1.2.2 Les réseaux de Hopfield

Les réseaux de Hopfield apparus aussi au début des années 80 (Hopfield, 1982) sont des réseaux de neurones récurrents, c’est-à-dire des réseaux dans lesquels il existe des boucles (figure 3.3). Cette caractéristique en fait des systèmes dynamiques, la modification de la valeur de sortie d’un neurone induisant la modification de ses entrées. Les notions de convergence, de stabilité, de point fixe (point de stabilité), d’attracteurs (point vers lequel converge le système) et de bassin attracteur (zone dans laquelle la convergence conduit au point attracteur) sont particulièrement importantes.

Ce système permet d’adresser une mémoire par son contenu. Ainsi, dans une première phase, un apprentissage de type hebbien permet de prendre en compte différentes observations en calculant les poids synaptiques entre les différents neurones. Ensuite dans une phase de récupération, il est possible de retrouver l’ensemble des caractéristiques définissant l’observation à partir seulement d’un petit nombre de caractéristiques éventuellement bruitées en activant les neurones correspondant à cette observation et en laissant converger le réseau vers un état stable correspondant à une des entrées donnée lors de l'apprentissage.

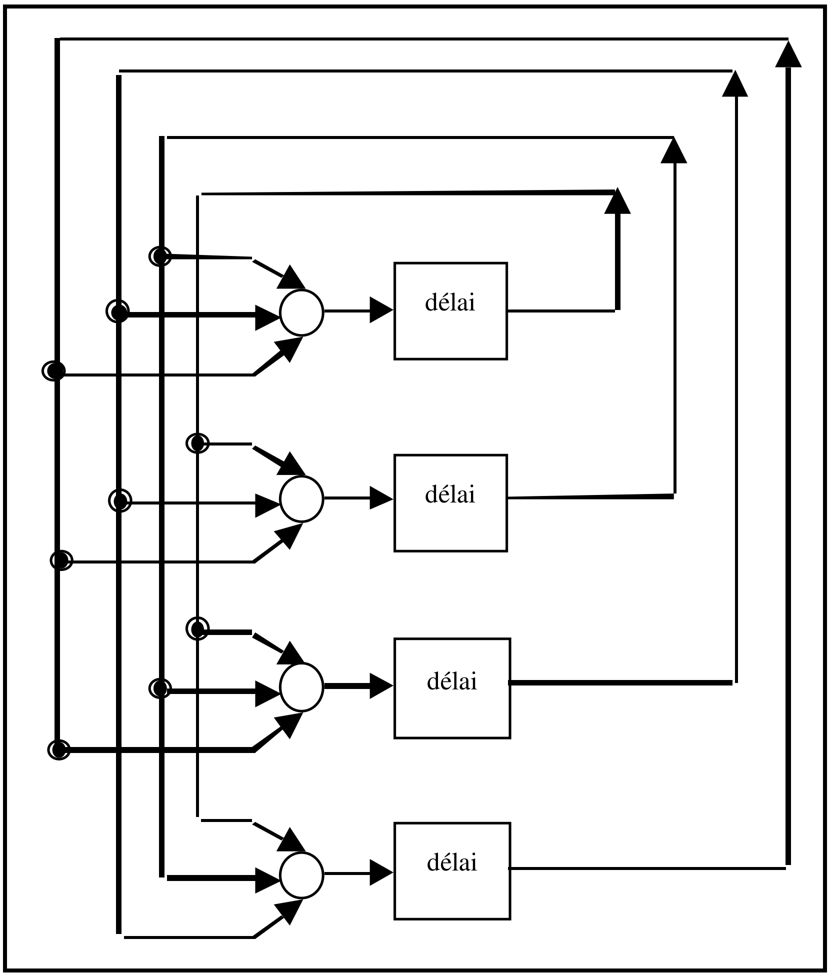


FIG. 3.3 Architecture d'un réseau de Hopfield composé de 4 neurones. On notera que la sortie de chaque neurone est une entrée (via un délai qui permet de synchroniser les opérations dans le réseau) pour les autres neurones.

Pour présenter une formalisation simplifiée, on considère ici que $O_{m,i}$ désigne la i^{ème} caractéristique de l'observation O_m et que l'on se place dans le cas discret où cette caractéristique est soit présente, soit absente. De plus, on associe à la i^{ème} caractéristique, le neurone i dont l'activation est +1 si la caractéristique est présente et –1 si elle est absente.

La phase d'apprentissage consiste à calculer le poids synaptique w_{ij} de la connexion le neurone i au neurone j. A l'issue de la prise en compte des M observations, on a :

$$w_{ij} = \begin{cases} \frac{1}{N}\sum_{m=1}^{M} O_{m,j}O_{m,i} & j \neq i \\ 0 & j = i \end{cases}$$

On notera qu'ici le poids est symétrique et que le facteur (1/N) est utilisé pour introduire des simplifications dans les calculs et qu'il n'a pas de raison profonde.

Dans la phase d'exploitation, on initialise ensuite les neurones avec une valeur correspondant à une nouvelle observation éventuellement incomplète et on repète, jusqu'à l'obtention d'un point fixe, la mise à jour asynchrone (c'est-à-dire une mise à jour à la fois et le neurone est choisi aléatoirement) de l'activation des neurones selon la formule suivante :

$$x_{j(n+1)} = \mathrm{sgn}\left[\sum_{i=1}^{N} w_{ji}x_{i(n)}\right]$$

où $x_{j(n+1)}$ désigne l'activation du neurone j à l'étape $n+1$, Nest le nombre d'entrées de chaque neurone (on a donc $N+1$ neurones) et où sgn désigne la fonction signe.

Ce type de réseau peut par exemple être utilisé pour reconnaître une forme (par exemple un chiffre) malgré une déformation par rapport à la version originale présentée lors de l'apprentissage. La forme est reconnue lorsque le réseau se stabilise dans l'état qui caractérise la forme originale. En Recherche d'Information, Chen et Ng (1995) ont proposé d'explorer un thésaurus dans l'objectif d'étendre une requête à un ensemble de mots plus important en utilisant ce type de techniques.

3.1.2.3 Les cartes cognitives floues (*Fuzzy Cognitive Maps*)

Le troisième et dernier réseau de neurone que nous présentons est de nature moins numérique que les deux précédents. Il s'agit d'un réseau dans lequel on associe sans apprentissage, à chaque neurone, un sens particulier. Ce type de réseau peut-être rattaché à la famille des méthodes neuro-floues. Les *cognitive maps* ont été introduites dans les années 70 pour représenter des connaissances de type politiques (figure 3.4). Dans ce réseau, les nœuds sont des concepts et les liens représentent des relations de causalité entre ces concepts et ces liens sont valués. Dans le cas des *Fuzzy Cognitive Maps* (FCMs), les concepts sont flous (ils ont des limites imprécises et sont représentés par des sous-ensembles flous) et les liens admettent des valeurs comprises entre -1 et 1. Dans les FCMs simples, la valeur des arcs est soit -1, soit 0, soit 1. La valeur 0 indique qu'il n'y a pas de causalité, la valeur 1 indique que lorsque le concept origine (Co) augmente, le concept arrivée (Ca) augmente aussi (plus le temps est mauvais, plus le nombre d'accidents de la route est important), et de même lorsque Co décroît, Ca décroît également. Une valeur -1 indique que, lorsque Co croît, Ca décroît et quand Co décroît, Ca croît.

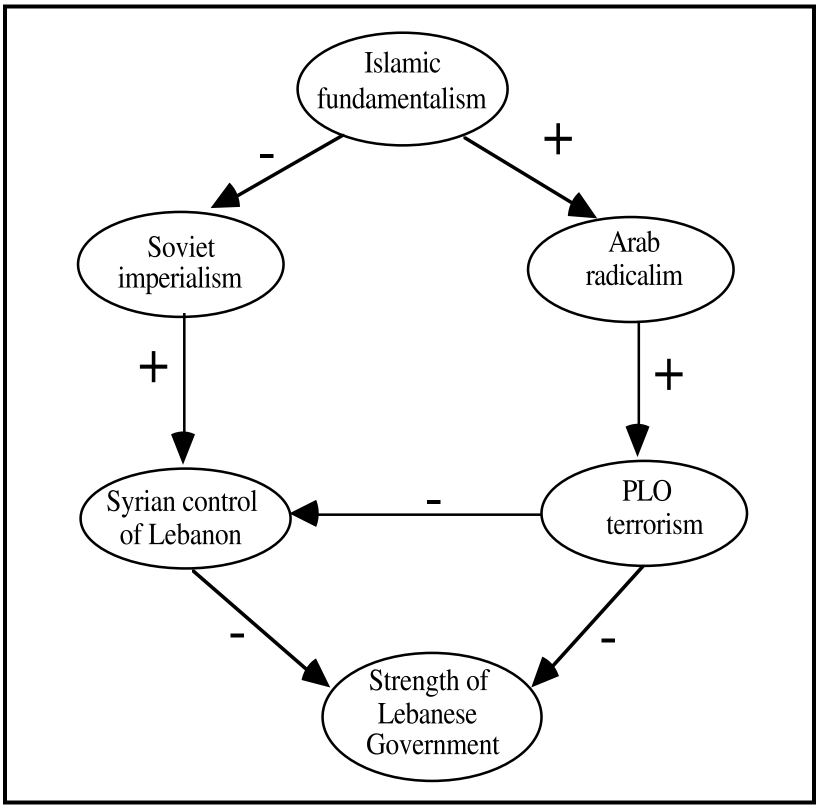


FIG 3.4 FCM construit à partir d'un article de Henry Kissinger. Extrait de Kosko (1986).

Ce type de système a donc la particularité de représenter des processus dynamiques. Le mode d'apprentissage utilisé est de type hebbien. Ainsi, la règle d'apprentissage consiste à modifier le poids synaptique entre le concept C_i et le concept C_j et peut s'écrire : $\frac{dw_{ij}}{dt} = -w_{ij} + C_i(t)C_j(t)$.

La modification du poids synaptique w_{ij} tient compte de deux termes différents. Le terme $-w_{ij}$ affaiblit la connexion (il correspond à un oubli). Le terme $C_i(t)C_j(t)$ correspond à l'activation simultanée des deux concepts. Il s'agit donc toujours du même type de règle auquel s'ajoute simplement ici la notion de temps. D'autres méthodes, cette fois-ci non locales, ont été récemment définies (Schneider & al., 1998). Dans les FCMs, le réseau de concepts peut comporter de nombreuses boucles et doivent donc être vus comme des systèmes dynamiques. On retiendra qu'ils permettent de prédire l'évolution d'un processus dynamique à partir de la donnée d'un état initial.

3.1.3 Les réseaux bayésiens

Les réseaux bayésiens (Pearl, 1988) sont des graphes, dont les nœuds correspondent à des variables aléatoires (par exemple la variable tabagisme qui peut prendre deux valeurs soit "fumeur" soit "non fumeur") et dont les arcs correspondent à des dépendances conditionnelles entre ces variables. Les arcs ne sont pas valués, ils indiquent simplement une dépendance. Les probabilités conditionnelles sont stockées au niveau de chaque nœud. Lorsqu'une information survient suite par exemple à la mesure d'une fréquence d'occurrence (proportion de fumeurs, par exemple), cette information est propagée dans le réseau, les probabilités des différentes variables sont mises à jour au niveau de chaque nœud.

La construction d'un réseau bayésien part de la décomposition de la loi de probabilité jointe en utilisant la formule de Bayes : $P(A \cap B) = P(A/B)P(B)$.

Ainsi, on peut écrire:

$$P(x_1,\ldots,x_n) = P(x_2,\ldots,x_n / x_1)P(x_1)$$

En renouvelant cette opération n fois on obtient finalement :

$$P(x_1,\ldots,x_n) = P(x_1)P(x_2/x_1)P(x_3/x_1,x_2)\ldots P(x_n/x_1,\ldots,x_{n-1})$$

En utilisant les indépendances conditionnelles entre certaines variables, il est possible de simplifier l'expression. On obtient alors une formule que l'on peut représenter par un graphe orienté où x_1 apparaît au sommet et constitue une source (aucun lien entrant) (figure 3.5). Ainsi, le graphe construit dépend de l'ordre d'énumération des différentes variables. Certains ordres engendrent des graphes de moins grande densité (nombre d'arcs par rapport au nombre de variables) que d'autres, ce qui entraîne des différences de lisibilité et de rapidité de calcul.

Deux types d'algorithmes ont été développés pour la mise à jour des probabilités dans le réseau. Le premier consiste en une transmission locale de messages entre variables voisines utilisant l'orientation des arcs (Kim & Pearl, 1985). La seconde opère préalablement des transformations dans le graphe afin d'obtenir un arbre de jonction non orienté (Lauritzen & Spiegelhalter, 1988).

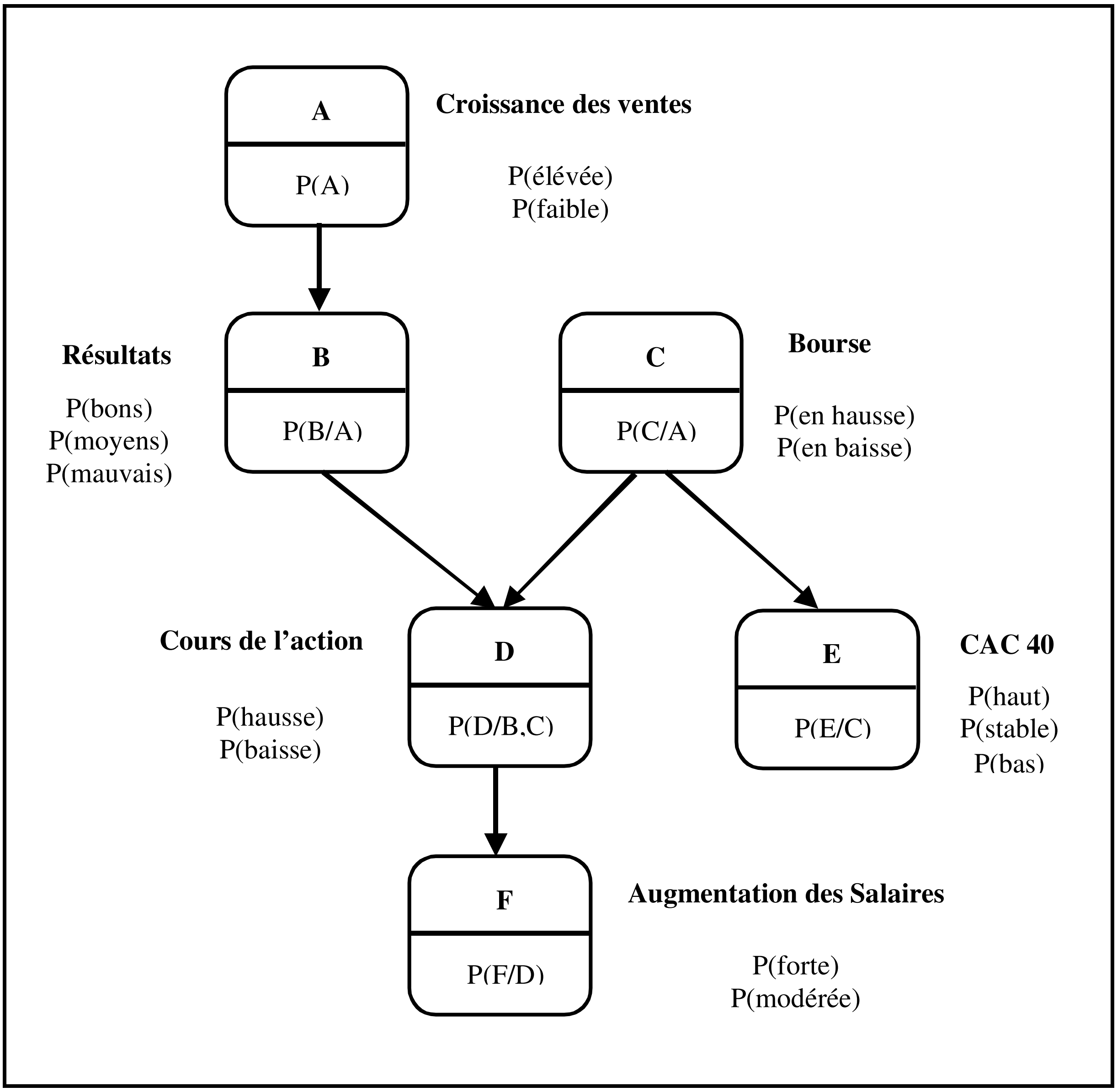


FIG. 3.5 Un exemple de réseau bayésien. L'évolution du cours de l'action d'une société (D) dépend de ses résultats (B) et du contexte boursier (C). La croissance des ventes (A) conditionne le niveau de ses résultats. Le contexte boursier influence la valeur du CAC 40 (E). Enfin, on suppose que l'augmentation du salaires des employés de la société (F) est liée à l'évolution du cours de l'action. En considérant, qu'il s'agit d'un énoncé exhaustif de toutes les dépendances, on a : P(A,B,C,D,E,F)=P(A)P(B/A)P(C)P(D/B,C)P(E/C)P(F/D). Extrait de Faÿ et Jaffray (1997).

Les réseaux bayésiens ont été largement utilisés en IA. Le système PROSPECTOR utilisé pour la prospection minière apporte par exemple une aide dans la recherche de zone de forage susceptible de correspondre à des gisements à partir de la donnée d'un ensemble de paramètres décrivant les différentes zones. Dans le domaine de la Recherche d'Information, ce type de représentation est aussi très répandu (Fuhr, 1992). Turtle et Croft (1990) utilisent par exemple un réseau bayésien dans lequel les documents apparaissent au sommet du réseau et sont reliés à un nœud représentant la requête via différents nœuds représentant des concepts. La recherche de documents pertinents vis-à-vis d'une requête consiste pour chaque document à définir une probabilité égale à 1 et à propager cette probabilité dans le réseau jusqu'au nœud représentant la requête. La pertinence du document est ensuite évaluée à partir de la probabilité de la requête.

3.1.4 Les systèmes à base de règles

3.1.4.1 Des systèmes pour la classification

Les règles sont une façon naturelle d'exprimer des dépendances entre les informations. Elles s'écrivent simplement sous la forme « SI *conditions* ALORS *conclusions* » où les conditions et les conclusions sont des conjonctions ou des disjonctions de propositions logiques. Bien qu'il existe des systèmes de règles utilisant des logiques d'ordre supérieur, nous limiterons ici notre propos à une logique propositionnelle considérant donc des couples attribut-valeur. Ces règles sont acquises auprès d'experts ou construites automatiquement à partir d'un ensemble d'observations.

Les algorithmes de génération de règles (Clark & Nibblet, 1989 ; Quinlan, 1986 ; Michalski, 1990) sont le plus souvent basés sur la prédiction de la valeur d'un attribut particulier (la classe). L'architecture qui en résulte en fait des systèmes dédiés à ce type de tâche et difficilement applicable à la prédiction de la valeur de multiples attributs. Ainsi, bien qu'il soit, par exemple, envisageable de construire un arbre de décision par objectif et d'utiliser ensuite un arbre particulier pour prédire chacun des attributs, l'information incomplète (s'il y a plusieurs attributs à prédire) constitue un obstacle. En effet, les attributs manquants peuvent apparaître dans l'arbre de décision d'un attribut que l'on cherche à prédire et ceci d'autant plus si l'approche est monothétique (comme les arbres de décision), c'est-à-dire, basée à chaque étape sur le choix de la valeur d'un attribut particulier (si l'attribut du sommet de l'arbre n'est pas renseigné, aucune prédiction n'est possible). En revanche, le système CHARADE (Ganascia, 1987) qui consiste en une exploration de l'espace des règles possibles plus exhaustive et qui ne se base pas systématiquement sur un attribut particulier à prédire est plus approprié pour la tâche d'extraction de dépendances entre les différentes propriétés caractérisant les observations.

3.1.4.2 Un système plus général : CHARADE

CHARADE se base sur la structure mathématique sous-jacente à tout problème d'induction. Cette structure est communément appelée « treillis de Galois ». Il s'agit plus précisément de deux treillis liés par une relation appelée correspondance de Galois (Barbut & Monjardet, 1970). On rappelle qu'un treillis est un ensemble ordonné tel que chaque couple d'éléments possède un plus petit majorant et un plus grand minorant et que, en considérant l'inclusion comme relation d'ordre, l'ensemble des parties d'un ensemble a une structure de treillis.

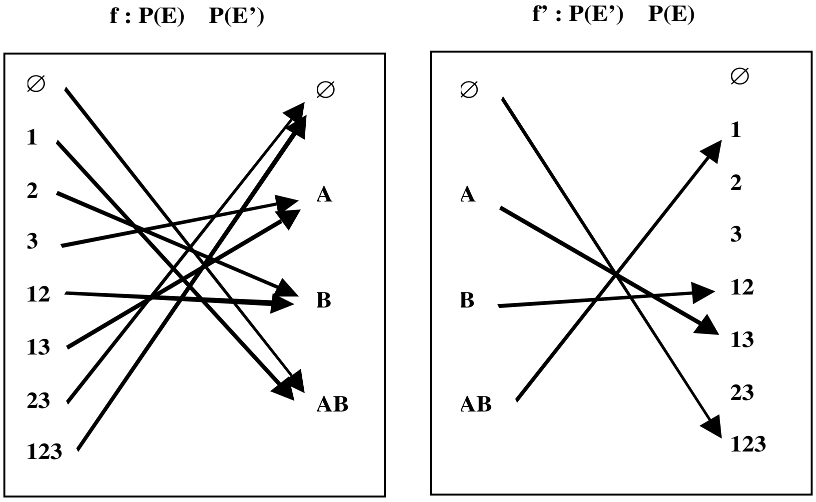
Dans le cadre de l'apprentissage inductif, l'ensemble d'apprentissage est la donnée d'une relation binaire liant deux ensembles, celui des exemples et celui des propriétés portées par les exemples. Cette relation se traduit par une correspondance entre deux treillis si l'on considère les parties de l'ensemble des exemples et celles de l'ensemble des propriétés.

Reprenons l'exemple proposé par Poitrenaud (1999) en considérant un ensemble d'apprentissage constitué de trois exemples : e_1 portant les propriétés A et B, e_2 portant la propriété B et e_3 portant la propriété A. On notera E=(1,2,3), l'ensemble des exemples et E'=(A,B) l'ensemble des propriétés (tableau 3.1).

	A	**B**
1	x	x
2		x
3	x	

TAB. 3.1 La relation binaire entre l'ensemble des exemples et l'ensemble des propriétés.

Considérons les applications f et f' qui lient P(E) (l'ensemble des parties de E) et P(E') (l'ensemble des parties de E') (figure 3.6).


f : P(E)   P(E')

f' : P(E')   P(E)

FIG. 3.6 Représentation des fonctions f et f' sur l'exemple.

On appelle le couple (f,f’) une correspondance de Galois. Si l’on construit h’=f’of, que l’on considère deux descriptions d, d’, correspondant à des conjonctions de propriétés, on peut constater que si h’(d) = d’, il est possible de conclure que d→d’, puisqu’alors, tous les exemples vérifiant les propriétés décrites dans d vérifient aussi les propriétés décrites dans d’. Partant de ce constat, CHARADE parcourt l’espace des descriptions (appliquant h’ à chaque nouvelle description) en partant de l’ensemble vide, puis en considérant successivement des propriétés simples puis des conjonctions de deux propriétés, de trois, etc. Différentes limitations ont été définies de façon à ne pas explorer la totalité des combinaisons de propriétés et à restreindre le nombre de règles.

Ainsi, si la règle d→c est trouvée au cours de l’exploration, il est inutile de rechercher les implications ayant comme prémisses d∧c car il s’agit en effet des mêmes que celles ayant d pour prémisse d’où l’intérêt de parcourir l’espace de recherche dans l’ordre croissant de la taille des conjonctions. On notera que cette première limitation n’introduit pas le risque de perdre des règles qui changeraient la conclusion du système. D’autres limitations sont envisageables, elles provoqueraient cependant une perte d'information. Ainsi, il est par exemple possible de se restreindre à des règles comprenant un nombre limité de propriétés en prémisse, s’appliquant à un minimum d’exemples, ou encore concluant sur un attribut particulier. De plus, il est possible d’associer des pondérations aux différentes règles contenant des exceptions et Ganascia (1987) propose de ne pas construire les règles redondantes qui ne modifient pas « suffisamment » la plausibilité des différentes conclusions.

Dans le cadre de la Fouille de Données, où les volumes de données sont particulièrement importants, des algorithmes limitant de façon plus drastique l’exploration ont été proposés. Il s’agit des algorithmes rapides de recherche d’associations.

3.1.4.2 Les algorithmes rapides de recherche d’associations

L’algorithme APRIORI proposé par Agrawal et al. (1995) constitue la référence de ce type d’algorithme à partir de laquelle différentes versions ont été développées. Il procède en deux étapes. Dans la première étape, il sélectionne toutes les conjonctions de propriétés dont l'effectif dans la base d'exemples (appelé support) dépasse un certain seuil. Pour construire ces conjonctions, il procède de manière incrémentale, considérant d'abord les conjonctions de n propriétés avant de considérer les conjonctions de n+1 propriétés. La simplification consiste à ne considérer qu'une partie des conjonctions possibles à l'étape n+1, en se basant sur les conjonctions retenues à l'étape n. Ainsi, si à l'étape 3 on a retenu les conjonctions, (1,2,3), (1,2,4), (1,3,4), (1,3,5), (2,3,4), la seule conjonction à examiner à l'étape suivante est (1,2,3,4). Il est par exemple inutile d'examiner (1,2,3,5) puisque (2,3,5) n'est déjà pas représenté suffisamment dans la base d'exemples.

Dans la seconde étape, il construit à partir des conjonctions résultantes, les règles dont la confiance (il s'agit simplement d'une fréquence conditionnelle, P(conclusion/condition)) dépasse un certain seuil (pas nécessairement le même que dans la première étape). Une nouvelle fois, une simplification permet de ne pas considérer toutes les règles. Cette simplification est basée sur l'observation que la confiance accordée à la règle ab→c est supérieure à celle attribuée à a→bc. Ainsi, pour une conjonction retenue par la première étape, on considère d'abord les règles ayant le plus grand nombre de conditions en prémisse et on envisage éventuellement ensuite, suivant la confiance calculée pour ces règles, celles ayant un nombre inférieur de conditions en prémisse.

L'Algorithme Apriori

Etape 1

<u>Principe de construction :</u>

Soit a, a' deux combinaisons de propriétés.

Si (a' inclus dans a) alors support(a')<support(a)

(par exemple a= « aile et plume et bec » et a'= « aile et plume »)

```
L1    Ensemble des singletons s (propriété simple) tels que support(s)>seuil1
k  2
Tant que $L_{k-1} \neq$ vide
Faire
$C_k$    Ensemble des combinaisons de taille k candidates à partir de $L_{k-1}$
Pour chaque combinaison c de $C_k$
Faire
Si (support(c)>seuil1) ajouter c à $L_k$
            Fin Faire
      k  k+1
Fin Faire
```

Etape 2

Principe de construction :

Soit l une combinaison de propriétés, x une combinaison incluse dans l et x’, une combinaison incluse dans x. On a nécessairement confiance(l-x →x)>confiance(l-x’ →x’)

```
$R_1$   Ensemble des règles comprenant un singleton en conclusion
(construite à partir de l’ensemble des combinaisons trouvées à l’étape 1)
Tant que $R_k \neq$ vide
Faire
$R_k$   générer les règles candidates comprenant k éléments en conclusion.
        Pour chaque r de $R_k$
        Faire
        Si confiance(r)>seuil2 alors ajouter c à $R_k$
        Fin Faire
k  k+1
Fin Faire
```

Cet algorithme permet de traiter de grands volumes de données (jusqu’à plusieurs millions d'exemples comportant chacun 20 attributs). Certaines variantes de cet algorithme ont été développées, notamment une version incrémentale (Cheung, 1996a ; 1996b) et une version permettant de traiter convenablement des données incomplètes (Ragel, 1998). Si l’on considère l’ensemble de ces règles comme un réseau associatif de propriétés, on se ramène à la notion de réseau sémantique que nous détaillons dans la suite.

3.2 Les réseaux sémantiques

L'acception du terme de réseau sémantique est en fait très large. Il s'agit d'un graphe (ou plus précisément d'un multigraphe si on considère que deux nœuds peuvent être reliés par plusieurs arcs) orienté et étiqueté (Kayser, 1997 , chap. 5). Pour le reste, Sowa (1991) considère les caractéristiques suivantes :

- Les nœuds dans le graphe représentent des objets, des propriétés, des événements ou des états.
- Les arcs dans le graphe, (souvent appelés relations conceptuelles), représentent des relations entre les concepts. L'arc est muni d'une étiquette qui indique le type de la relation.

Les relations spécifient souvent le rôle des objets les uns par rapport aux autres (agent, objet, instrument) et on notera, à ce sujet que la représentation des rôles par des relations ou des concepts est un sujet de divergence. Par ailleurs, la relation hiérarchique notée IS_A spécifiant qu'un concept est plus général ou spécifique qu'un autre, aussi appelée relation de subsomption a été particulièrement étudiée. Elle permet d'organiser les connaissances en hiérarchies de catégories auxquelles on peut attacher des mécanismes d'héritage. Ainsi, si Socrate est un humain, il hérite des propriétés propres aux humains mais aussi des propriétés propres aux animaux (si la hiérarchie considère que les humains sont des animaux). On notera aussi que le lien IS_A traduit une inclusion et que, de ce fait, il renvoie à la notion d'implication.

3.2.1 Historique

On attribue à Quillian (1968), pour ses travaux sur la mémoire sémantique, l'introduction des réseaux sémantiques (voir chapitre 2). Sowa (1991) précise néanmoins que l'idée de représenter les connaissances par un graphe dans lequel les nœuds représentent des concepts et les arcs, des relations entre ces concepts, n'est pas récente. En effet, au troisième siècle déjà, le philosophe grec Porphyre représentait les relations hiérarchiques des catégories d'Aristote par un arbre (figure 3.7).

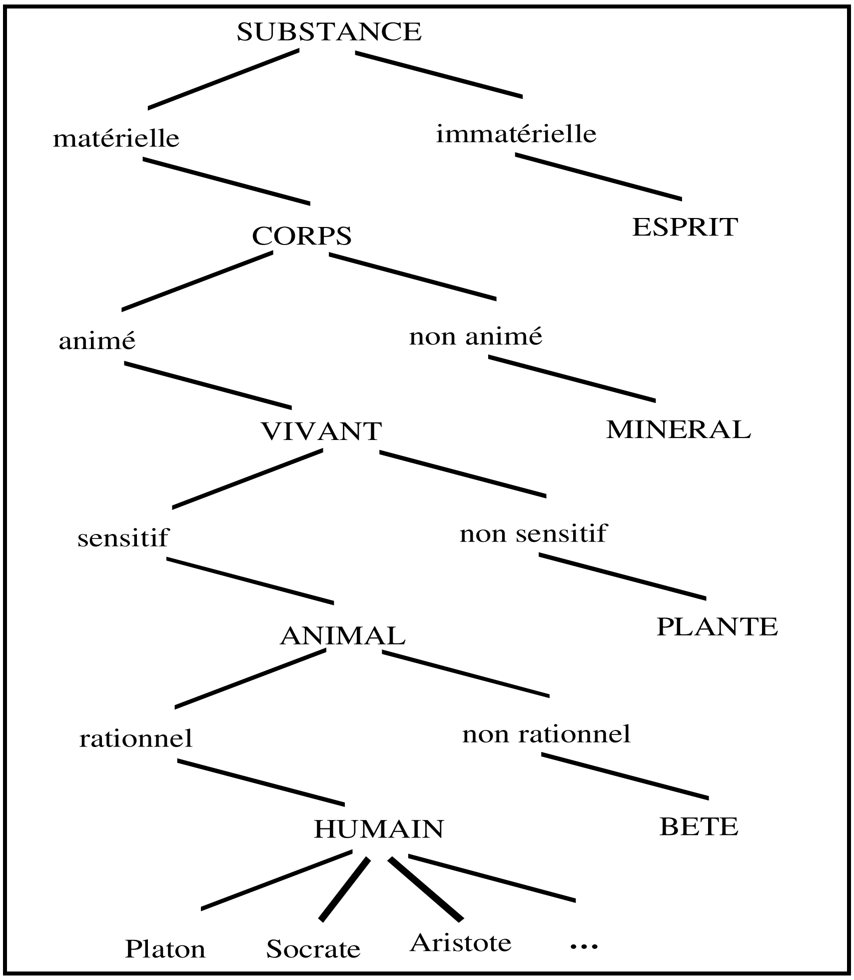


FIG 3.7 Arbre de Porphyre

Suite aux travaux de Quillian (1968), de nombreux types de réseaux sémantiques ont été développés dans lesquels la diversité du sens attribué aux nœuds et aux liens a été l'objet d'un certain nombres de questions (Woods, 1975 ; Brachman, 1977). Dans l'objectif de définir avec plus de précision les réseaux sémantiques, certains formalismes particulièrement complets et rigoureux ont alors été développés. Les plus connus sont KL-ONE (Brachman &Schmolze, 1985) et le formalisme des graphes conceptuels (Sowa, 1991). On s'est aussi attaché à définir une correspondance avec la logique du premier ordre. Tout graphe conceptuel pouvant par exemple s'écrire comme une formule de la logique du premier ordre (la réciproque est fausse) (Sowa, 1984).

Les domaines d'utilisation des réseaux sémantiques sont : la modélisation de l'organisation des connaissances en mémoire et de leur mode d'accès, le traitement du langage naturel et plus précisément la représentation de la connaissance transmise par un énoncé.

Il est à noter que ce système de représentation demeure tout de même peu utilisé en Intelligence Artificielle dans le cadre de la modélisation du raisonnement pour laquelle on lui préfère de simples systèmes de règles. Ceci est très certainement lié au manque de méthodes permettant d'utiliser ce type de représentation des connaissances comme le fait remarquer Bill Woods « A major gap, I believe, is the lack of sufficient emphasis on algorithmic uses of network representations to support various kinds of inference » (extrait de Sowa, 1991, p. 21).

3.2.2 Méthodes pour la construction

Certains systèmes sont en mesure de construire automatiquement la représentation d'un énoncé sous la forme d'un réseau sémantique. Compte tenu de leur aspect naturel, les réseaux sémantiques sont aussi souvent construits manuellement.

Nous nous intéressons ici plus particulièrement aux méthodes proposées dans le cadre d'un apprentissage inductif, c'est-à-dire aux méthodes de construction basées sur la donnée d'un ensemble d'observations.

On notera d'abord qu'il est possible de considérer un ensemble de règles comme un réseau, chaque prémisse et conclusion apparaissant comme les nœuds du réseau et chaque implication apparaissant comme un lien entre ces nœuds. De plus, parce que la conjonction ou la disjonction de propriétés peut être considérée comme la définition intensionnelle d'un concept, on se ramène à la notion de réseau sémantique, même si certains préfèrent parler de réseau associatif compte tenu de la spécificité de ce type de réseau. De ce point de vue, les systèmes d'induction de règles d'associations peuvent être considérés comme des méthodes de construction de réseaux sémantiques.

D'autres méthodes portent sur la recherche d'une structure de nature hiérarchique (voir Bournaud, 1996, pour un exposé plus complet). Si l'on reprend la structure sous-jacente à l'induction exposée lors de la description de CHARADE, la correspondance de Galois entre le treillis des exemples et celui des propriétés, il est possible de générer une hiérarchie de concepts.

On reprend ici l'exemple donné par Poitrenaud (1998), où E est l'ensemble des exemples (E=(1,2,3)), E' est l'ensemble des propriétés (E'=(A,B)), f est la fonction qui lie les parties de E aux parties de E' et f' est la fonction qui lie les parties de E' aux parties de E.

La méthode consiste à considérer les parties de E (P(E)) fermées par h=fof' et celles de E' (P(E')) fermées par h'=f'of, soit (1), (1,2), (1,3), (1,2,3) pour P(E) et (Ø), (A), (B), (A,B) pour P(E') (figure 3.8).

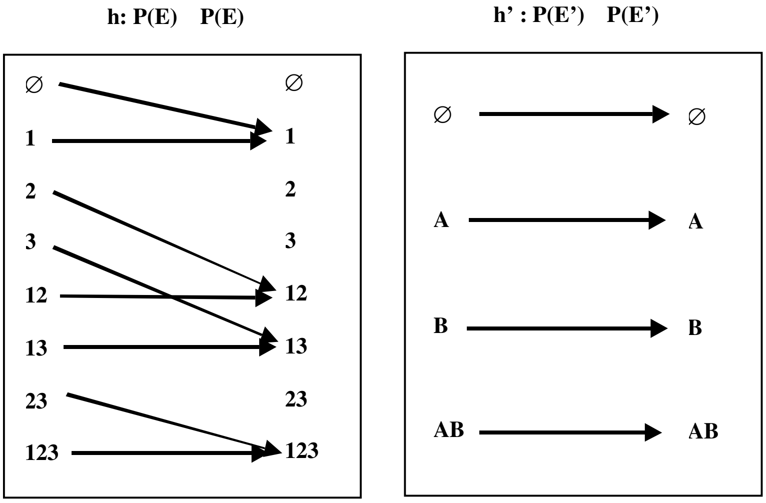


FIG. 3.8 Représentation des fonctions h et h' pour l'exemple.

En terme de correspondance entre treillis, on obtient alors la configuration suivante et le treillis de Galois qui en découle:

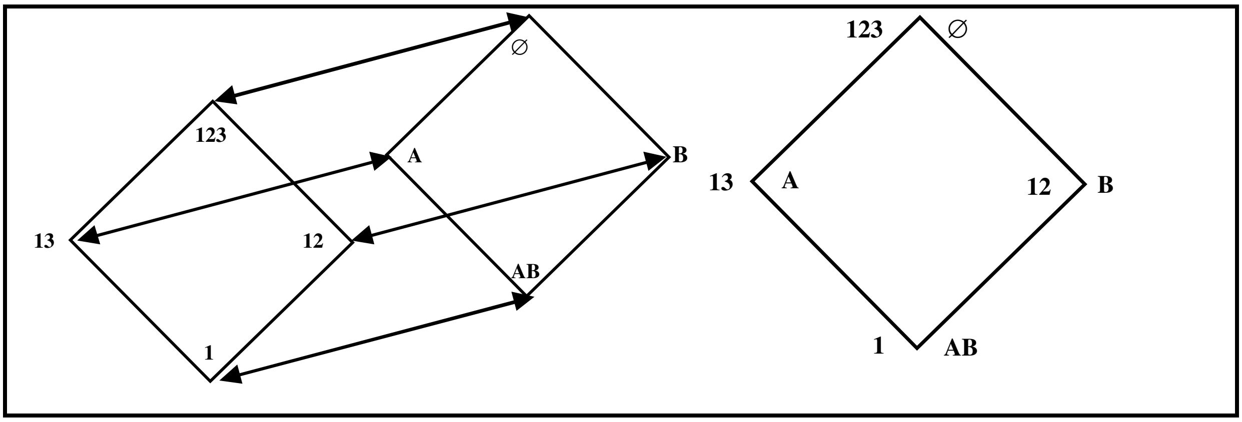


FIG. 3.9 Correspondance des parties fermées de l'ensemble des exemples et de celui des propriétés et treillis de Galois associé.

La structure ainsi construite peut être considérée comme une hiérarchie de catégories définies à la fois en intension (conjonction de propriétés) et en extension (ensemble d'exemples). Une catégorie correspond donc en termes de propriétés, à un ensemble maximal de propriétés partagées par des exemples c'est-à-dire tel qu'il n'existe pas un ensemble de propriétés plus grand qui s'applique aux mêmes exemples. Une catégorie correspond en termes d'exemples, à un ensemble maximal d'exemples qui partagent un ensemble de propriétés c'est-à-dire tel qu'il n'existe pas un ensemble d'exemples plus grand partageant ces propriétés. On notera que ce type de structure est à la base du système PROCOPE (Poitrenaud, 1995) qui simule la construction psychologique des catégories (Richard & al., 1993 ; Tijus & Moulin, 1997).

On notera que les réseaux d'implications et les réseaux hiérarchiques sont construits à partir d'un même ensemble d'exemples. Ils représentent tous deux la structure de propriétés et sont équivalents. Il est possible de lire les implications dans le réseau hiérarchique et il est aussi possible de lire des liens hiérarchiques à partir du réseau d'implications. Si d→d', cela signifie que tous les exemples vérifiant d vérifient d', d' est donc plus générale que d.

Il existe différents algorithmes permettant de construire automatiquement le treillis de Galois (Guenoche, 1990). Certains permettent de construire cette hiérarchie de façon incrémentale (Gaudin & al., 1995), d'autres s'appliquent à des exemples structurés (Mineau, 1990 ; Girard & Ralambondrainy, 1996). Lorsque le nombre d'observations est important et la structure des propriétés complexe, le treillis devient vite illisible compte tenu de la multiplicité des liens et des nœuds. Bernard et Poitrenaud (1999) ont développé une méthode de simplification du treillis basée sur la notion de quasi-implication revenant à ne pas considérer certaines conjonctions de propriétés sous-représentées.

D'autres algorithmes de construction de hiérarchies de concepts plus heuristiques ont été définis. Parmi les plus connus, on trouve le système COBWEB (Fisher, 1987). Dans ce système, la représentation d'une catégorie consiste en la donnée de l'ensemble des probabilités des différentes propriétés. Par exemple la catégorie des oiseaux sera représentée par P(plume) = 1, P(vole)=0.9, P(nage)=0.1, etc. L'algorithme procède de façon incrémentale. La prise en compte de chaque nouvelle observation consiste à modifier la hiérarchie construite à partir des opérateurs suivants :

Placer : Cet opérateur place l'observation dans une catégorie et met à jour les probabilités associées aux différentes catégories mais qui ne modifie pas la structure.

Créer : Cet opérateur crée une nouvelle catégorie et y place la nouvelle observation.

Fusionner : Cet opérateur crée une nouvelle catégorie qui regroupe deux sous-catégories.

Diviser : Cet opérateur détruit une catégorie qui regroupait deux sous-catégories.

L’algorithme consiste pour chaque nouvelle observation à parcourir la hiérarchie en partant de la racine et à appliquer à chaque niveau l’opérateur le plus adapté jusqu’à atteindre une feuille dans laquelle la nouvelle observation est placée. Le choix de l’opérateur se fait sur le critère de la « Category Utility » (CU) (Gluck & Corter, 1985), qui est une mesure originellement développée dans le souci de rendre compte des catégories les plus naturellement utilisées d’un point de vue psychologique, c’est-à-dire les catégories du niveau de base décrites par Mervis et Rosch (1981). C’est une mesure globale qui s’applique à une hiérarchie et qui rend compte en quelque sorte de sa qualité.

Elle combine deux critères : l’un indique à quel point chaque catégorie permet de prédire les attributs des objets y appartenant. L’autre indique à quel point la valeur des attributs des objets permet de prédire leur catégorie. On notera une nouvelle fois que cette mesure de qualité renvoie à une double implication entre l’objet à qualifier (la catégorie) et la situation (ici, l’ensemble des observations). Ceci peut être reformulé comme une optimisation maximisant à la fois la similarité entre les observations d’une même catégorie et la différence entre les observations de catégories différentes.

$$CU = \frac{\sum_{k=1}^{n} P(C_k)\left[\sum_i \sum_j P(A_i = V_{ij} / C_k)^2 - \sum_i \sum_j P(A_i = V_{ij})^2\right]}{n}$$

On peut aussi citer d’autres systèmes comme CLUSTER-2 (Michalski & Stepp, 1983) basé sur le principe des nuées dynamiques (Diday, 1971) qui consiste à considérer k observations initiales qu’on appelle « noyaux » à partir desquelles on constitue k regroupements en considérant pour toutes les autres observations le noyau le plus proche et en réitérant la même procédure sur les k nouveaux regroupements ou encore AUTOCLASS (Cheeseman & al., 1990) qui considère une représentation des catégories semblable à COBWEB et qui repose sur une méthodologie bayésienne.

3.2.3 Méthodes pour l'exploitation

3.2.3.1 Diverses méthodes

Les inférences dans les réseaux sémantiques peuvent prendre différentes formes. Elles peuvent correspondre à des opérations sur les graphes comme des projections ou des jointures ou encore des simplifications. Chevallet (1992) propose par exemple de représenter la spécificité d’un document D vis-à-vis d’une requête Q dans le formalisme des graphes conceptuels par une projection partielle du graphe conceptuel représentant le contenu de Q sur le graphe représentant le contenu de D.

Elles peuvent correspondre à des procédures adaptées à des problèmes particuliers permettant par exemple d'inférer les propriétés d'un objet en considérant des mécanismes d'héritage, ou encore de retrouver un objet ayant une relation précise avec l'objet courant en suivant des relations particulières dans le réseau.

Elles peuvent aussi être moins abstraites et plus locales, lorsqu'elles proviennent d'une analogie avec le fonctionnement neuronal. Ainsi les méthodes de propagation d'activation ou encore de propagation de marqueurs s'inspirent de données neurophysiologiques.

Concernant les méthodes de propagation de marqueurs, on peut citer les travaux de Fahlman (1979) et son système NETL qui utilise la propagation de marqueurs et dont le principe consiste dans un premier temps à placer un marqueur sur un ou plusieurs nœud(s) du graphe, à propager ce ou ces marqueurs dans le réseau selon des règles qui dépendent de ce que l'on cherche à inférer et enfin lorsque la propagation est terminée, à observer les nœuds marqués pour en tirer des conclusions. Une application connue concerne le traitement des règles à exception avec l'utilisation de liens inhibiteurs. Enfin, on peut également citer les travaux de Shastri (Kayser, 1997, p. 150) portant sur la propagation de signaux périodiques.

3.2.3.2 Les méthodes basées sur la propagation d'activation

Concernant les méthodes de propagation d'activation, la méthode proposée par Collins et Loftus (1975), qui reprennent le travail de Quillian (1968), considère pour expliquer différents phénomènes psychologiques (temps d'accès, amorcage, voir chapitre 2) les principes suivants (que nous présentons ici sous une forme résumée):

i) L'intensité décroît sur le chemin de la propagation,

ii) Aussi longtemps qu'un concept est activé, il propage une activité à intervalles réguliers,

iii) L'activité d'un concept décroît suivant le temps,

iv) Lorsque deux activations se propagent en un même nœud, ses deux activations s'ajoutent. Et ce n'est que lorsque cette somme dépasse un certain seuil, qu'elle est prise en compte.

Ces méthodes de propagation ont donné assez peu d'applications en Intelligence Artificielle. Le domaine où elles sont le plus représentées est sans doute le domaine de la Recherche d'Information (Crestani, 1997). On comprend aisément pourquoi puisque la problématique de la Recherche d'Information est l'organisation et l'accès aux informations, c'est-à-dire précisément ce sur quoi portent les travaux relatifs à la mémoire sémantique qui sont à l'origine de ce type de méthodes.

Cohen et Kjeldsen (1987), avec le système GRANT, obtient par propagation d'activation de meilleurs résultats que les systèmes classiques de recherche par mots clés. Ce système est chargé de trouver un organisme de financement pour un sujet de recherche donné. La méthode suivie consiste dans un premier temps à représenter dans un réseau sémantique les connaissances portant sur les sujets de recherche et les agences de financements, puis dans un second temps, à activer les connaissances relatives au sujet de recherche pour lequel doit être trouvé un financement et à laisser cette activation se propager dans le réseau vers les nœuds représentant les différentes agences de financement. Certaines limitations ont dû être ajoutées pour que ce système donne de bons résultats. La distance de propagation est limitée (deux arcs semble la distance optimale), les nœuds à forte connectivité sont évités (ces nœuds correspondent à des concepts généraux comme "vie" qui sont reliés à de très nombreux autres nœuds) et enfin il existe des chemins privilégiés, d'autres au contraire pénalisés (figure 3.10).

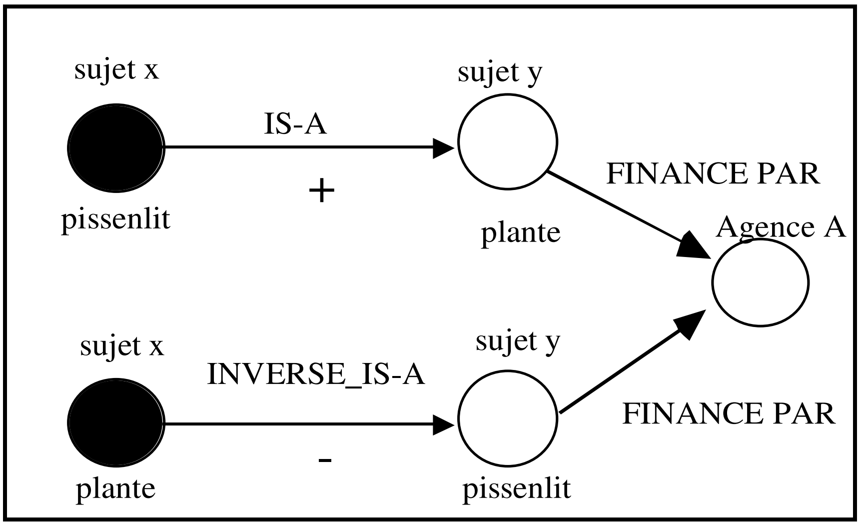


FIG. 3.10 : Illustration d'un chemin privilégié (+) et d'un chemin pénalisé (-). Si l'étude porte sur les pissenlits et que l'agence finance des sujets portant sur les plantes, il est possible que l'agence puisse convenir. Dans le cas contraire, c'est moins probable.

Dans sa synthèse sur les méthodes de propagation pour la Recherche d'Information, Crestani (1997) distingue les méthodes de propagation « pures » et des méthodes de propagation plus contrôlées. Les méthodes de propagation « pures », consistant à laisser l'activation se propager sans introduire de contrainte, posent selon cet auteur les problèmes suivants:

- La propagation de l'activation à l'ensemble du réseau,
- Le fait que la sémantique attachée aux différents liens soit inutilisée,
- La difficulté à implémenter des inférences basées sur la sémantique des liens.

Les contraintes communément ajoutées pour contourner ces problèmes sont :

- Les contraintes de distance

Il s'agit de faire décroître l'activation lorsque l'on s'éloigne du nœud source d'activation comme le proposent Collins et Loftus (1975). On peut aussi de façon plus drastique considérer une distance maximale de propagation comme dans le système GRANT.

- Les contraintes de connectivité

Certains nœuds ont une connectivité particulièrement forte. Il s'agit de nœuds représentant des concepts très généraux et par conséquent liés à de très nombreux autres concepts. Ces nœuds sont en partie la source du problème de la propagation à l'ensemble du réseau. Une solution consiste à ne pas propager l'activation aux nœuds à trop forte connectivité.

- Les contraintes de chemin

Suivant l'application, le type d'inférence souhaité peut être différent. Ces préférences peuvent être représentées par une pondération des liens suivant leur types ; ce qui revient à l'extrême à ne considérer que certains types de liens.

- Les contraintes d'activation

Un nœud est activé si la somme de l'activation provenant de ses entrées dépasse un certain seuil. Une solution pour éviter la propagation à l'ensemble du réseau consiste parfois à augmenter ce seuil.

Bien que le système GRANT qui utilise ce type de contraintes obtienne de bons résultats, on peut cependant considérer que l'intérêt de ce type de méthode réside précisément dans leur aspect automatique et local, et que le fait d'introduire des règles pour contrôler la propagation rend la méthode un peu artificielle puisque l'on se ramène finalement à un système de règles.

Si le système GRANT propage l'activation jusqu'à l'information directement utile pour l'utilisateur, des méthodes de propagation d'activation ont aussi été utilisées pour étendre les requêtes (Chen & Ng, 1995). Ces dernières considèrent un nombre de termes plus important que ceux de la requête initiale, ce qui permet ensuite de sélectionner des documents qui contiennent les termes qui n'étaient pas initialement présents dans la requête. Ces méthodes ont aussi été utilisées pour la recherche d'informations à partir d'hypertextes (Savoy, 1992 ; Lucarella, 1990) où la notion de réseau apparaît immédiatement.

3.2.4 Les réseaux sémantiques flous

Un concept est souvent décrit par une conjonction de propriétés. Certaines de ces propriétés sont vérifiées de façon binaire (avoir ou ne pas avoir d'ailes par exemple), d'autres propriétés que nous utilisons quotidiennement ne sont au contraire pas clairement délimitées (être grand, avoir les cheveux long, être jeune). La théorie des ensembles flous (Zadeh, 1965 ; Bouchon-Meunier, 1995), qui généralise la théorie des ensembles en autorisant une appartenance graduelle des éléments aux ensembles (exprimée par une valeur appartenant à l'intervalle [0,1]), a pour objectif de représenter ce type de catégories.

Considérant des catégories floues, on peut donc être amené, si l'on considère les relations entre ces catégories, à utiliser le terme de « réseau sémantique flou ».

3.2.4.1 Des hiérarchies de concepts intégrant la notion d'imprécision

Les travaux permettant d'intégrer de l'imprécision dans la construction de réseau de concept sont assez rares et relèvent d'approches très différentes.

L'approche de Rossazza (1992) basée sur la théorie des possibilités part de la définition d'une classe par un ensemble d'attributs et de l'association pour chaque attribut d'une zone possible et d'une zone typique. La zone possible, notée Z(a,C) (zone possible de l'attribut a pour la classe C), est un ensemble flou représentant les valeurs plus ou moins permises sur l'attribut donné pour la classe donnée. La zone typique, notée T(a,C) (zone typique de l'attribut a pour la classe C), est un ensemble flou représentant les valeurs plus ou moins usuelles des éléments de la classe. Par ailleurs, on impose que le noyau de Z(a,C) soit inclus dans le support de T(a,C) de façon à ce qu'une valeur typique soit au moins possible avec un degré 1. La classe des mammifères peut être ainsi décrite sur l'attribut locomotion en considérant Z(locomotion,Mammifère) = {0.2/vole, 1/marche, 0.3/saute, 0.5/nage} et T(locomotion,mammifère) = {1/marche}. Pour les objets, l'auteur définit similairement une zone possible et une zone crédible pour chaque attribut.

Les classes définies manuellement par la donnée de leurs attributs en termes de zones possibles et typiques peuvent ensuite être agencées dans une hiérarchie à partir d'un calcul d'inclusion pour tout couple de classe. Le calcul d'inclusion d'une classe C2 dans une classe C1 repose sur l'agrégation des degrés d'inclusion des sous-ensembles flous définissant les zones des différents attributs. Quatre degrés d'inclusion peuvent être calculés suivant la zone considérée pour chacune des classes. Si l'on considère les zones possibles pour les deux classes et si l'on note NG la mesure d'inclusion entre deux sous-ensembles flous, le degré d'inclusion est donné par :

$$\min\left\{N_G(Z(a,C_1),Z(a,C_2)) \mid a \text{ est un attribut de } C_1\right\}$$

A partir des différentes inclusions, différents types de raisonnement par héritage peuvent être définis. Rossazza (1992) distingue l'héritage normal, typique et atypique. Bien que cette méthode intègre l'imprécision et permette de calculer des degrés d'inclusion, on notera qu'elle ne permet pas de construire automatiquement la hiérarchie de classes à partir des observations puisque les classes doivent être préalablement données. Omri (1994) reprend ce type de réseau sémantique flou pour représenter les connaissances relatives à l'utilisation de dispositifs techniques.

Itzkovich et Hawkes (1994) ont introduit la notion de réseau de concepts flous « fuzzy concept network ». Ils étendent ainsi la notion de hiérarchie classique en considérant d'une part des degrés d'inclusion et en considérant d'autre part, une relation de similarité (elle aussi pondérée). Ainsi, ils définissent deux relations floues définies sur un ensemble de concepts C, $C=\{c_1,c_2,\ldots,c_n\}$ qui à tout couple de concept c_i, c_j font correspondre une valeur de l'intervalle [0,1]. La relation de similarité, notée R_{sim}, est réflexive ($\mu_{sim}(c_i,c_i)=1$), symétrique ($\mu_{sim}(c_i,c_j)=\mu_{sim}(c_j,c_i)$ et transitive ($\mu_{sim}(c_i,c_k)\geq\vee_{cj}(\mu_{sim}(c_i,c_j)\wedge\mu_{sim}(c_j,c_k)$)). La relation de généralisation R_g est réflexive ($\mu_g(c_i,c_i)=1$), antisymétrique (si $\mu_g(c_i,c_j)>0$ et si $\mu_g(c_j,c_i)>0$ alors $c_i=c_j$) et transitive ($\mu_g(c_i,c_k)\geq\vee_{cj}(\mu_g(c_i,c_j)\wedge \mu_g(c_j,c_k)$)). Un réseau de concepts flou est défini par l'ensemble des concepts et les deux relations (figure 3.11).

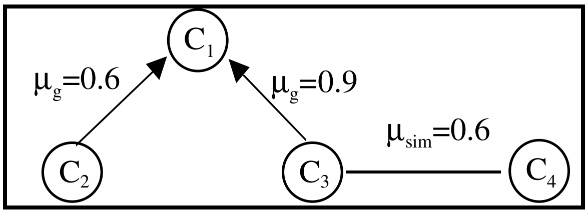


FIG. 3.11 Un exemple de réseau de concepts flou

Les auteurs proposent également un algorithme qui consiste à considérer un seuil pour les relations de similarité et de généralisation se ramenant ainsi à des relations binaires. Cet algorithme regroupe ainsi certains concepts dont la similarité est égal à 1 pour former des classes de concepts et organise des hiérarchies à l'intérieur de ces groupes. Ce type de réseau a été généralisé par Horng et Chen (1999) pour des relations définies par des nombres flous (et non plus simplement des réels) de forme trapézoïdale. On notera que la construction proposée, d'une part, ne correspond qu'à un réarrangement du réseau pour lequel sont déjà définies des relations de généralisation et des relations de similarité et, d'autre part, supprime tout aspect imprécis en considérant des seuils. De plus, on notera que l'imprécision dans ces réseaux ne concerne pas les concepts, mais les relations entre ces concepts.

Concernant la construction automatique de hiérarchies rendant compte de l'imprécision, la solution introduite par Girard et Ralambondrainy (1996) est très satisfaisante puisque non seulement elle permet d'introduire une imprécision mais que de plus, elle permet sa construction automatique à partir d'objets éventuellement structurés et mal définis. La solution adoptée consiste à définir un treillis de valeurs qualifiant les attributs (figure 3.12).

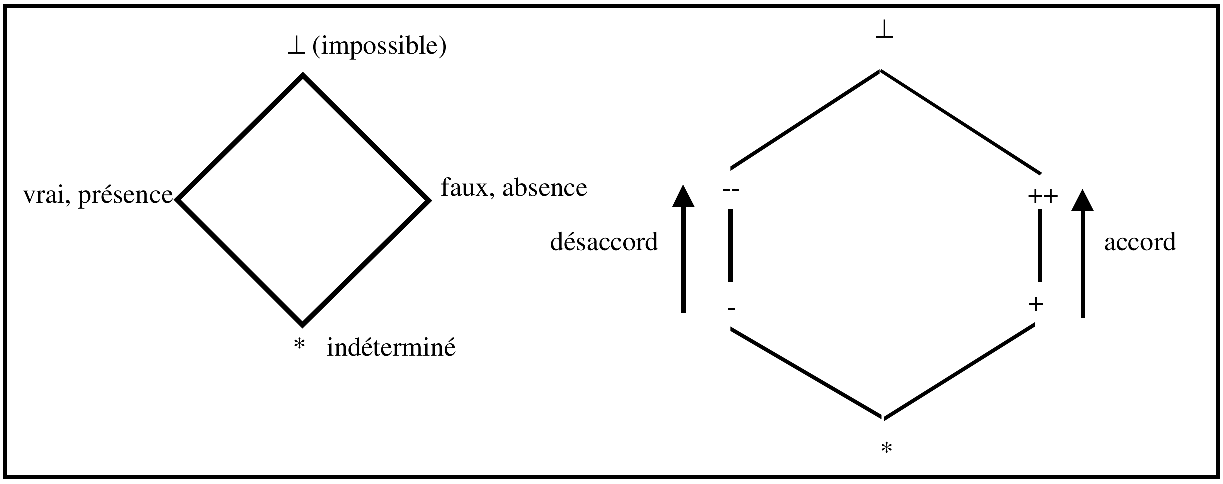


FIG 3.12 Deux exemples de treillis de valeurs qualifiant les attributs

Les observations à partir desquelles la hiérarchie est construite correspondent donc à une liste d'attributs valués comme par exemple ((a,*)(b,+)(c,--)). A partir de l'ensemble des observations, il est possible de construire le treillis de Galois puisqu'on dispose d'une relation binaire entre un ensemble d'exemples et un ensemble de propriétés structurées et valuées qui ont tous deux une structure de treillis. Soit par exemple pour deux observations simples e_1, e_2 le treillis de Galois suivant :

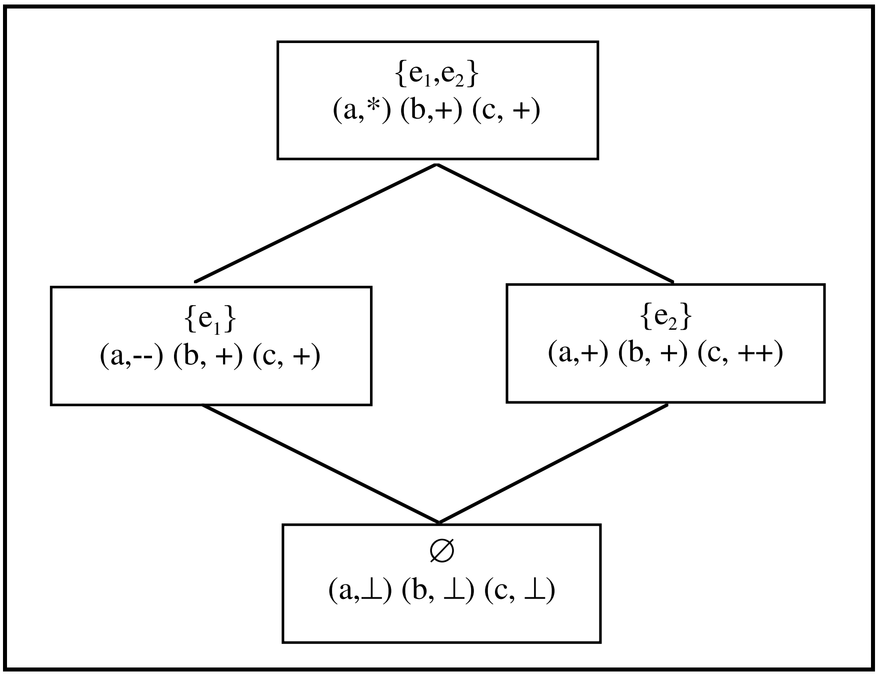


FIG. 3.13 Exemple de treillis de Galois construit pour deux observations décrites par des propriétés valuées.

3.2.4.2 Les thésaurus flous en Recherche d'Information

Nous abordons ici des réseaux utilisés dans le cadre de la Recherche d'Information qui permettent de représenter des relations de nature associative entre les différents termes. Les thésaurus permettent par exemple d'étendre une requête à des termes synonymes. On peut soit considérer des relations binaires entre les termes, soit envisager des relations floues comme Radecki (1976) l'a proposé. Dans ce dernier cas, on parle de thésaurus flous. Lucarella et Morara (1991) présentent par un exemple, un système de recherche d'information basé sur des réseaux de concepts flous. Ils considèrent un réseau de concepts dans lequel les liens sont orientés et valués et dans lequel figurent aussi des nœuds représentant les documents (figure 3.14). La valuation des liens représente le degré avec lequel le concept ou document pointé est pertinent vis-à-vis du nœud origine.

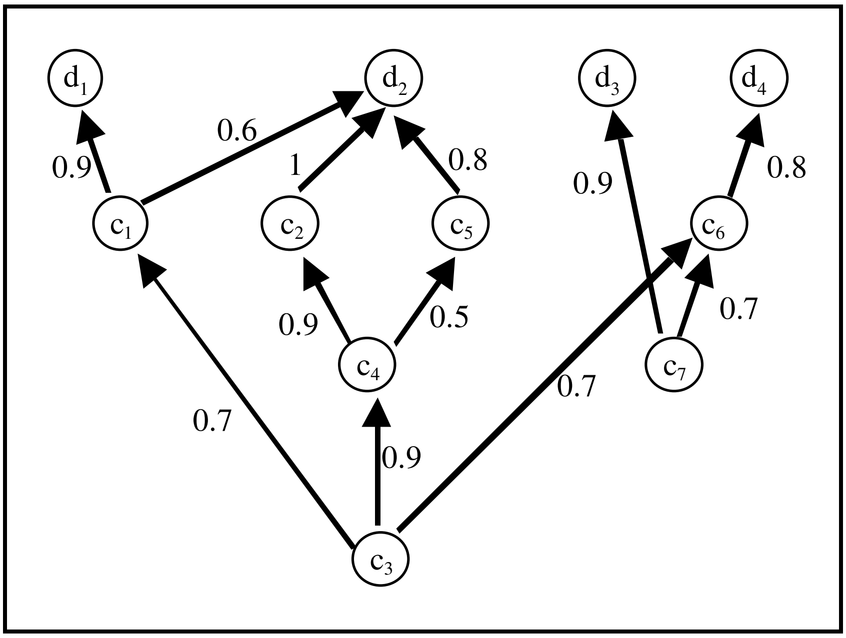


FIG. 3.14 Un exemple de réseau de concepts « flou ». Les nœuds notés d_i représentent des documents.

Dans ce réseau, un document est donc décrit comme un sous-ensemble flou de concepts. Par exemple, d'après la figure 3.14, $d_2 = \{(c_1,0.6), (c_2,1), (c_5,0.8)\}$.

Si la requête porte sur le concept c_3 par exemple, le nœud représentant le document d_2 peut être atteint selon trois chemins $c_3 \rightarrow c_1 \rightarrow d_2$, $c_3 \rightarrow c_4 \rightarrow c_2 \rightarrow d_2$ et $c_3 \rightarrow c_4 \rightarrow c_5 \rightarrow d_2$. En considérant une agrégation des valuations des liens d'un même chemin par le minimum, on obtient respectivement pour les trois chemins : 0.6, 0.9, 0.5. La pertinence du document d_2 vis-à-vis du concept c_3 est définie par max(0.6,0.9,0.5) = 0.9.

Les concepts sont généralement des mots (les mots-clefs choisis pour la représentation des documents) et la construction automatique du thésaurus consiste le plus souvent à mesurer la co-occurence des mots dans les différents documents. Ainsi lorsque deux mots apparaissent dans les mêmes documents, ils sont parfaitement liés. Miyamoto (1990) par exemple, considère deux types de relation. Partant de la fonction h qui associe à chaque concept w le sous-ensemble flou des documents dans lequel il est plus ou moins présent, il définit :

$$R(v,w) = \frac{|h(v) \cap h(w)|}{|h(v) \cup h(w)|}$$

$$N(v,w) = \frac{|h(v) \cap h(w)|}{|h(v)|}$$

où $|A|$ désigne la cardinalité de l'ensemble flou A qui s'écrit aussi $\sum count(A) = \sum_{i=1}^{n} \mu_A(o_i)$ où $\mu_A(o_i)$ correspond au degré d'appartenance de l'observation o_i à l'ensemble flou A. $R(v,w)$ est une relation qui mesure à quel point, lorsque v ou w est présent, on peut s'attendre à ce que l'autre concept soit aussi présent. $N(v,w)$ mesure à quel point, lorsque v est présent, on peut s'attendre à ce que w le soit aussi.

Le plus souvent, il s'agit donc d'association positive. On notera cependant que Chen et Horng (1999) ont proposé d'étendre le concept de thésaurus flous en proposant d'autres types de liens. En plus des associations positives, ils considèrent en effet des associations négatives, des liens de généralisation et des liens de spécialisation. Ces différents types de relation peuvent être composés et utilisés pour la formulation de requête et la recherche de documents pertinents.

Nous proposons dans le chapitre suivant un nouveau type de réseau sémantique flou, ainsi qu'une méthode de construction et d'exploitation associée.

Chapitre 4

Relief : Ce qui fait partie saillante d'un ensemble. Ensemble des saillies et des creux d'une surface. Fig. Apparence plus vive d'une chose.
Dictionnaire Encyclopédique Larousse, édition 1983

4 PRESENTATION DU SYSTEME RELIEFS

Nous avons proposé dans le second chapitre, un ensemble de principes pour l'extraction d'informations pertinentes. Dans le chapitre suivant, nous avons décrit un ensemble de systèmes respectant ces principes en insistant plus particulièrement sur les réseaux sémantiques. Nous présentons maintenant le système RELIEFS (pour RELevant Information Extraction Fuzzy System) permettant la construction et l'exploitation de réseaux sémantiques flous.

Considérant que les régularités présentes dans les données ne sont pas toujours triviales dans le sens où elles peuvent faire intervenir des propriétés aux limites imprécises et qu'elles peuvent être soumises à exception comme dans "les grands appartements ont en général un loyer élevé", RELIEFS extrait des règles valuées (c'est-à-dire auxquelles sont attachés des coefficients de vraisemblance) dans lesquelles apparaissent en prémisse ou en conclusion des conjonctions de propriétés floues. Ces règles correspondront aux arcs du réseau.

La valuation des règles fait l'objet de la première partie de la présentation du système. Dans la seconde partie, nous traitons de la construction globale du réseau, notre méthode proposant des solutions au problème de l'explosion combinatoire liée à la recherche de conjonctions de propriétés. Dans une troisième partie, nous proposons une méthode de construction automatique des sous-ensembles flous définissant les propriétés qui renvoie au problème de la discrétisation. Nous traitons ensuite de l’exploitation de tels réseaux pour l’extraction d’information pertinente. Ces quatre parties commencent par une introduction du problème posé comprenant un bref état de l’art, donnent la méthode proposée et l’illustrent sur un exemple jouet.

Nous terminons la présentation de RELIEFS par une discussion générale qui inclut un parallèle de la méthode avec le système cognitif.

4.1 Evaluer la règle « SI A ALORS B » à partir d’un ensemble d’observations

L’un des principes définis à l’issue du second chapitre concernait la capacité du système à extraire des dépendances entre les différentes informations, puisque la pertinence renvoyait à l’évaluation d’une dépendance. Nous avons dans le chapitre suivant, mis en avant différents systèmes en mesure d’extraire ces dépendances. Nous proposons une nouvelle méthode qui a la particularité d’être incrémentale et qui permet de traiter des données incomplètes et/ou imprécises.

4.1.1 Introduction

4.1.1.1 Etat de l’art

Ce problème n'est pas nouveau puisque déjà, dans le système MYCIN (Shortliffe, 1976), les différentes règles étaient pondérées par l'expert pour rendre compte de leur vraisemblance. Plus récemment, des systèmes construisant et pondérant automatiquement des règles comme CHARADE (Ganascia, 1987) ou encore des algorithmes rapides de recherche d'associations comme APRIORI (Agrawal & al., 1995) ont utilisé différents critères d'évaluation de l'implication.

Par ailleurs, différents indices statistiques mesurant la dépendance entre variables (Goodman & Kruskal, 1959) ont été proposés. Ces indices sont calculés à partir d’une table de contingence dans laquelle, si l’on ne considère par exemple que 2 variables, chaque case indique le nombre d’observations n_{ij} ayant la $i^{ème}$ modalité de la première variable et la $j^{ème}$ de la seconde. Le plus usuel est l’indice du χ^2 basé sur une différence du nombre d’observations réelles avec l’effectif théorique considérant une indépendance entre les deux variables. Cet effectif théorique correspond à $\tilde{n}_{ij} = \frac{n_i n_j}{n}$ où n_i et n_j représentent respectivement le nombre d’observations ayant la $i^{ème}$ modalité de la première variable et le nombre d’observations ayant la $j^{ème}$ modalité de la seconde variable et n représente le nombre total d’observations. L’indice s’écrit : $\chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{(n_{ij}-\tilde{n}_{ij})}{\tilde{n}_{ij}}$ où p et q sont respectivement le nombre de modalités des deux variables.

Lorsque ces indices sont locaux (c'est-à-dire qu'ils portent sur la dépendance entre deux modalités et non sur la dépendance entre deux variables) et orientés, il est possible de parler d'implication (Bernard, 1996 ; Gras & Lahrer, 1993).

Par exemple, L’indice de Loevinger (1948) sur lequel repose l’indice proposé par Bernard (1996), est un indice simple qui peut être utilisé dans le cas où l’on dispose de deux variables binaires, c’est-à-dire pour lesquelles on considère seulement deux modalités exclusives et complémentaires (blanc/noir, riche/pauvre). Ainsi, si l’on considère pour deux variables A et B les modalités respectives a et a’ (a’=¬a), b et b’(b’=¬b), la mesure de l’implication « a→b » est définie par l'indice $H_{ab} = 1 - \frac{f_{ab'}}{f_a f_b}$ qui varie dans l'intervalle $]-\infty,1]$ et qui vaut 1 lorsque les évènements a et b' ne sont jamais survenus ensemble. Cet indice mesure la petitesse de la fréquence de l’événement « a et b’ » relativement aux fréquences marginales de a et de b. On notera qu’il s’appuie donc sur le seul cas pour lequel l’implication matérielle est fausse, considérant donc les contre-exemples.

En logique floue, différentes implications ont été proposées, généralisant (pour la plupart) l'implication matérielle sur l'intervalle [0,1], c’est-à-dire attribuant une valeur à la relation d’implication lorsque les valeurs de vérité de la prémisse ($f_A(x)$) et de la conclusion ($f_B(y)$) est un degré compris dans l’intervalle [0,1]. On peut citer par exemple, l’implication de Lukasiewicz min($1-f_A(x)+f_B(y)$,1) ou l’implication de Rescher-Gaines qui vaut 1 si $f_B(x) \geq f_A(x)$ et 0 sinon. Leur utilisation dans le cadre de l'induction et leurs différentes sémantiques ont été étudiées par Bandler et Kohout (1980). Il faut aussi signaler les travaux portant sur les quantificateurs flous (Zadeh, 1983 ; Dubois & Prade, 1988 ; Kacpryzk & Iwanski, 1992) où l'on cherche à évaluer des énoncés du type "la plupart des A sont B". Le quantificateur « La plupart » peut être représenté par un sous-ensemble flou défini sur l’ensemble des proportions (figure 4.1).

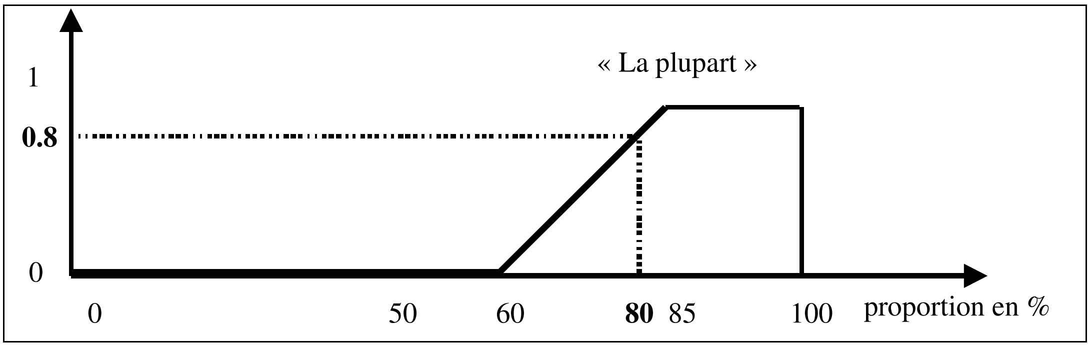


FIG 4.1 « la plupart » peut se représenter par un sous-ensemble flou sur l'ensemble des proportions.

Ainsi pour évaluer la proposition « la plupart des observations sont A» à partir d'un ensemble d'observations O={o_1,o_2,…,o_n} où A représente un ensemble éventuellement flou, on calcule la cardinalité de A notée $\sum Count(A)$ définie par $\sum Count(A) = \sum_{i=1}^{n} \mu_A(o_i)$ et on rapporte ensuite cette cardinalité au nombre d'observations, pour obtenir la proportion $r = \frac{1}{n} \sum_{i=1}^{n} \mu_A(o_i)$. Si l'on note μ_Q la fonction d'appartenance définissant le quantificateur Q, $\mu_{Q(r)}$ correspond alors à l'évaluation de l'énoncé (dans l'exemple, 80% des observations sont A et « la plupart des observations sont A » est vrai avec un degré 0.8). On peut évaluer de la même façon des énoncés du type « la plupart des A sont B » en calculant la cardinalité relative $r = \frac{\sum Count(AetB)}{\sum Count(A)} = \frac{\sum_{i=1}^{n} (\mu_A(o_i) \wedge \mu_B(o_i))}{\sum_{i=1}^{n} \mu_A(o_i)}$ (Zadeh, 1983).

4.1.1.2 Trois critères d'évaluation

Globalement, on retrouve dans les mesures proposées, la prise en compte de trois aspects distincts. Un aspect descriptif (au sens statistique du terme) porte sur les fréquences observées (comme, par exemple, 90% des carrés sont bleus) et un aspect inductif (au sens statistique du terme) porte sur le nombre d'observations sur lequel s'appuie la règle (10 carrés ont permis d'évaluer la règle). La spécificité des recherches en logique floue concerne pour sa part un troisième aspect, la prise en compte de l'imprécision. On notera aussi une scission entre ces différentes mesures, certaines évaluant une fréquence conditionnelle, les autres une implication matérielle. Les premières prennent le parti de se fier à la présence simultanée des deux propriétés (principe d'association), les secondes se fient à la petitesse du nombre d'observations telles que la prémisse soit vraie et la conclusion fausse (la case « 0 » de l'implication) se basant ainsi sur les contre-exemples.

Nous plaçant dans le cadre de l'induction, nous cherchons donc à mesurer à quel point une propriété P prédit une propriété Q, c'est-à-dire à quel point il est possible d'affirmer que, lorsque la propriété P est présente, la propriété Q l'est aussi. Il ne s'agit donc pas d'une implication matérielle qui considèrera l'implication "P→Q" vraie lorsque P est faux. Ici lorsque P est faux nous ne considérons pas l'observation pour l'évaluation. La considérer comme validant l'implication, comme le propose la table de l'implication matérielle, conduirait à une mesure optimiste et la considérer comme invalidant l'implication nous conduirait au contraire à une mesure pessimiste. Cet aspect discutable de l'implication matérielle pour la représentation des règles est bien connu (Dubois & Prade, 1991) et nous rejoignons ici le propos des logiques de la pertinence et les logiques conditionnelles (voir chapitre 2). L'implication ne nous satisfaisant pas pour ces raisons, nous nous en tenons à l'évaluation d'une fréquence conditionnelle. Nous dénombrons en fait trois critères dans notre évaluation de "P→Q" :

1- la fréquence conditionnelle de Q sachant P.
2- l'imprécision des propriétés.
3- le nombre d'observations permettant d'évaluer "P→Q".

Le second critère qui concerne l'imprécision des propriétés est en effet une donnée importante à prendre en considération. Ramenons-nous, par exemple, à l'évaluation d'implications pour des formes géométriques simples présentes dans un dessin, soit par exemple : "grand→carré". Supposons que l'on observe un cercle d'une taille plutôt importante. L'évaluation de l'implication "grand→carré" sera alors d'autant plus affaiblie par notre observation que la taille du cercle est véritablement grande.

Enfin le troisième critère concerne l'aspect inductif de la mesure au sens statistique du terme par opposition aux deux premiers critères qui sont descriptifs. Plus le nombre d'observations sur lequel se base la règle est important, plus l'évaluation de la force de l'implication doit être élevée. Le nombre d'observations permettant d'évaluer la règle "P→Q" est le nombre d'observations vérifiant P que l'on appellera par la suite, *soutien* de la règle.

4.1.2 Méthode

4.1.2.1 Une règle d'apprentissage incrémentale et associative

Considérons un réseau dans lequel les nœuds sont associés à des conjonctions de propriétés et les connexions à des implications. L'apprentissage consiste à mettre à jour, pour chaque nouvelle observation, les implications entre conjonctions de propriétés. Nous posons donc le problème de l'évaluation des implications. Si l'on se contente des deux premiers critères, (1) et (2), il est possible de proposer une règle d'apprentissage qui, appliquée à chaque nouvelle observation, permet de réévaluer les différentes implications. Si l'on considère que $I_{PQ(k-1)}$ et I_{PQk} sont respectivement l'évaluation de "$P \rightarrow Q$" avant et après la k$^{\text{ème}}$ observation, que O_i est la i$^{\text{ème}}$ observation et que $\mu_P(O_i)$ est le degré (compris entre 0 et 1) avec lequel la i$^{\text{ème}}$ observation vérifie P, cette règle s'écrit :

$$I_{PQk} = \frac{\alpha I_{PQ(k-1)} + \mu_P(O_k)\mu_Q(O_k)}{\alpha + \mu_P(O_k)} \qquad \text{(R1)}$$

$$avec \;\; \alpha = \sum_{i=1}^{k-1} \mu_P(O_i)$$

Il s'agit simplement d'une moyenne pondérée entre $I_{PQ(k-1)}$ (avec la pondération correspondant au nombre des précédentes observations à partir desquelles $I_{PQ(k-1)}$ a été évaluée) et la contribution de la k$^{\text{ème}}$ observation, soit $\mu_Q(O_k)$ (qui est le degré avec lequel la k$^{\text{ème}}$ observation vérifie la propriété Q) pondérée par $\mu_P(O_k)$ (c'est-à-dire le degré avec lequel la k$^{\text{ème}}$ observation vérifie la propriété P). On ajoutera par ailleurs, que tant qu'aucune observation n'est telle que $\mu_P(O) \neq 0$, la connexion n'est pas construite et n'a donc pas besoin d'être évaluée. Enfin, on remarquera que si l'on avait voulu évaluer une implication matérielle, on aurait remplacé $\mu_Q(O_k)$ par une des différentes implications floues entre P et Q.

Le numérateur de la règle correspond précisément à la forme de base de la règle de Hebb couramment utilisée (Haykin, 1999). L'un de ses désavantages est de croître indéfiniment. Le dénominateur qui apparaît dans la règle et qui permet une normalisation de I_{PQ}, apporte une solution à ce problème.

On notera que la valeur de I_{PQ_n} correspond à la cardinalité relative, dans laquelle on a utilisé le produit comme t-norme, ce qui nous ramène à la probabilité conditionnelle d'événements flous (Zadeh, 1968).

$$I_{PQn} = \frac{\sum_{i=1}^{n} \mu_P(O_i)\mu_Q(O_i)}{\sum_{i=1}^{n} \mu_P(O_i)}$$

Ce résultat peut être démontré par récurrence :

Considérons la première observation o_k telle que $\mu_P(O_k) \neq 0$.

$\forall i < k, \mu_P(O_i) = 0$ donc $\alpha = 0$.

d'où en appliquant (R1) :

$$I_{PQk} = \frac{\mu_P(O_k)\mu_Q(O_k)}{\mu_P(O_k)} = \frac{\sum_{i=1}^{k} \mu_P(O_i)\mu_Q(O_i)}{\sum_{i=1}^{k} \mu_P(O_i)}$$

Par ailleurs, d'après (R1) on a : $I_{PQ(n+1)} = \dfrac{\alpha I_{PQn} + \mu_P(O_{(n+1)})\mu_Q(O_{(n+1)})}{\alpha + \mu_P(O_{(n+1)})}$

Si on considère l'hypothèse de récurrence $I_{PQn} = \dfrac{\sum_{i=1}^{n} \mu_P(O_i)\mu_Q(O_i)}{\sum_{i=1}^{n} \mu_P(O_i)}$,

on obtient :

$$I_{PQ(n+1)} = \frac{\sum_{i=1}^{n} \mu_P(O_i) \dfrac{\sum_{i=1}^{n} \mu_P(O_i)\mu_Q(O_i)}{\sum_{i=1}^{n} \mu_P(O_i)} + \mu_P(O_{(n+1)})\mu_Q(O_{(n+1)})}{\sum_{i=1}^{n} \mu_P(O_i) + \mu_P(O_{(n+1)})}$$

$$I_{PQ(n+1)} = \frac{\sum_{i=1}^{n+1} \mu_P(O_i)\mu_Q(O_i)}{\sum_{i=1}^{n+1} \mu_P(O_i)}$$

4.1.2.2 Prise en compte du nombre d'observations

Enfin, pour prendre en compte le troisième critère, plusieurs solutions sont possibles :

Une première solution consiste à introduire un terme d'oubli dans la règle locale. Ce terme dépend du degré de satisfaction de P (moins P est vérifiée, plus l'oubli est important) et de la précédente valeur de la connexion (plus la valeur de la connexion est importante plus il est possible de la réduire), soit :

$$J_{PQk} = I_{PQk} - oubli(\mu_P(O_k), J_{PQ(k-1)}) \qquad (R2)$$

$$(\text{où } I_{PQk}\ est\ défini\ en\ (R1))$$

On notera que l'on retrouve le même type de fonction d'oubli que dans le cas d'un apprentissage hebbien. Cette solution permet bien de réévaluer les implications appuyées par le plus grand nombre d'observations puisque le terme d'oubli pénalise plus celles qui ont le plus petit soutien. Cependant, elle introduit un biais lié à l'ordre de présentation des observations. En effet, les premières implications rencontrées seront oubliées plus de fois que les dernières. Ce biais peut être souhaitable (permettant une adaptation) si l'on considère que le système est dans un environnement changeant où les régularités peuvent évoluer. Mais si l'on souhaite par exemple, extraire les régularités d'un dessin composé de formes géométriques, il n'y a aucune raison de considérer un ordre sur ces formes.

Une autre solution, permettant d'éviter l'effet d'ordre, consiste à utiliser la règle (R1) jusqu'à épuisement des n observations, puis à agréger le résultat I_{PQ_n} obtenu, avec le nombre d'observations vérifiant P.

On écrit : $I_{PQ} = A(I_{PQn}, \sum_{i=1}^{n} \mu_P(O_k))$ (R3)

A titre d'illustration, dans l'exemple jouet qui suit, nous considérons la fonction $f_{soutien}$ (figure 4.2) qui favorise les implications dont le soutien est important avec : $I_{PQ} = f_{soutien}(\sum_{i=1}^{n} \mu_P(O_i)) I_{PQn}$ qui correspond donc à considérer une proportion plus ou moins grande de I_{PQ_n} suivant la valeur du soutien .

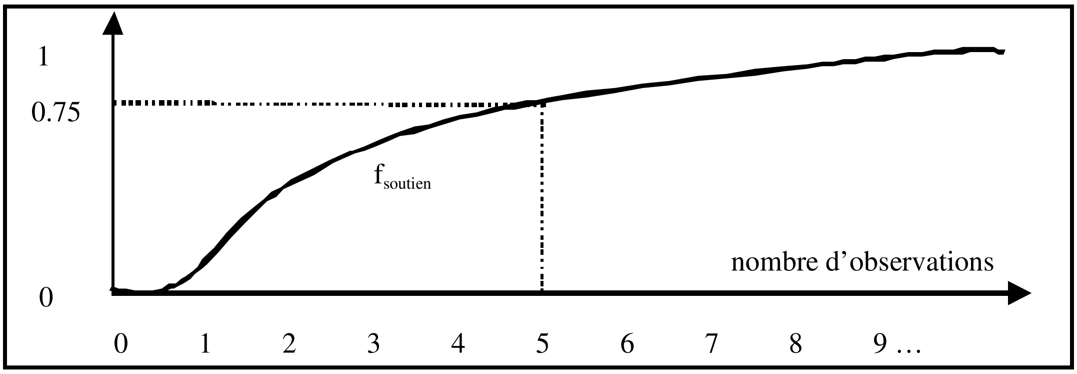


FIG. 4.2 Si une règle est appuyée par 5 observations, l'évaluation de l'implication est réduite à 75% de sa valeur.

Cependant l'agrégation de ces deux valeurs est délicate car elle renvoie à deux notions très différentes. En effet, il est problématique de considérer une règle sûre ayant un soutien faible comme une règle plus incertaine mais ayant un soutien fort. Nous proposerons une solution de nature très différente dans la suite.

4.1.3 Exemple

Dans cet exemple, le but est d'extraire les régularités d'un dessin (figure 4.4) dont les figures varient en forme, couleur, position verticale et position horizontale. On a considéré ici que les propriétés de forme et de couleur étaient discrètes. Les propriétés de position sont représentées par des sous-ensembles flous (figure 4.3).

En l'absence de considération d'ordre, nous avons utilisé (R3) avec l'agrégation définie ci-dessus (voir figure 4.2). L'étape correctrice permet de mettre en évidence, parmi les nombreuses régularités obtenues par (R1), celles qui ont le plus grand soutien (tableau 4.1).

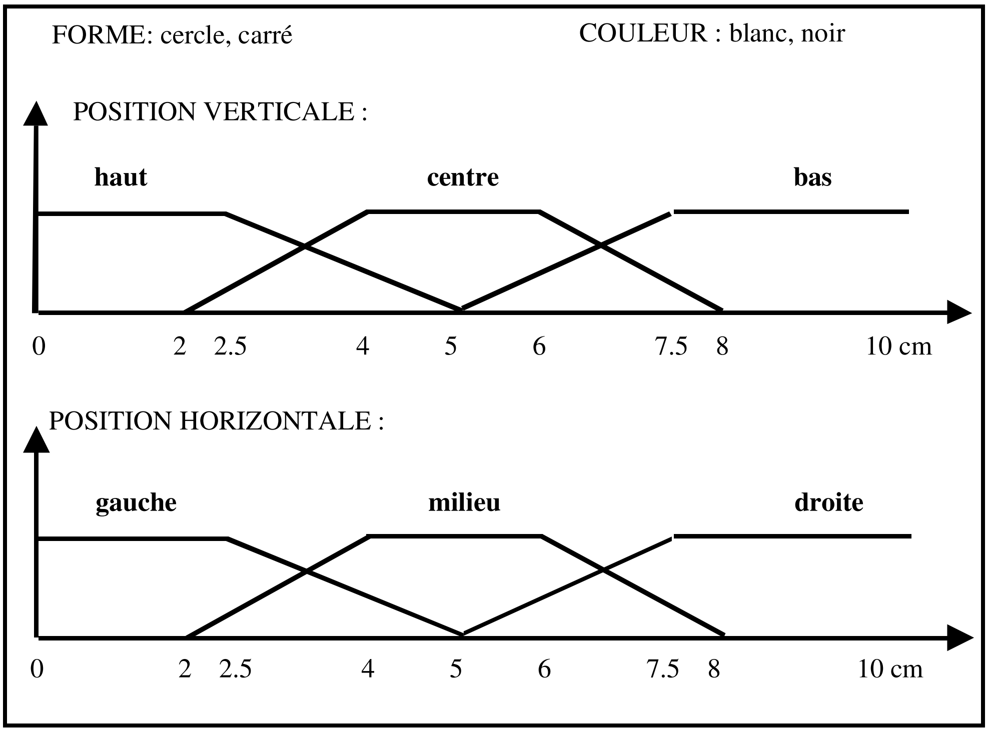


FIG. 4.3 Les différentes propriétés d'une figure.

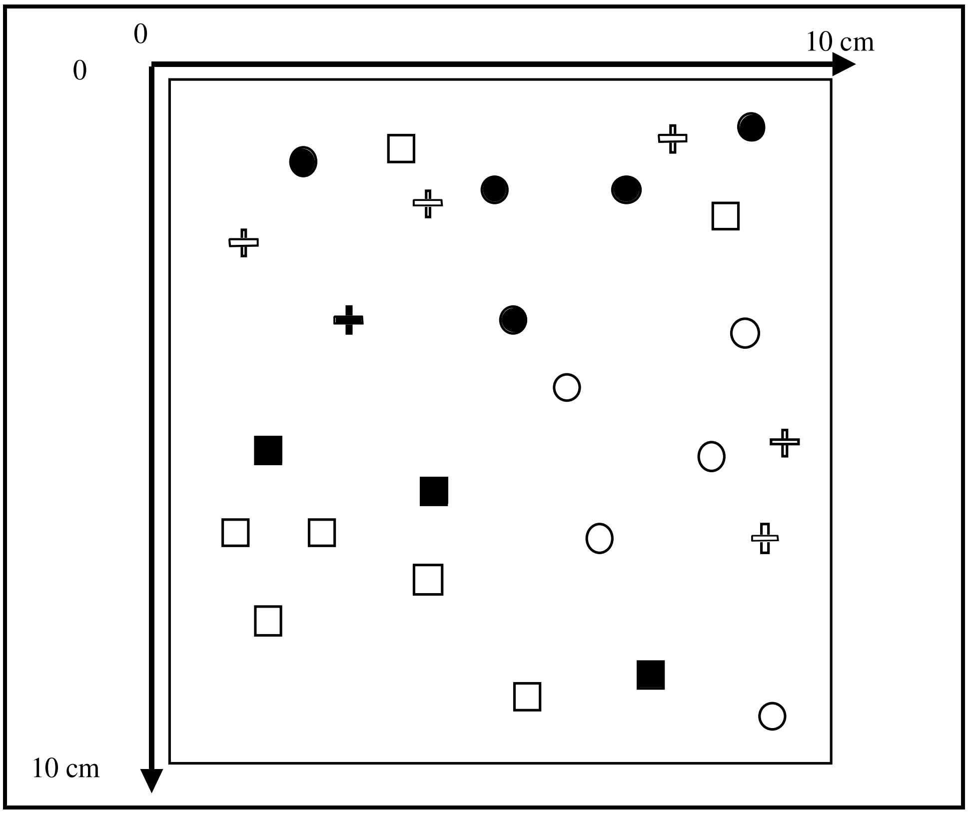


FIG. 4.4 Dessin à partir duquel les régularités doivent être extraites.

bas	→	blanc	0.77
droite	→	blanc	0.76
droite&centre	→	blanc	0.75
cercle&noir	→	haut	0.74
croix	→	blanc	0.71
noir&haut	→	cercle	0.71
carré	→	blanc	0.70
croix&droite	→	blanc	0.70
gauche&bas	→	carré	0.70
haut&cercle	→	noir	0.69

TAB. 4.1 - Les dix premières implications découvertes par le système avec R3.

4.2 Construction du réseau

4.2.1 Introduction

4.2.1.1 Le choix de la structure des règles

Les règles les plus couramment extraites par les algorithmes d'apprentissage automatique associent des conjonctions de propriétés et non des disjonctions. C'est aussi ce que nous considèrerons ici, et on peut noter que la conjonction de propriétés apporte généralement plus d'information que la disjonction de ces mêmes propriétés. En effet, dans le cas de règles sûres par exemple, $A \vee B \rightarrow C$ peut se réécrire $A \rightarrow C$ et $B \rightarrow C$ alors que $A \wedge B \rightarrow C$ ne permet pas d'inférer la vérité de $A \rightarrow C$ et $B \rightarrow C$. On notera aussi que deux propriétés prises séparément peuvent ne fournir aucune règle de décision sur un attribut particulier alors que leur conjonction peut parfaitement déterminer la valeur de cet attribut.

Enfin une autre limitation est de ne pas construire de connexion orientée vers des conjonctions, ces connexions étant moins informatives. En effet, si l'on a $A \rightarrow B$ avec un degré d_1 et $A \rightarrow C$ avec un degré d_2 et que l'on considère que d_1 et d_2 représentent des fréquences conditionelles, il est possible d'écrire $A \rightarrow B \wedge C$ avec un degré d tel que $d<\min(d_1,d_2)$ puisque la condition exprimée en conclusion est plus restrictive. Par contre, il n'est pas possible d'évaluer le degré d'implication attaché à "$A\&B \rightarrow C$" à partir de ceux attachés à "$A \rightarrow C$" et "$B \rightarrow C$" ce qui justifie que l'on construise des conjonctions en prémisse. Cette limitation est utilisée par Ganascia (1987) et est aussi reprise dans les premiers algorithmes rapides de recherche de d'associations (Agrawal & al., 1993) pour être abandonnée par la suite.

Pour résumer, nous considérerons donc des règles comprenant des conjonctions de propriétés, et on se limite éventuellement à des conclusions composées de simples propriétés. L'objectif général suivi dans cette partie est de limiter la complexité du réseau construit, tout en perdant un minimum d'information.

4.2.1.2 L'explosion combinatoire liée à la construction de conjonctions

Nous allons donc considérer un réseau dans lequel les nœuds représentent des conjonctions de propriétés. Le problème évident de ce type de réseau concerne donc l'explosion combinatoire liée à la construction des conjonctions. En effet, supposons qu'une observation porte p propriétés, le nombre de nœuds à générer correspond alors à toutes les conjonctions de 1, 2,…, p propriétés, soit :

$$C_p^1 + C_p^2 + \ldots + C_p^p = 2^p - 1$$

Cependant plus la prémisse de la règle comprend de propriétés, plus elle est spécifique, moins elle est utilisable et une solution peut consister à limiter le nombre maximal de propriétés dans une conjonction (Ganascia, 1987 ; Machado & Rocha, 1990).

4.2.2 Méthode

4.2.2.1 Une première méthode

A la différence de la plupart des réseaux de neurones, nous ne fixons ni l'architecture du réseau (celui-ci se développe au fur et à mesure de la prise en compte de nouvelles observations), ni son orientation (du fait que tous les nœuds sont à la fois des sorties et des entrées potentielles). Rappelons que dans ce réseau, les nœuds représentent des conjonctions de propriétés et que les connexions représentent des implications. Une première idée concernant la méthode de construction, est de considérer pour chaque observation, toutes les conjonctions de propriétés présentes et d'associer ces conjonctions. L'algorithme s'écrit alors :

Algorithme 1 :

```
Pour chaque observation faire :
Début
1- Construction des nœuds représentant toutes les conjonctions de propriétés présentes dans l’observation (les nœuds sont créés uniquement s’ils n’existent pas déjà dans le réseau).
2- Construction des connexions entre les nœuds correspondant aux conjonctions de propriétés présentes dans l’observation (les connexions sont créées uniquement si elles n’existent pas déjà dans le réseau).
3- Mise à jour de toutes les connexions (selon (R1) ou (R2)).
Fin
```

Cet algorithme a une complexité trop importante pour permettre de traiter un grand nombre de variables sans limiter le nombre de propriétés pouvant apparaître dans les conjonctions. Une solution doit donc être trouvée pour limiter la taille du réseau sans pour autant perdre d’information.

4.2.2.2 Le principe d'économie

Dans l'objectif de limiter la taille du réseau sans perdre d'information, nous utilisons un principe d'économie basé sur le constat que si tous les A sont B, les connexions ayant pour origine A valent pour la conjonction A&B et qu'il est donc inutile de construire un nœud pour représenter cette conjonction (on notera que CHARADE (Ganascia, 1987) utilise le même principe général pour limiter la combinatoire). Il s'agit donc d'introduire une condition pour la création des nœuds représentant les conjonctions. Supposons que la conjonction p_1 & p_2 & … & p_n soit vérifiée par l'observation courante. La condition nécessaire et suffisante pour la construction du nœud correspondant est :

[(le nœud représentant (p_1 & p_2 … & p_{n-1}) existe) ET ((p_1& p_2 & … & $\neg(p_n)$) a déjà été observée)] OU [(le nœud représentant (p_1 & p_2 & … & p_{n-2} & p_n) existe) ET ((p_1 & p_2 & … & p_{n-2} & $\neg(p_{n-1})$ & p_n) a déjà été observée)] OU … OU [(le nœud représentant (p_2 et … et p_{n-2} et p_n) existe) ET (($\neg(p_1)$ et p_2 et … et p_n) a déjà été observée)].

4.2.2.3 Implémentation du principe d'économie

Ainsi, lorsqu'un nœud est activé par l'observation courante (c'est-à-dire que la conjonction de propriétés qu'il représente est portée par l'observation), les prédictions qu'il permet de faire sont comparées aux propriétés vérifiées par l'observation pour la même variable. Si une prédiction certaine (implication évaluée à 1) est en contradiction avec la propriété portée par l'observation, les deux nœuds correspondant respectivement à la conjonction avec la propriété portée par l'observation et la conjonction avec la propriété prédite sont créés (figure 4.5).

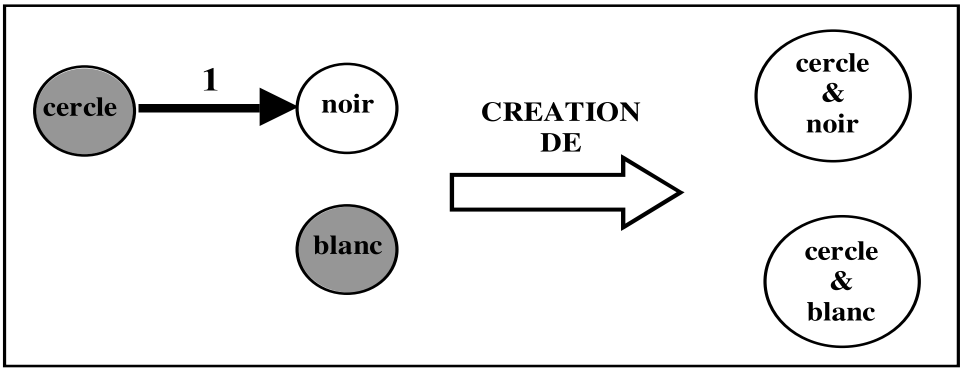


FIG. 4.5 L'observation remet en cause une règle certaine. Deux nouveaux nœuds sont créés. Les nœuds apparaissant en grisé sont les nœuds activés (c'est-à-dire que la conjonction de propriétés qu'il représente est portée par l'observation).

De plus, de façon à ne perdre aucune information, toutes les connexions du nœud courant activé sont recopiées pour le nœud créé par conjonction avec la propriété prédite avec certitude (figure

4.6). De cette façon le nouveau nœud est créé avec précisément les mêmes connexions qu'il aurait eu s'il avait été créé dès la première apparition et que les connexions avaient été mises à jour sur les différentes observations précédant celle-ci.

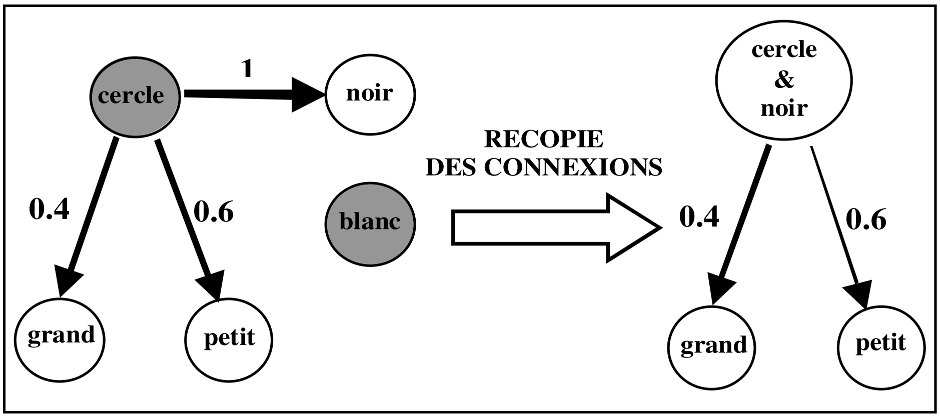


FIG. 4.6 L'observation remet en cause une règle certaine. Deux nouveaux nœuds sont créés. On recopie ensuite, les connexions du nœud à l'origine de la création pour le nœud créé par conjonction avec la propriété prédite avec certitude. Les nœuds apparaissant en grisés sont les nœuds activés (c'est-à-dire que la conjonction de propriétés qu'il représente est portée par l'observation).

S'il n'y a pas de prédictions certaines en contradiction avec l'observation mais que la propriété portée par l'observation n'est pas prédite (elle n'est jamais survenue simultanément avec la propriété représentée par le nœud), soit il n'existe pas encore de prédiction sur la variable correspondant à la propriété et dans ce cas aucun nœud n'est créé, soit il existe plus d'une propriété prédite pour la même variable (figure 4.7) et dans ce cas, le nœud correspondant à la conjonction avec cette propriété est créé (les autres conjonctions étant déjà construites).

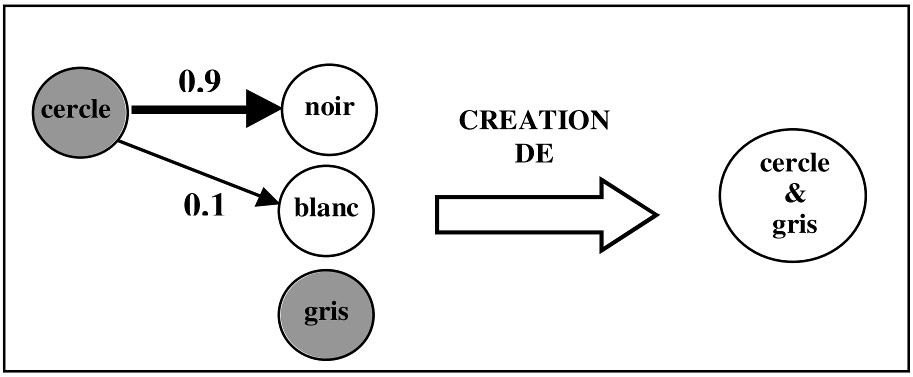


FIG. 4.7 L'observation porte une propriété qui n'est pas prédite, un nouveau nœud est créé. Les nœuds apparaissant en grisés sont les nœuds activés (c'est-à-dire que la conjonction de propriétés qu'il représente est portée par l'observation).

Le test de la condition création ainsi que ces mécanismes de création sont réalisés par la procédure Prédiction_Création. L'algorithme final s'écrit donc :

Algorithme 2 :

```
Pour chaque observation faire :
Début
1-Construction des nœuds simples représentant toutes les propriétés vérifiées par l'observation (les nœuds sont créés uniquement s'ils n'existent pas déjà dans le réseau).
2- Activation des nœuds représentant des conjonctions de propriétés vérifiées par l'observation.
3- Pour tous les nœuds activés, appeler la procédure Prédiction_Création.
4- Mise à jour de toutes les connexions (selon (R1) ou (R2)).
Fin
```

4.2.2.3 Economie réalisée

Cette procédure permet une limitation importante du nombre de nœuds et de connexions dans le réseau. Elle le rend encore plus dépendant du nombre d'observations puisque moins le nombre d'observations est important, moins il est probable d'avoir observé les conjonctions de propriétés qui conditionnent la création des nœuds. On notera de plus que plus les conjonctions de propriétés sont longues et moins il est probable qu'un nœud les représentant soit construit. En effet, la condition de création imposant l'observation d'une conjonction de longueur juste inférieure à celle devant faire l'objet de la création du nœud, plus cette conjonction comprend un nombre élevé de propriétés, moins il est probable que cette conjonction ait été observée.

Enfin, il peut arriver, comme dans le cas d'un problème de classification, que la recherche de prédictions porte préférentiellement sur certaines propriétés. Dans ce cas, une limitation du nombre de nœuds s'ensuit, puisqu'on peut alors construire le réseau uniquement à partir des observations contenant lesdites propriétés et que seule la mise à jour des implications portant sur ces propriétés est nécessaire.

4.2.3 Exemple

Nous comparons les deux algorithmes sur les 50 premières observations d'une base d'apprentissage comprenant 10 variables, relative à des mammographies (figure 4.8). On peut constater que le second algorithme permet un gain très substantiel en terme de nœuds construits, bien que ce nombre reste très grand. De manière générale, le gain est d'autant plus grand que les observations ne partagent pas de propriétés.

En effet dans le cas extrême où des propriétés toujours nouvelles apparaissent dans les observations successives, on construira avec le second algorithme, un nombre de nœuds égal au nombre de propriétés, alors qu'avec le premier algorithme, chaque observation donne lieu à la création de 2^p nœuds (p étant le nombre de propriétés portées par l'observation). Ainsi plus l'univers d'observation est ordonné et plus la diversité des propriétés est importante, plus le gain est important.

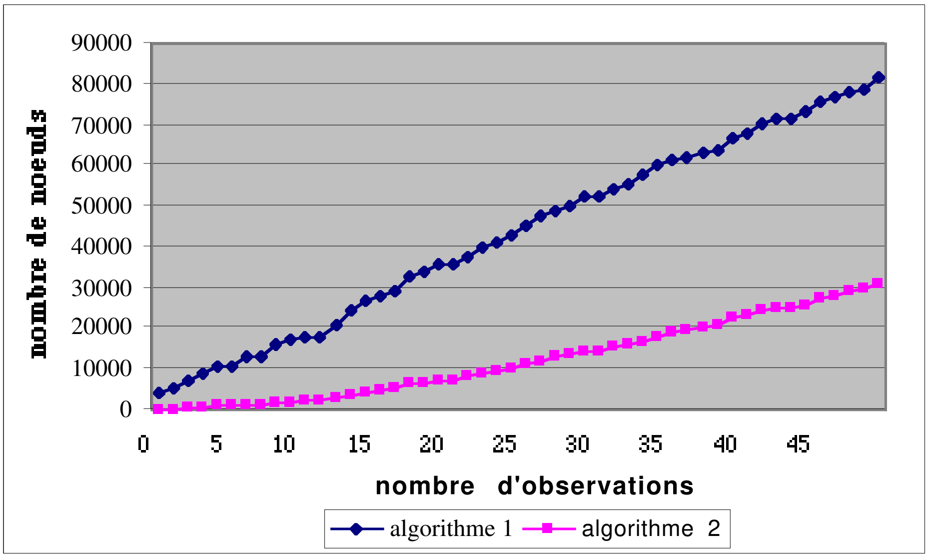


FIG. 4.8 Comparaison du nombre de nœuds construits par les deux algorithmes.

4.3 Construction automatique des sous-ensembles flous

4.3.1 Introduction

4.3.1.1 Le problème posé

Jusqu'à maintenant, nous avons supposé que les sous-ensembles flous étaient donnés a priori, c'est-à-dire, par exemple, acquis auprès d'experts (Aladenise & Bouchon-Meunier, 1997). Les méthodes d'acquisition posent cependant de nombreux problèmes et dans le cadre de la découverte de règles, une construction automatique peut être préférable. Considérons par exemple, dans l'illustration donnée pour l’évaluation des implications, la régularité "cercle&haut→noir" qui apparaît dans les dix premières implications obtenues avec (R3) (tableau 4 .1). Ainsi, le problème posé porte sur la construction automatique d'une définition de « haut » adaptée à cette implication.

4.3.1.2 Etat de l'art

Nous nous ramenons donc au problème de la discrétisation d'une variable continue. Pour ce problème, différentes solutions ont été proposées. Dans le cas où la discrétisation ne s'appuie pas sur un attribut particulier à prédire, on retient généralement deux solutions. Soit on considère des intervalles réguliers, soit on considère des intervalles de taille inversement proportionnelle au nombre d'observations qui s'y trouvent. Dans le cas où l'on s'appuie sur la distribution des observations sur un attribut particulier (la classe à prédire), de nombreuses solutions ont été proposées. On peut les classer selon deux approches opposées : l'approche dite ascendante qui considère une décomposition initiale et qui procède par la suite à des regroupements et l'approche dite descendante qui considère initialement un seul ensemble et qui le décompose ensuite. Pour l'approche ascendante, on peut citer le système ChiMerge (Kerber, 1992) qui part d'une décomposition totale considérant un intervalle par valeur et qui regroupe à chaque étape deux intervalles adjacents. Ce regroupement se base sur la statistique du χ^2 calculée pour chaque paire d'intervalles adjacents qui rend compte de la différence de répartition des observations dans les différentes classes pour les deux intervalles. Ainsi, les deux intervalles adjacents pour lesquels le χ^2 est le plus faible (c'est-à-dire les deux intervalles pour lesquels la différence de répartition sur les classes est la moins importante) sont regroupés. Le critère d'arrêt correspond à un seuil de significativité du test statistique, si les deux intervalles sont significativement différents, ils ne sont pas regroupés. Pour l'approche descendante, on peut citer le système de Fayyad et al. (1993) qui considère initialement un seul intervalle et qui à chaque étape choisit parmi les « points frontières » préalablement déterminés celui qui donne lieu à la bi-partition qui respecte le mieux la partition définie par les classes.

Par ailleurs, on peut considérer que la discrétisation par intervalle est trop brutale et préférer avoir recours à une discrétisation floue. Dans ce cadre, on peut citer les méthodes neuronales qui incluent l'apprentissage des paramètres déterminant les fonctions d'appartenance (Ichikawa & al., 1992), ou encore les méthodes basées sur la recherche d'une partition floue la plus discriminante possible vis-à-vis de la classe (Marsala, 1998) utilisant des mesures d'entropie floue.

4.3.2 Méthode

4.3.2.1 Principe général

C’est une approche ascendante que nous adoptons ici. Dans une première phase, nous construisons donc le réseau d'implications en considérant pour les variables continues, les propriétés obtenues par un découpage en petits intervalles de l'univers X associé à la variable. Ensuite, nous nous basons sur l'ensemble des implications ayant pour origine les intervalles P_i partitionnant les valeurs d’une même variable V_P et aboutissant à une propriété Q donnée. Chaque degré d'implication I_{P_iQ} correspondant après normalisation au degré d'appartenance à la propriété générale P_G ainsi définie (figure 4.9) . Le sous-ensemble flou P_G de X s’écrit alors :

$$P_G = \sum_i \frac{I_{P_iQ}}{\max\limits_i (I_{P_iQ})} / P_i \qquad (1)$$

L’implication résultante est $P_G \rightarrow Q$ avec une pondération égale à $\max\limits_i (I_{P_iQ})$ de façon à retrouver les différentes fréquences conditionnelles liant les différentes propriétés P_i à Q.

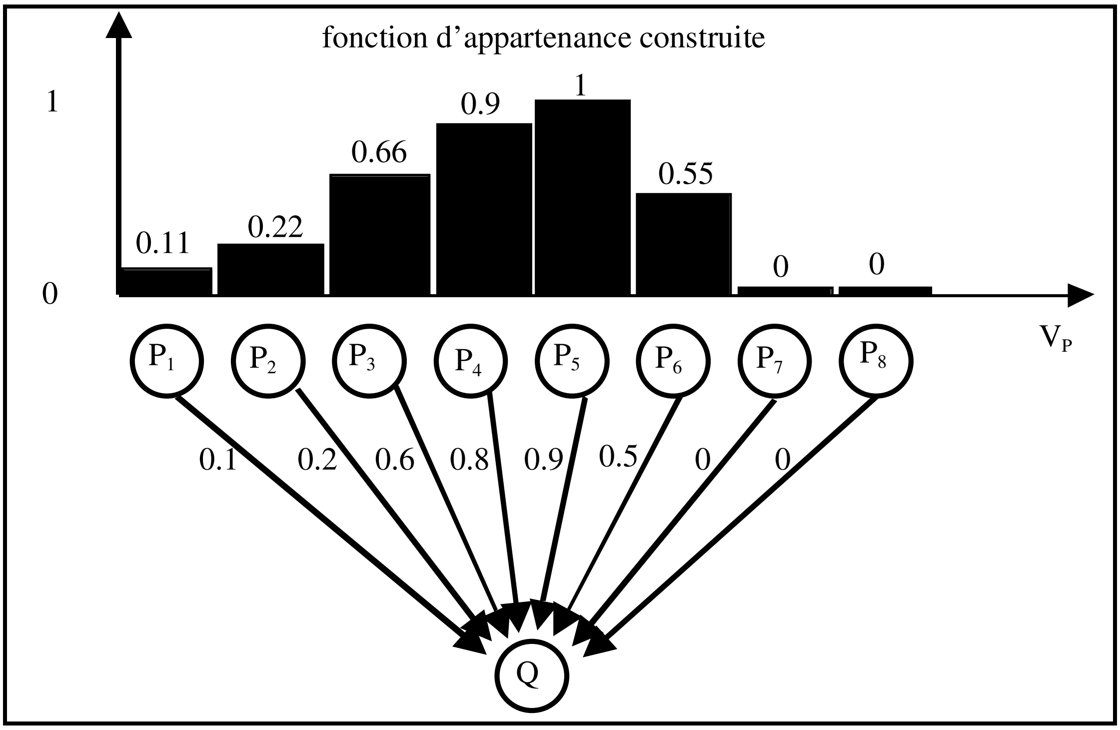


FIG. 4.9 Construction automatique d’une fonction d’appartenance sur la variable V_p par normalisation de l’évaluation des implications des intervalles P_i concluant sur une propriété Q donnée.

Ainsi, si l’on reprend l’exemple des figures géométriques, que l’on ne considère que les cercles et que l’on découpe l’univers des positions verticales en 10 intervalles réguliers non flous, une définition de « haut » adaptée à l'énoncé « les cercles en haut sont noirs » émerge automatiquement de la répartition des valeurs des implications des différents intervalles de position verticale vers la propriété « noir » (figure 4.10).

On notera que l'on définit ainsi des propriétés floues différentes pour une même variable, suivant la propriété Q considérée.

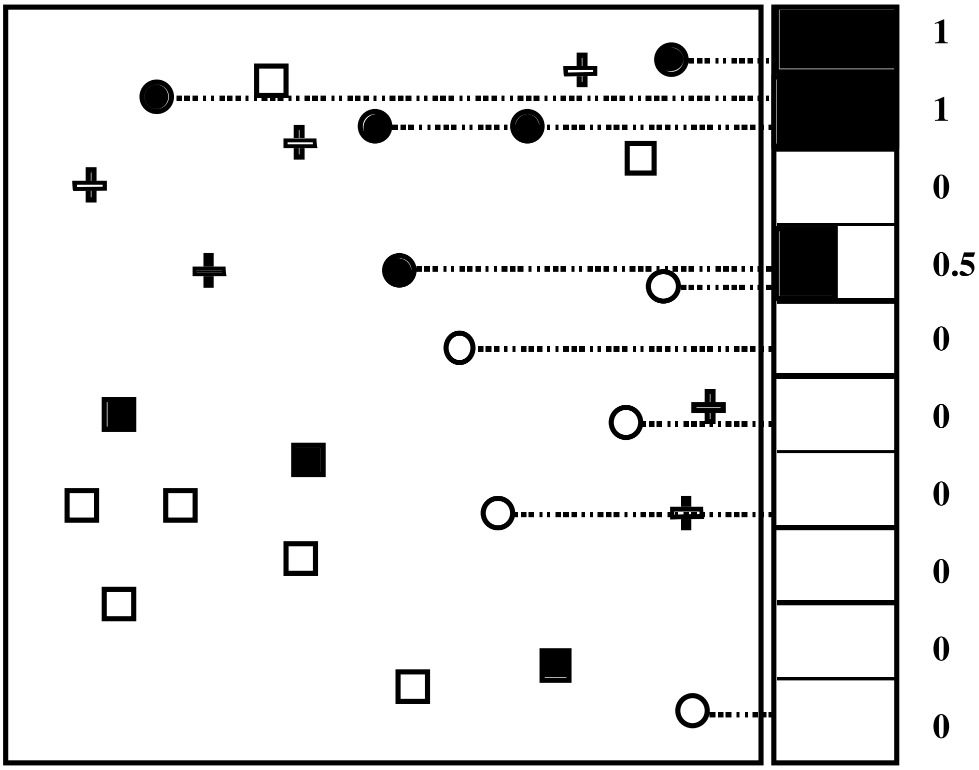


FIG. 4.10 Une définition de « haut » émerge de la répartition des valeurs des implications des différents intervalles de position verticale pour les cercles vers la propriété « noir ».

4.3.2.2 Résoudre le problème des données manquantes et harmoniser les fonctions d'appartenances

Par ailleurs, pour résoudre le problème des données manquantes (certains intervalles de valeurs, comme le troisième intervalle de la figure 4.10, peuvent ne jamais être observés) et celui plus pratique de l'harmonie des fonctions d'appartenance, nous proposons une méthode basée sur la notion de filtre qui consiste simplement à considérer non pas un découpage en intervalles bien définis mais un découpage en intervalles flous qui se recouvrent partiellement. Ainsi, une observation va préférentiellement correspondre à un intervalle, mais aussi, avec un degré moindre, aux intervalles voisins (voir Huang, 1997, pour une étude théorique de ce principe appliqué au traitement de données incomplètes). Cette prise en compte du voisinage a pour effet de rapprocher l'évaluation des implications partant d'intervalles voisins et donc de lisser la fonction d'appartenance. Il est à noter que des méthodes différentes mais aussi basées sur la notion de filtre avaient déjà été employées pour traiter ce problème (Marsala, 1998). Concernant la formalisation de ce principe, on utilise toujours (R1) mais on considère simplement des intervalles P_i flou et I_{P_iQ} est calculée suivant les mêmes principes que ceux donnés en 4.1.

Ainsi, sur le même exemple, en considérant cette fois qu'une observation compte totalement pour l'intervalle correspondant mais aussi avec un degré 0.5 pour le ou les intervalles adjacents, une définition plus harmonieuse de « haut » adaptée à l'énoncé « les cercles en haut sont noirs » émerge (figure 4.11).

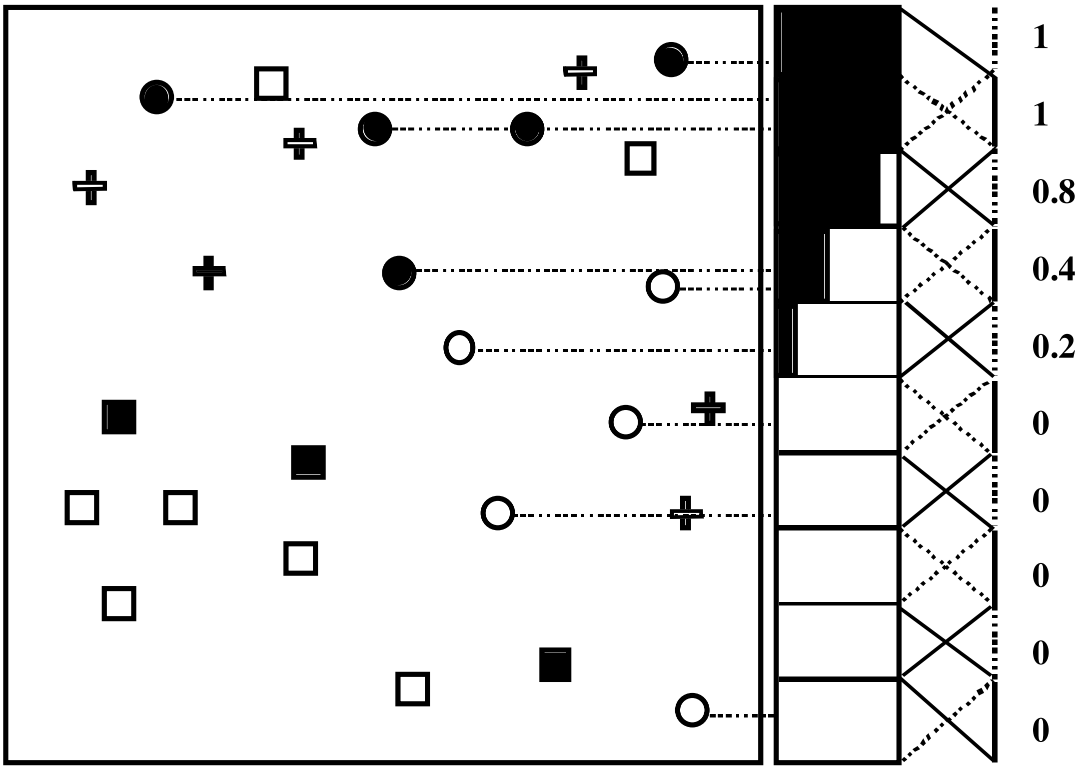


FIG. 4.11 Une définition plus harmonieuse de « haut » émerge de la répartition des valeurs des implications des différents intervalles de position verticale vers la propriété « noir » lorsque l'on prend en compte le voisinage (représenté ici par les trapèzes).

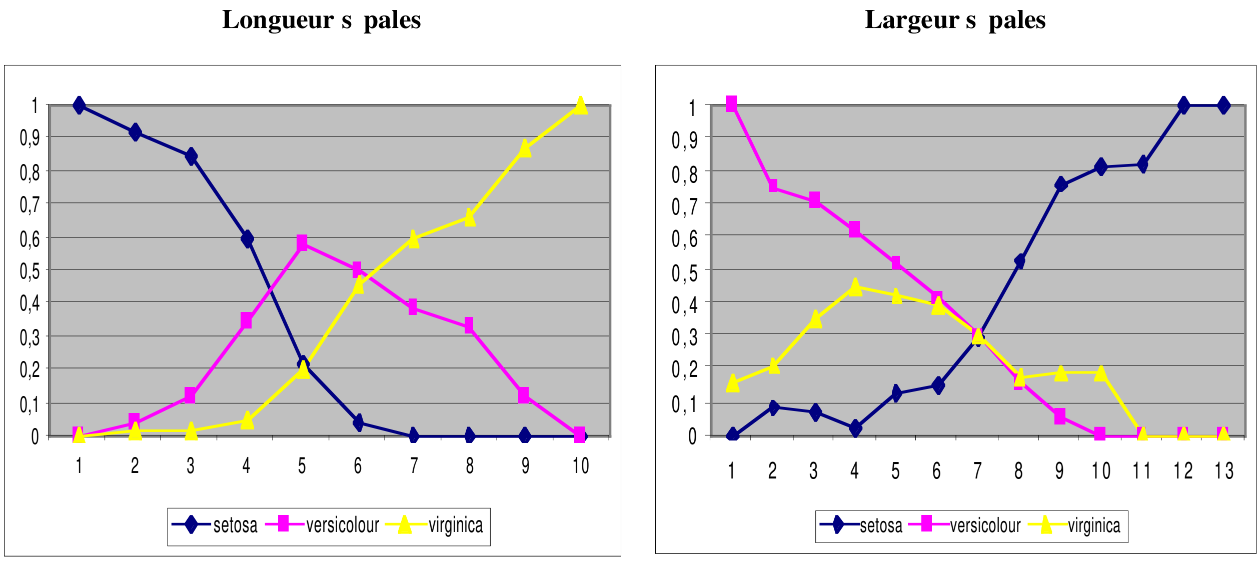
Ce principe de diffusion a différents avantages. Comme nous l'avons signalé, il permet de traiter les valeurs manquantes (dans la mesure où les intervalles voisins ne sont pas vides). De ce point de vue, il constitue une alternative à la première solution que nous avons donnée pour tenir compte du soutien dans l'évaluation des implications. En effet, plutôt que de faire décroître l'implication en fonction du nombre d'observations l'appuyant, ce qui amène lorsque le nombre d'observations est faible à la considérer comme une implication mal vérifiée (ce qui est faux), la diffusion attribue à l'implication sous représentée dans les observations, la valeur des implications voisines ayant un soutien plus fort. Il permet aussi d'harmoniser les fonctions d'appartenance de façon adaptée puisque c'est le voisinage qui détermine cette harmonisation et que la modification de la distribution d'origine est donc faite à partir des observations les plus proches. Un effet intéressant est celui de l'augmentation de l'évaluation d'une implication dont les implications voisines à gauche et à droite sont fortes comme nous le verrons dans le chapitre 5.

4.3.3 Exemple

Nous avons appliqué cette méthode de discrétisation à la base UCI des iris (figure 4.12). Nous avons utilisé un découpage par filtres trapézoïdaux (tels que pour toute valeur, il existe un filtre dont le noyau l'inclut) en prenant 10 filtres pour la largeur des sépales (ce qui correspond à une largeur de 0,4 cm pour le noyau de la fonction d'appartenance), 13 pour la largeur des sépales (correspondant à 0,2 cm), 12 pour la longueur des pétales (correspondant à 0,5 cm), et 13 pour la largeur des pétales (correspondant à 0,2 cm).

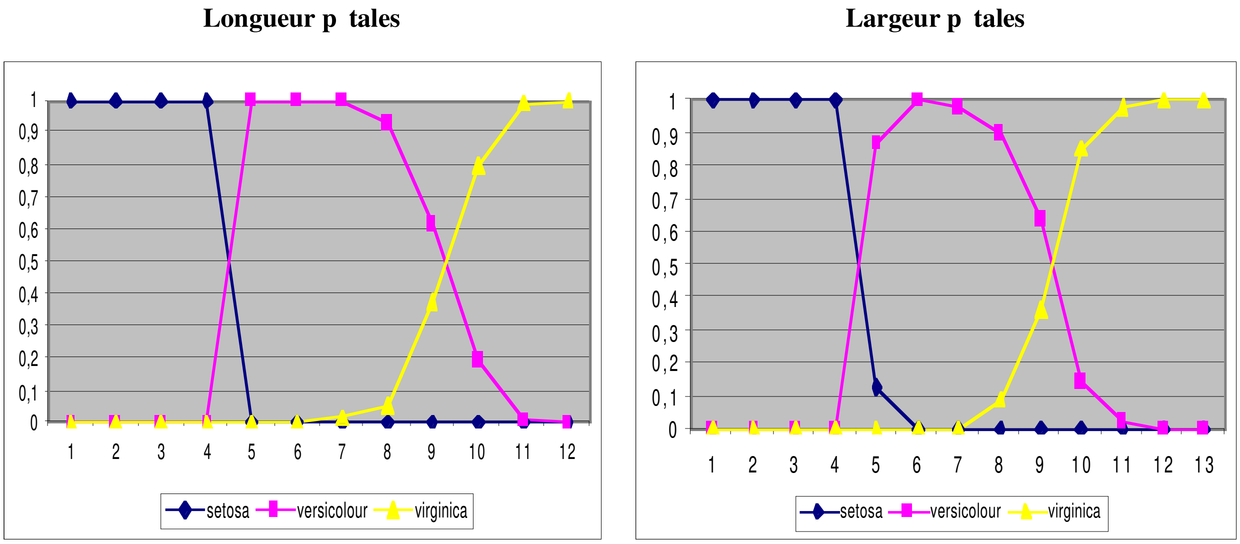
Les résultats obtenus corroborent le fait que la corrélation entre les dimensions des pétales et la classe d'iris est plus forte que celle des sépales. On notera aussi qu'à la différence de Hong et Lee (1996), notre méthode nous permet d'extraire des règles pour les sépales. Notre problème ici n'étant pas la classification, nous ne nous étendons pas sur les performances de cet ensemble de règles pour prédire la classe des iris. Par ailleurs nous n'avons pas représenté ici les distributions obtenues pour des conjonctions de propriétés, mais la méthode s'applique sans restriction à ce cas, les filtres sont simplement alors en plusieurs dimensions.

On notera (figure 4.12) que l'étiquette linguistique est attribuée a posteriori et manuellement. On peut cependant envisager pour chaque utilisateur, la possibilité de définir un certain nombre d'adjectifs par des sous-ensembles flous et un certain nombre d'adverbes comme « plutôt » par des coefficients qui, s'ils s'apparient bien avec respectivement un sous-ensemble flou construit et un coefficient extrait, peuvent être réutilisés dans la formulation automatique des règles.



SI longueur sépales = petite ALORS setosa
SI longueur sépales = moyenne ALORS plutôt versicolour
SI longueur sépales = grande ALORS virginica

SI largeur sépales = petite ALORS versicolour
SI largeur sépales = moyenne ALORS plutôt virginica
SI largeur sépales = grande ALORS setosa



SI longueur pétales = petite ALORS setosa
SI longueur pétales = moyenne ALORS versicolour
SI longueur pétales = grande ALORS virginica

SI largeur pétales = petite ALORS setosa
SI largeur pétales = moyenne ALORS versicolour
SI largeur pétales = grande ALORS virginica

FIG. 4.12 Nous avons représenté ici les fréquences conditionnelles évaluées pour chaque filtre à partir de (R1). Les fonctions d'appartenance s'obtiennent par normalisation des distributions de probabilités, ce qui n'est nécessaire ici que pour les propriétés "moyenne" pour les sépales. Nous avons attribué une étiquette linguistique à chaque fonction d'appartenance construite et nous avons modéré le "SI ALORS" par "SI ALORS plutôt" lorsque la règle n'était pas certaine.

4.4 Exploitation du réseau sémantique flou

4.4.1 Introduction

4.4.1.1 Une utilisation naturelle des méthodes de propagation d'activation

La règle d'inférence classiquement utilisée en logique est le Modus Ponens qui à partir d'une proposition P et d'une implication P→Q, permet d'inférer Q. Dans le cadre de la logique floue où les propositions peuvent être imprécises, le Modus Ponens généralisé a été défini par Zadeh (1976). Dans ce cadre, on considère une règle floue « SI V est P ALORS W est Q » unissant deux variables V et W respectivement définies sur les univers X et Y, avec P et Q respectivement définis par les fonctions d'appartenance μ_P et μ_Q. On considère de plus une observation P' éventuellement floue définie par $\mu_{P'}$. Le résultat Q' de l'inférence par le Modus Ponens généralisé est donné par :

$$\forall y \in Y \quad \mu_{Q'}(y) = \sup_{x \in X} T(\mu_{P'}(x), I(x,y))$$

où T est une t-norme et I une implication floue.

Cependant ici, d'une part, on ne considère pas une implication matérielle mais une fréquence conditionnelle et d'autre part, on dispose d'un réseau. La façon la plus naturelle d'utiliser ce type de représentation semble donc de propager les probabilités dans le réseau, dans l'idée des méthodes de propagation d'activation. Par ailleurs, bien que nous le puissions, nous n'utilisons pas le réseau pour résoudre un problème de classification. Nous cherchons plus globalement à extraire l'information pertinente, c'est-à-dire à découvrir des régularités dans les données et à sélectionner les informations liées à une information particulière exprimée par exemple, dans une requête. Les méthodes de propagation d'activation s'appliquent de façon naturelle à ce type de problème.

4.4.1.2 Considérer la spécificité

L'un des écueils de ce type de méthode est sa tendance à propager l'activation sur l'ensemble du réseau (Crestani, 1997). Pour faire face à ce problème, différentes solutions ont été proposées (voir chapitre 3). Il s'agit, soit de faire décroître l'activation lors de son parcours lorsqu'elle s'éloigne de sa source, soit de privilégier certains types de chemins, soit de ne pas propager l'activation arrivant aux nœuds à forte connectivité, soit enfin, d'exiger que l'activation arrivant au nœud dépasse un certain seuil pour être propagée.

Or, il semble que la définition de la pertinence en terme de spécificité qu'on peut aussi associer à l'idée de résonance, permette de traiter plus globalement ce problème. En effet, la pertinence d'une information I vis-à-vis d'un objectif ou plus généralement d'une situation S est particulièrement liée à la spécificité de cette information vis-à-vis de la situation ce qui renvoie donc à l'évaluation de « I→S » (voir chapitre 2). Bien que l'évaluation de l'implication inverse puisse avoir un rôle, elle semble être secondaire et la sélection des informations sur le critère de l'évaluation de « I→S » est plus révélatrice d'une véritable relation.

Ainsi, étant donné une requête portant sur un concept C, le fait de considérer les concepts C' tels que C'→C plutôt que ceux tels que C→C' permet d'éviter les concepts C' à forte connectivité. En effet, si C' est un concept très général, il survient en présence de nombreux autres concepts et l'implication C→C', s'explique par la fréquence marginale élevée de C' au contraire de l'implication opposée.

Considérons par exemple, une requête portant sur une maladie « x ». Compte tenu de la grande densité de population à Paris, on a toutes les chances d'avoir une implication « maladie x→Paris » particulièrement forte relativement aux autres implications portant sur les autres départements. Il est pourtant possible que cette maladie « x » ne soit pas particulièrement présente à Paris si on la rapporte au nombre d'habitants. C'est l'idée que l'on retrouve dans le calcul d'indices statistiques tenant compte de la fréquence marginale (tel que l'indice de Loevinger) ainsi que dans le terme noté idf du modèle vectoriel en recherche d'information (voir chapitre 3).

Au contraire, la grandeur caractérisant l'implication « Paris→x » relativement aux autres implications « département→x » est plus révélatrice puisqu'elle traduit la forte présence de la maladie x à Paris relativement à sa présence dans d'autres départements. C'est ce principe que nous utilisons pour la sélection d'information pertinente.

4.4.2 Méthode

4.4.2.1 Principe général

Considérons une propriété représentée par un nœud Q dans le réseau. La recherche de l'information pertinente vis-à-vis de cette propriété se décompose en deux temps.

Dans un premier temps, on propage l'activation aux nœuds P_i en suivant le sens de l'implication $Q \rightarrow P_i$. Les nœuds P_i auxquels on accède sont donc tels que $Q \rightarrow P_i$ est non nul (sinon la connexion n'existe pas) et aussi tels que $P_i \rightarrow Q$ est non nul . En effet $Q \rightarrow P_i$ est non nul si et seulement si les propriétés P_i et Q sont au moins portées avec un degré différent de 0 par une même observation ce qui est équivalent à $P_i \rightarrow Q$ non nul.

Dans un second temps, on considère les nœuds P_i en les regroupant par variable et pour chaque variable similairement à la méthode de discrétisation, on considère la distribution des implications $P_i \rightarrow Q$. Comme nous l'avons vu, cette distribution s'écrit comme un sous-ensemble flou P_G défini sur l'univers des propriétés caractérisant la variable considérée, où l'on associe à chaque propriété P_i l'évaluation de $P_i \rightarrow Q$ notée I_{P_iQ}, Soit :

$$P_G = \sum_i \frac{I_{P_iQ}}{\max_i(I_{P_iQ})} / P_i .$$

Considérant la relation de spécificité comme l'indice de pertinence de l'information P_i vis-à-vis de Q comparée aux autres informations P_j, on mesure ensuite le relief de ce sous-ensemble flou normalisé pour chaque variable (figure 4.13). Pour mesurer ce relief, nous avons choisi une mesure dite de "spécificité" qui mesure le degré avec lequel le sous-ensemble flou se rapproche de la spécificité parfaite qui consiste en un singleton (un élément appartenant à l'ensemble avec un degré 1, tous les autres éléments n'appartenant pas à l'ensemble).

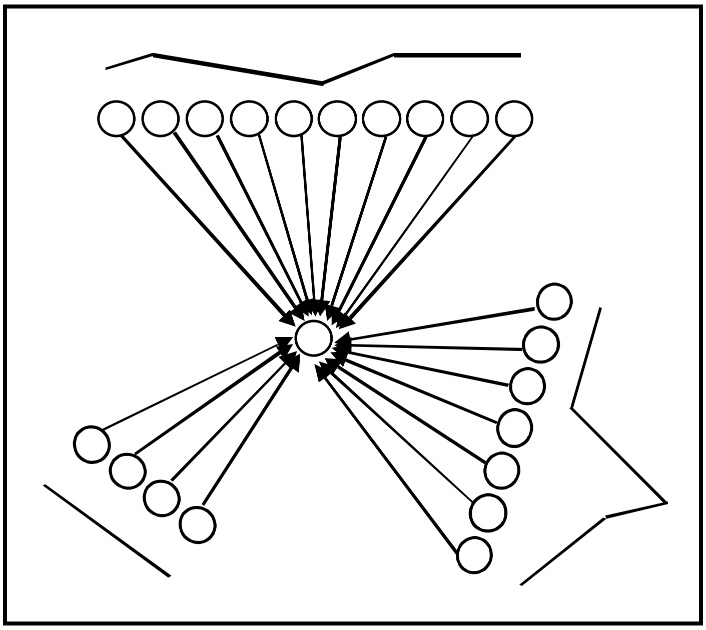


FIG. 4.13 On calcule la spécificité des sous-ensembles flous construits pour chaque variable sur la base des implications entrantes vis-à-vis d'un nœud particulier.

La normalisation appliquée au sous-ensemble flou permet de considérer le rapport de chaque degré d'appartenance relativement au plus grand de ces degrés de façon à obtenir un résultat identique par exemple pour $A = 0.8/x_1 + 0.4/x_2 + 0.2/x_3$ et $B = 0.4/x_1 + 0.2/x_2 + 0.1/x_3$ dont la différence peut être simplement liée à un découpage plus ou moins fin de la variable correspondant à la propriété pour laquelle on considère les implications entrantes.

4.4.2.2 Choix d'une mesure de spécificité

Une mesure de spécificité est une fonction (notons-la Spe) qui associe à tout ensemble flou une valeur dans [0,1] et qui a de plus les propriétés suivantes :

Soient F, F' deux sous-ensembles flous.

- Si $F \subseteq F'$ alors $Spe(F) \geq Spe(F')$
- Spe(F) = 1 si et seulement si F est un sous-ensemble réduit à un élément.

Différentes mesures de spécificité ont été proposées (Yager, 1982), nous considérons ici la mesure suivante :

Soit m un indice tel que $\forall i \;\; \mu(e_m) \geq \mu(e_i),\; Spe(F) = \mu(e_m) - \frac{1}{(n-1)} \sum_{i \neq m} \mu(e_i)$

Cette mesure est un cas particulier simple des mesures de spécificité linéaires strictes proposées par Yager (1992). Si nous ajoutons l'hypothèse que le sous-ensemble flou est normalisé, on notera que $\mu(e_m) = 1$. Contrairement à d'autres mesures de spécificité, elle permet de s'abstraire des découpages plus ou moins fins des différentes variables pour lesquelles on considère les implications sortantes. En outre, elle ne fait pas intervenir d'alpha-coupes qui, du point de vue de l'implémentation, sont toujours coûteuses. On notera enfin qu'elle se visualise très bien puisqu'elle s'interprète comme l'aire de la négation du sous-ensemble flou (figure 4.14).

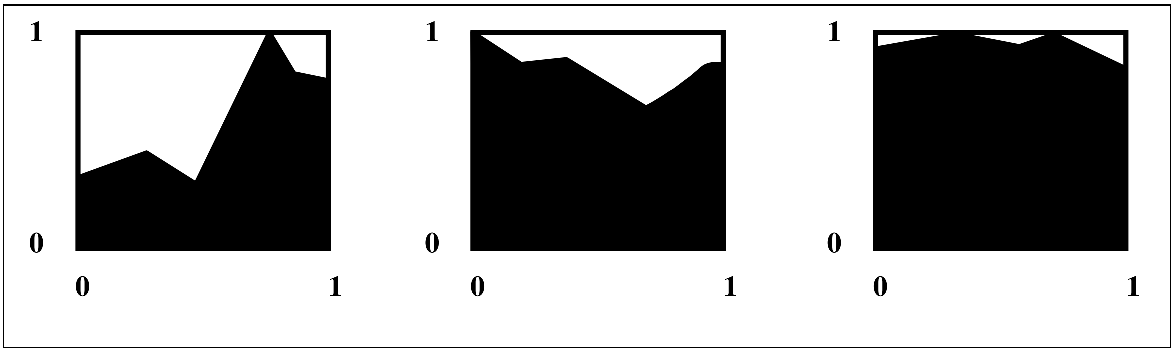


FIG. 4.14 Trois exemples de sous-ensembles flous de gauche à droite dans un ordre de spécificité décroissante. La spécificité correspond à l'aire de la partie blanche.

4.4.2.3 Compléments

4.4.2.3.1 Divers

La première propagation permet une première sélection des informations liées et le calcul de la spécificité prenant en compte les implications inverses raffine cette sélection. L'intérêt de l'utilisation du réseau qui s'implémente particulièrement bien avec des pointeurs et des langages à objets est la rapidité avec laquelle il est possible d'accéder à l'information liée.

Par ailleurs, on notera que le traitement des valeurs manquantes par le principe de diffusion est en accord avec le principe de spécificité maximale contrairement à l'utilisation du terme d'oubli introduit dans R3 ou R2 qui aurait introduit une spécificité artificielle liée à la diminution de l'évaluation des implications soutenues par un faible nombre d'observations.

4.4.2.3.2 Prise en compte des relations transitives

Enfin, nous avons souhaité tenir compte du cas dans lequel une implication A→C s'explique par un lien transitif soit : A→B et B→C en considérant que C est plus dépendant de B que de A dans ce cas et que A ne doit donc pas apparaître comme information pertinente vis-à-vis de C. Nous avons donc adopté l'heuristique qui consiste à considérer que A→B et B→C expliquent A→C si $d_{ab}.d_{bc} + (1-d_{ab}).d_c > d_{ac}$ où d_{ab}, d_{bc}, d_{ac} représentent respectivement les fréquences conditionnelles évaluant l'implication et d_c représente la fréquence marginale des C. Ce test estime la proportion des A qui sont B et qui de ce fait sont C (à laquelle s'ajoute la proportion des A qui ne sont pas B et qui sont C) et la compare avec celle des A qui sont C. Si la proportion des A qui sont C est inférieure à celle des A qui sont B et qui de ce fait sont C (à laquelle s'ajoute la proportion des A qui ne sont pas B et qui sont C), on considère que A→C s'explique par A→B et B→C (figure 4.15). Ce test est empirique puisqu'il fait l'hypothèse qu'on trouvera autant de C parmi les B qui sont A que parmi les autres B.

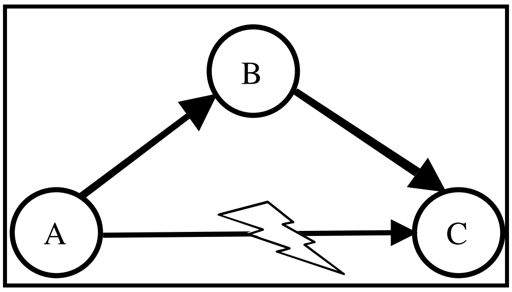


FIG. 4.15 L'implication A→C est expliquée par A→B et B→C.

On ne tient ensuite pas compte de telles implications qui ont de plus la particularité d'être des valeurs maximales pour les distributions considérées dans l'évaluation de la spécificité (figure 4.16).

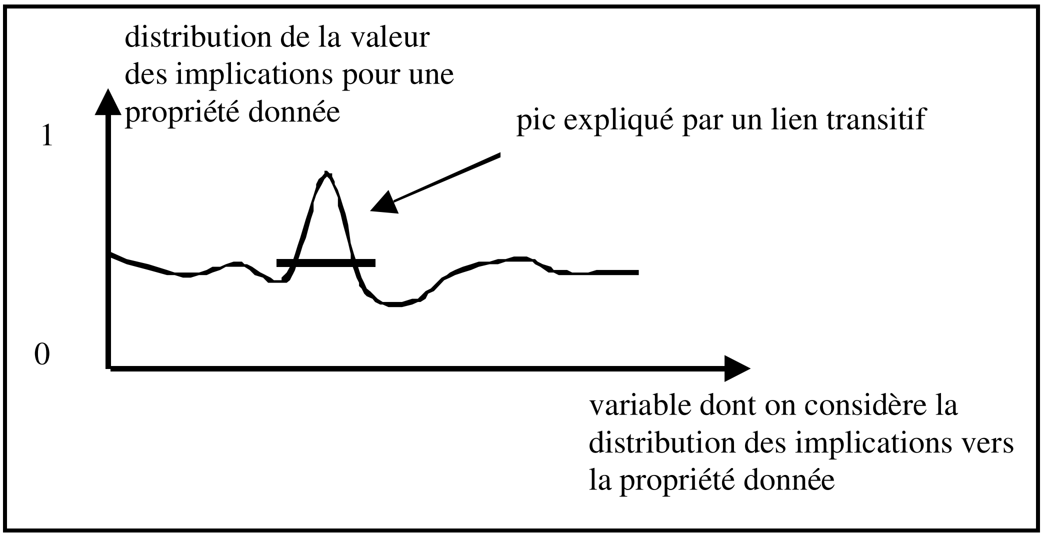


FIG. 4.16 Le relief de la distribution des implications disparaît lorsque l'on considère les liens transitifs des valeurs maximales.

4.4.3 Exemple

Nous reprenons ici le même exemple que précédemment (figure 4.4.bis) en ne considérant que les implications simples ne faisant pas intervenir de conjonctions de propriétés. Nous comparons les dix règles les plus fortes du point de vue de la fréquence conditionnelle (calculée par R1) et du point de vue de la spécificité. Pour la spécificité, la valeur correspond à la spécificité calculée pour la variable et la propriété retenue en prémisse correspond à la propriété dont l'implication est la plus forte. Pour la spécificité, A→B s'interprète comme A est spécifique à B et donc A est pertinent vis-à-vis de B ou encore dans le cadre d'une décision sur B, A est une information utile à la prédiction de B.

Ainsi, lorsque l'on calcule les fréquences conditionnelles, les quatre premières implications extraites concluent sur la propriété « blanc » (tableau 4.2). Pourtant, ni « carré » ni « droite » ne sont des propriétés spécifiquement liées à « blanc », ces implications sont simplement dûes à la sur-représentation des figures blanches. Comme on peut le constater, la prise en compte de la spécificité permet de tenir compte de ce problème sans pour autant défavoriser systématiquement les propriétés fréquemment observées. On notera aussi qu'aucune spécificité n'est expliquée par transitivité.

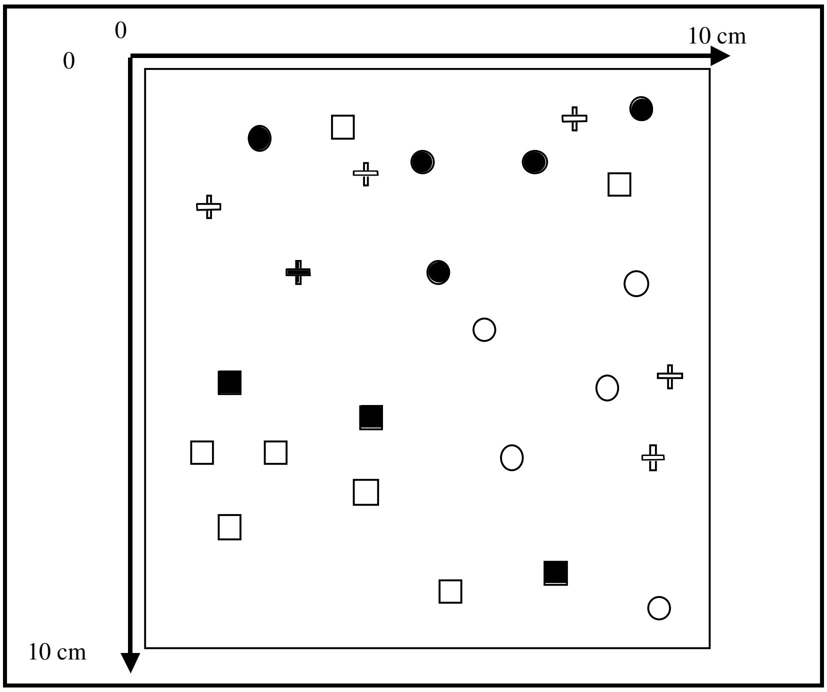


FIG. 4.4.bis Rappel du dessin à partir duquel les régularités sont extraites.

Les dix règles les plus fortes

fréquences conditionnelles				spécificités			
croix	→	blanc	f : 0.83	carré	→	bas	s: 0.72
droite	→	blanc	f : 0.76	blanc	→	croix	s: 0.62
carré	→	blanc	f : 0.70	blanc	→	bas	s: 0.57
gauche	→	blanc	f : 0.62	carré	→	gauche	s: 0.54
gauche	→	carré	f : 0.61	cercle	→	noir	s : 0.53
croix	→	haut	f : 0.60	gauche	→	carré	s : 0.53
carré	→	gauche	f : 0.58	haut	→	croix	s : 0.48
noir	→	haut	f : 0.58	noir	→	cercle	s : 0.47
carré	→	bas	f : 0.58	haut	→	noir	s : 0.46
milieu	→	blanc	f : 0.56	noir	→	haut	s : 0.41

TAB. 4.2 Comparaison des règles les plus fortes du point de vue de la fréquence conditionnelle et du point de vue de la spécificité. Le second point de vue permet d'éliminer des règles qui ne traduisent pas de dépendances réelles entre propriétés mais qui proviennent simplement de la sur-représentation de certaines propriétés (par exemple, blanc)

4.5 Discussion

4.5.1 Discussion générale

L’architecture du système RELIEFS est résumé dans la figure 4.17. Ainsi, la prise en compte de chaque observation se traduit par l’activation des différents nœuds du réseau à des degrés divers (c’est à ce niveau qu’intervient le principe de diffusion). Par comparaison des prédictions de propriétés permises par le réseau et celles portées par l’observation, des nouveaux nœuds et de nouvelles connexions sont créés. Enfin, les connexions des nœuds activés simultanément sont renforcées. Dans cette première étape, on considère des propriétés résultant du découpage le plus fin possible des différentes variables. Dans une seconde étape, on regroupe ces propriétés pour construire des sous-ensembles flous de façon automatique et on sélectionne les plus spécifiques (soit globalement, soit par rapport à une requête particulière).

La première limitation de notre approche est liée à la taille du premier réseau construit (avant la sélection des sous-ensembles flous les plus spécifiques). La solution généralement retenue, et qui est celle sur laquelle reposent les algorithmes de recherche rapide d'association, est de supprimer les implications dont l'évaluation est inférieure à un certain seuil. Cette solution est discutable car elle défavorise les implications portant sur des propriétés rares et favorise celles portant sur les propriétés les plus fréquentes. Même si, par la suite, des indices statistiques peuvent être utilisés pour sélectionner les règles les plus significatives, certaines implications sont irrémédiablement perdues dans la première phase.

Les algorithmes rapides de recherche d'associations ont aussi le désavantage de procéder en deux étapes, ce qui rend impossible la prise en compte simultanée des aspects inductif et descriptif de la mesure de l'implication. De plus, le seuil imposé dans la première phase est exprimé en pourcentage. Ainsi, lorsque la base comprend dix millions d'observations, le taux de 0.25% (qui est l'exemple donné dans le meilleur des cas pour l'algorithme APRIORI) correspond à 25 000 observations. Or, il doit être possible d'extraire une règle vérifiée sur simplement cent observations, même si cette règle s'applique plus rarement. Notre méthode, bien que limitée dans ses applications sur de grandes base de données, apporte donc des éléments qui ne sont pas présents dans les algorithmes rapides de recherche d'associations.

Si l’on compare RELIEFS aux autres systèmes d’extraction de règles floues (Hong & Lee, 1996 ; Wang & Mendel, 1992 ; Wu & Chen, 1999). On notera que, comme le font remarquer Hong et

Lee (1996), peu de méthodes permettent de construire automatiquement les sous-ensembles flous. Nous ajouterons également que ces méthodes, lorsqu'elles existent, font des hypothèses sur la forme des fonctions d'appartenance (triangle ou trapèze). De plus, elles se décomposent en un grand nombre d'étapes, alors qu'ici l'ensemble de l'apprentissage est intégré dans la règle locale. Enfin, elles font le plus souvent intervenir une mesure de distance donnée a priori qui rajoute une hypothèse supplémentaire.

Comparativement au système proposé par Girard et Ralambondrainy (1996), on notera, que nous ne partons pas d'un ensemble de nuances données mais que nous les construisons automatiquement. D'autre part, notre réseau est de nature associative et non hiérarchique et cela tient au fait que nous ne considérons pas une relation binaire qui est à la base du treillis de Galois mais que nous souhaitons garder l'information exprimée par les degrés d'implication et que nous ne voulons donc pas considérer de seuil.

Concernant l'exploitation du réseau, et la prise en compte des distributions marginales, on notera que, contrairement aux différents indices statistiques proposés, nous introduisons ce critère, non pas au niveau des propriétés, mais au niveau de la variable.

Enfin, pour la conception de ce système, nous nous sommes inspirés de principes cognitifs. Nous détaillons, dans le paragraphe suivant, plus amplement ces principes en établissant un parallèle avec le système cognitif. Notre système ne constitue pas un modèle dans le sens où tout ceci reste très empirique et n'est pas confronté à des données expérimentales. Il s'agit simplement d'établir une correspondance entre un intérêt algorithmique et une donnée concernant le système cognitif. La nature est une source d'inspiration et certains principes généraux sur notre fonctionnement sont maintenant connus. Lorsqu'on abandonne cette recherche d'imitation au profit de la recherche d'une solution pragmatique et viable, il arrive que l'on retrouve une donnée concernant le système cognitif.

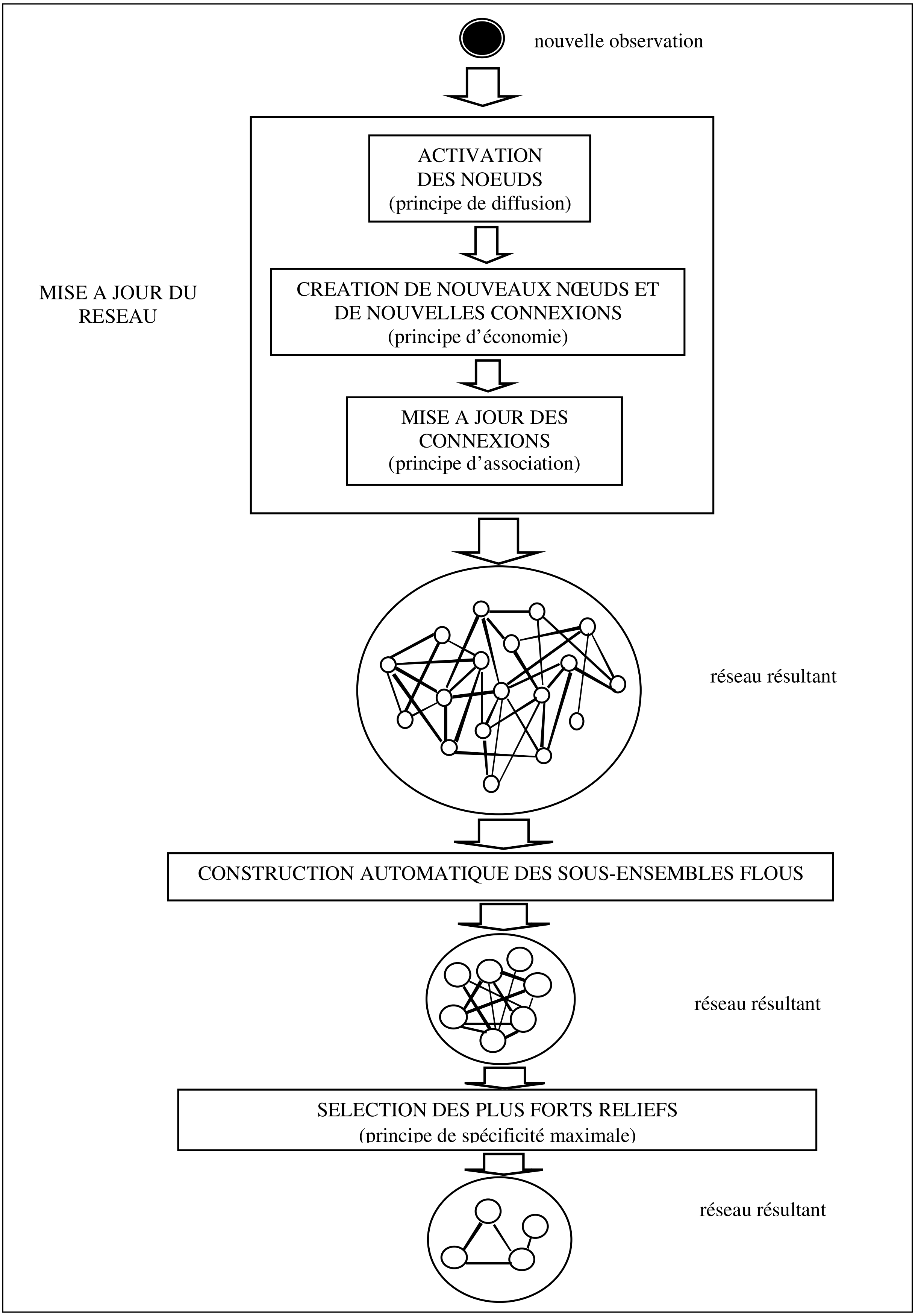


FIG. 4.17 Architecture du système RELIEFS (on pourra se reporter à l'annexe pour le détail des principales procédures de RELIEFS).

4.5.2 Point de vue cognitif

4.5.2.1 Principes liés à la règle d'apprentissage

On sait que le principe d'association joue un rôle important dans le fonctionnement de la mémoire. Aristote décrivait déjà un ensemble de conditions selon lesquelles deux événements étaient liés en mémoire dont celle portant sur le déroulement simultané des deux événements. De même, dans le cadre de l'étude de la mémoire sémantique, des expériences d'amorçage dans lesquelles on étudie par exemple l'effet de la présentation d'un mot sur le traitement d'un autre mot ont permis de montrer que le traitement du second mot est facilité lorsque celui-ci est précédé d'un mot associé (voir § 2.3.1.1).

Hebb (1949) a révélé l'existence de ce principe à un niveau neurophysiologique : "Quand un axone de la cellule A est assez proche pour exciter une cellule B et quand, de façon répétée et persistante, il participe à son activation, un certain processus de croissance ou un changement métabolique s'installe dans une cellule ou dans les deux, tel que l'efficacité de A, en tant qu'elle est une des cellules qui active B, est augmentée".

Par ailleurs, bien qu'une controverse existe à ce sujet, l'apprentissage automatique et précis des fréquences conditionnelles par des sujets placés dans certaines circonstances a été mis en évidence (Hasher et Zacks, 1984).

Enfin, l'oubli et les effets sériels de l'oubli, en particulier l'effet de récence sont aussi des phénomènes pour lesquels il existe une littérature abondante (Baddeley, 1992). Il existe également un effet dit de primauté dont nous ne rendons pas compte, les premières observations étant aussi mieux rappelées. Pour en rendre compte, une solution simple consiste à ne pas diminuer linéairement la plasticité de la synapse comme nous le faisons lorsque nous considérons $\alpha = \sum_{i=1}^{n-1} \mu_P(o_i)$, mais à la faire décroître plus rapidement pour les premières observations. Cependant, contrairement à l'oubli, l'intérêt de considérer ce phénomène pour l'extraction d'information pertinente n'est pas évident.

4.5.2.2 Principes liés à la construction des catégories

D'après Rosch (1976), la construction des catégories trouve sa justification dans les prédictions qu'elle permet. Or, c'est ce principe qui justifie que nous construisions la conjonction des A&B lorsque tous les A ne sont pas B, cette conjonction ne permettant pas de faire des prédictions différentes de A ou de B. Plus globalement, les conjonctions de propriétés construites respectent le compromis effet/effort énoncé par Rosch (1976). La conjonction est construite si elle apporte une nouvelle information, et ne l'est pas dans le cas contraire. De même que pour les conjonctions de propriétés, la construction des sous-ensembles flous se justifie du point de vue de la construction des catégories. Ainsi on construira la catégorie des grands pétales si cette construction nous permet de prédire d'autres propriétés et particulièrement celles en rapport avec nos objectifs.

4.5.2.3 La notion de filtre

Un filtre peut être matérialisé par une fonction définie sur un univers X et à valeur dans [0,1], exprimant pour chaque x appartenant à X, un degré de réponse. La notion de filtre renvoie à celle de sélection, des valeurs sont préférées (celles pour lesquelles la réponse est totale), d'autres sont plus ou moins prises en compte (celles pour lesquelles le degré de réponse est compris entre 0 et 1), et enfin certaines sont ignorées (celles pour lesquelles le degré de réponse est nulle). Le filtre est plus ou moins sélectif suivant l'évasement de sa fonction caractéristique (figure 4.18).

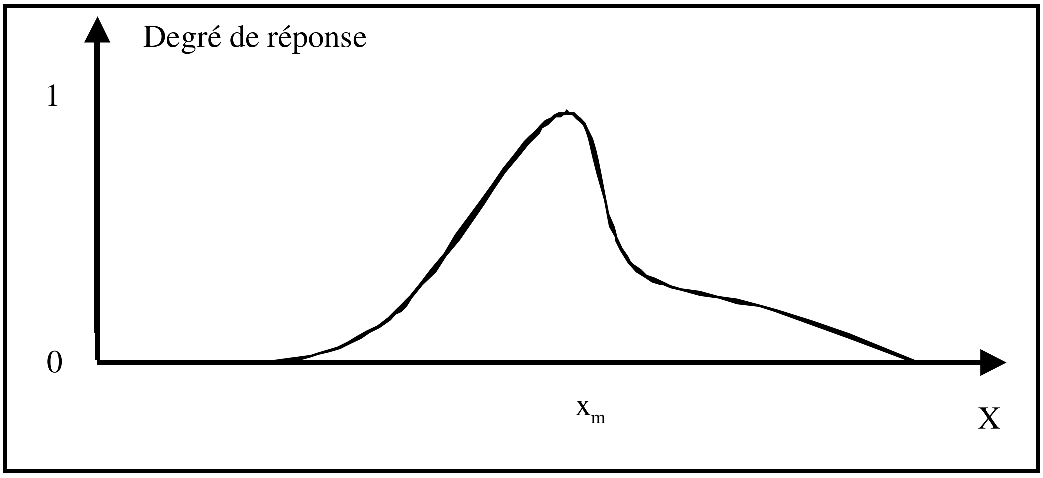


FIG. 4.18 Exemple de filtre. La réponse est maximale aux environs de x_m et décroît en s'éloignant de cette valeur.

La notion de filtre utilisée pour harmoniser les sous-ensembles flous, a aussi une correspondance au niveau cognitif. La réalité des filtres dans le système perceptif a été mise en évidence par deux types de méthodes. Un premier type correspond aux méthodes électrophysiologiques qui mesurent directement l'activité d'un neurone (on mesure la variation de sa fréquence de décharge) suivant le stimulus placé en entrée (Bagot, 1996). Il faut noter deux phénomènes différents suivant que l'on considère la valeur du stimulus (la fréquence de l'onde sonore si on considère la hauteur d'un son), ou son intensité (amplitude de l'onde sonore liée au volume perçu). Indépendamment de la valeur du stimulus, plus l'intensité est importante, plus le neurone répond (sa fréquence de décharge augmente). Indépendamment de l'intensité, plus la valeur se rapproche de la valeur à laquelle le neurone répond préférentiellement, plus le neurone répond. Dans ce dernier cas, les courbes obtenues que l'on appelle courbes d'accord du neurone (« neural tuning curves » en anglais) rendent donc compte de la spécificité neuronale. Elles sont obtenues en définissant pour chaque valeur du stimulus la valeur de l'intensité juste nécessaire à une variation de l'activité du neurone par rapport à son activité spontanée. Ces courbes d'accord indiquent une valeur minimale pour la valeur du stimulus préférentielle. Pour obtenir les filtres, il faut donc les inverser.

Les expériences de masquage constituent un second type de méthode permettant de déterminer des courbes d'accord (et donc des filtres). Dans ces expériences, on détermine la courbe d'accord pour une valeur centrale x du stimulus en mesurant l'intensité nécessaire de chaque stimulus de valeur y pour qu'il masque le stimulus de valeur x (dans le cas d'un son, on peut considérer qu'un son est masqué lorsqu'on ne le perçoit plus). Plus y est proche de x moins l'intensité nécessaire au masquage est importante. Il est remarquable de constater que l'on retrouve les mêmes courbes d'accord que celles obtenues par la méthode électrophysiologique (McAdams, 1994).

Ainsi, un stimulus correspond à une valeur particulière mais aussi, bien que dans une moindre mesure, aux valeur voisines avec un degré défini par le filtre centré sur la valeur du stimulus (figure 4.19). Par ailleurs, il existe un filtre différent pour chaque valeur du stimulus. Il faut donc plus globalement imaginer un ensemble de filtres distribués sur un axe représentant les différentes valeurs du stimulus (figure 4.19). On peut considérer qu'un stimulus correspondra à un filtre particulier mais aussi avec un degré moindre aux filtres voisins.

On peut penser que la présence des filtres dans le système perceptif est à l’origine de l’imprécision. Ainsi si l’on considère qu’au niveau neuronal les patterns d’activation correspondant à deux stimuli sont la base du jugement de similarité, ces patterns sont plus ou moins proches suivant la superposition des filtres correspondant respectivement aux deux stimuli.

L'identification des filtres avec les sous-ensembles flous est très tentante car la notion de degré d'appartenance semble assez proche de celle de degré de spécificité de réponse. Les filtres ont, d'un point de vue pratique, l'intérêt de permettre de lisser les sous-ensembles flous de façon à obtenir des découpages plus harmonieux. On peut aussi penser que les filtres jouent aussi pour le système cognitif ce rôle intégrateur.

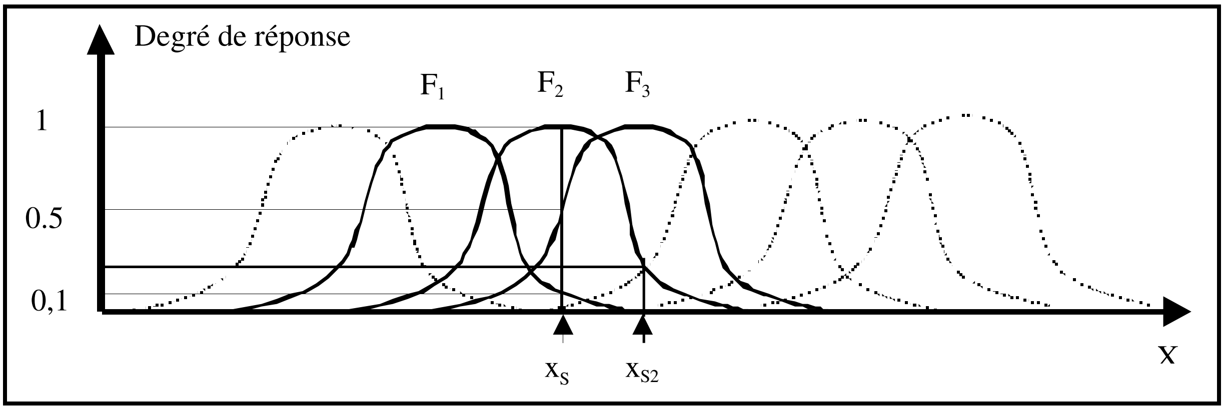


FIG. 4.19 Un stimulus correspond à une valeur particulière (x_s) mais aussi aux valeurs voisines (par exemple x_{s2}) suivant le filtre qui lui est associé (ici F_2). Un stimulus correspond à un filtre particulier (ici F_2), mais aussi dans une moindre mesure aux filtres voisins (ici F_1 avec un degré 0.1 et F_3 avec un degré 0.5).

4.5.2.4 Principes liés à l'exploitation du réseau sémantique

On notera que les méthodes de propagation d'activation sont fondées sur un neuro-mimétisme et qu'elles ont permis de rendre compte d'un certain nombre de phénomènes psychologiques (voir chapitre 2).

Par ailleurs, le principe de résonance a donné lieu au développement d'une théorie de la résonance (*l'Adaptative Resonance Theory*) (Carpenter & Grossberg, 1987) dans le cadre des réseaux de neurones. Or, il se trouve que cette théorie de forte inspiration neurophysiologique s'applique particulièrement bien à la modélisation de l'attention et au rappel d'information stockée en mémoire qui sont, comme nous l'avons vu, particulièrement liés à la notion de pertinence. Ce type de réseau fonctionne sur un principe de résonance entre deux couches de neurones. La première couche F_1, concerne la représentation des stimuli (neurones activés par la perception des stimuli) et la seconde couche F_2, concerne la représentation des traces en mémoire laissées par les précédentes observations. A la fois, l'apprentissage et l'exploitation du réseau repose sur l'oscillation de l'activation entre les deux couches. Ainsi, l'activation produite par le stimulus dans la couche F_1 se propage à la couche F_2 (représentant ici le traitement ascendant de l'information), l'activation de la couche F_2 se rétropropage ensuite (après compétition interne) à la couche F_1 (représentant ici le traitement descendant de l'information) et ainsi de suite jusqu'à ce que l'activation rétropropagée par F_2 s'apparie bien avec l'activation produite par le stimulus (ce qui caractérise un état de résonance). Le processus s'arrête donc lorsque l'activation Y de F_2 (réponse du système) est spécifique à l'activation X de F_1, soit

Y→X. Le degré d'appariement des deux activations exigé pour arrêter le processus est un paramètre du système qui représente un niveau de vigilance (le système est plus ou moins tolérant).

Chapitre 5

5 APPLICATIONS

Les trois applications que nous présentons dans ce chapitre permettent de mettre en avant différents aspects du système RELIEFS. Il s'agit dans les trois cas d'extraire l'information pertinente à partir d'un ensemble de données. La diversité des applications que nous présentons permet de mettre en avant la généralité de notre système.

La première application est la plus directe et la plus générale des trois. Il s'agit de l'utilisation de RELIEFS pour l'extraction d'information pertinente à partir d'une base de données provenant de l'Institut de Veille Sanitaire (InVS) et portant sur l'enregistrement de cas de méningite. Dans un premier temps, RELIEFS construit le réseau sémantique flou qui rend compte des régularités présentes dans les données. Ce réseau peut être ensuite interrogé pour extraire les informations liées à la requête.

La seconde application est plus particulièrement dédiée à la notion de filtre et au principe de diffusion. Elle montre en particulier son intérêt dans l'analyse de réponses concernant l'évaluation de produits alimentaires (des crèmes dessert) faites en utilisant des items verbaux comme « légèrement sucré », « compacte », « très claire », etc.

La dernière application porte sur la dimension « interactive » de la définition de la pertinence. Elle présente une expérience dans laquelle des sujets ont été mis en situation de communication. Dans cette situation, un sujet doit faire reproduire un dessin au moyen de commandes verbales à un autre sujet. Nous montrons qu'une nouvelle fois, l'information omise dans les commandes de dessin s'explique bien en termes de pertinence si nous la définissons comme une relation de spécificité. D'autre part, nous proposons un modèle de l'interprétation de commande dans lequel le réseau sémantique est intégré pour compléter l'information manquante et dans lequel l'aspect flou des représentations laisse la situation guider l'action.

5.1 Application de RELIEFS à des données épidémiologiques

Cette application se situe dans le cadre de la fouille de données. L'objectif est d'extraire des relations entre les informations présentes dans une base de données décrivant des cas de méningites en France sur une période de 12 ans (1985-1997) afin de caractériser la distribution de cette maladie dans le temps, géographiquement et en termes de facteurs individuels associés à sa survenue. Enfin, on souhaite aussi permettre à l'utilisateur d'interroger la base de données sur des aspects particuliers au moyen de requêtes.

5.1.1 Introduction

L'étude a été réalisée en collaboration avec l'Institut de veille Sanitaire(InVS) dont une des missions principales est la surveillance épidémiologique de la population française. Cette surveillance est rendue possible par le caractère obligatoire de la déclaration d'un certain nombre de maladies comme le choléra, la diphtérie, la légionellose, le paludisme , la tuberculose ou encore les infections à méningocoque.

Concernant les infections à méningocoque, les formes cliniques sont soit une méningite, soit une septicémie à méningocoque. Le critère de déclaration de cette maladie est l'isolement de la bactérie « Neisseria meningitidis » dans le liquide céphalo-rachidien (lcr) et/ou le sang, ou la présence d'antigènes solubles de « N. meningitidis » dans le lcr, le sang ou les urines. Il existe différents sérogroupes de méningocoques. Le sérogroupe B est dominant en France, vient ensuite le sérogroupe C, puis beaucoup plus rarement le sérogroupe A et exceptionnellement les sérogroupes X, Y, U et Z. Il existe une forme de méningite particulièrement foudroyante appelée « purpura fulminans » dont l'évolution est particulièrement grave (décès, séquelles cutanées ou neurologiques).

La démarche généralement suivie dans l'analyse de ces données est double. Elle consiste :

- dans un but d'alerte, à comparer les cas observés sur la période la plus récente aux observations historiques, et à tester les éventuels écarts avec les valeurs historiques par des tests statistiques pour différents niveaux géographiques.

- dans un but plus descriptif, à étudier la distribution de la maladie dans la population, à formuler des hypothèses qui sont validées par des tests statistiques.

La démarche suivie ici est tout autre ; il s'agit en effet de générer automatiquement des hypothèses qui pourront éventuellement ensuite faire l'objet de tests statistiques. Nous nous restreignons au calcul d'indices statistiques simples (fréquences conditionnelles) mais nous considérons l'ensemble des dépendances potentielles.

Enfin, contrairement à la recherche d'information ou à l'interrogation classique de base de données, il ne s'agit pas de sélectionner un ensemble de documents ou d'enregistrements mais plus généralement un ensemble d'informations pertinentes. Les travaux dans ce domaine intégrant l'imprécision sont assez rares. On peut néanmoins citer ceux de Bosc et ses collaborateurs (1998) qui proposent un algorithme pour la recherche de dépendances fonctionnelles dans les bases de données intégrant l'imprécision.

5.1.2 Description des données

Le fichier traité comprend les 4671 cas d'infections à méningocoques déclarés en France entre l'année 85 et l'année 97. Chaque cas est décrit à partir de 11 variables :

- l'année
- le département de déclaration
- le mois
- l'âge du patient
- le sérogroupe du méningocoque (A, B, C, X, Y, U, Z)
- la présence d'antigène soluble
- l'évolution (guérison, séquelles, décès)
- la présence du méningocoque dans le lcr
- le caractère foudroyant (pupura fulminans) de la maladie
- la présence du méningocoque dans le sang
- le sexe du malade

Recherchant des dépendances pour des zones géographiques ne respectant pas nécessairement les frontières administratives des départements, nous avons remplacé cette variable par une variable indiquant la longitude (x) et une variable indiquant la lattitude (y) en considérant une grille de 26 sur 26 recouvrant la totalité du territoire français (figure 5.1). Pour des raisons que nous spécifierons par la suite, nous avons de plus ajouté une variable indiquant la présence de la maladie (qui est par définition vérifiée pour toutes les observations puisqu'il s'agit de déclaration de cas). De plus, nous avons introduit certaines connaissances et objectifs dans la discrétisation que nous avons réalisée sur les différentes variables.

Pour l'année

Nous considérons pour les années une frontière floue assez réduite. Une observation compte pour 1 pour l'année où elle a été déclarée et compte aussi pour l'année précédente et l'année suivante avec un degré 0.5. L'objectif est tout de même d'extraire des périodes comprenant plusieurs années.

Pour l'âge

Nous avons introduit ici une discrétisation floue et un découpage plus fin pour les premiers âges exprimant ainsi qu'une même différence d'âge se traduit par une différence physiologique plus grande pour les enfants que pour des adultes.

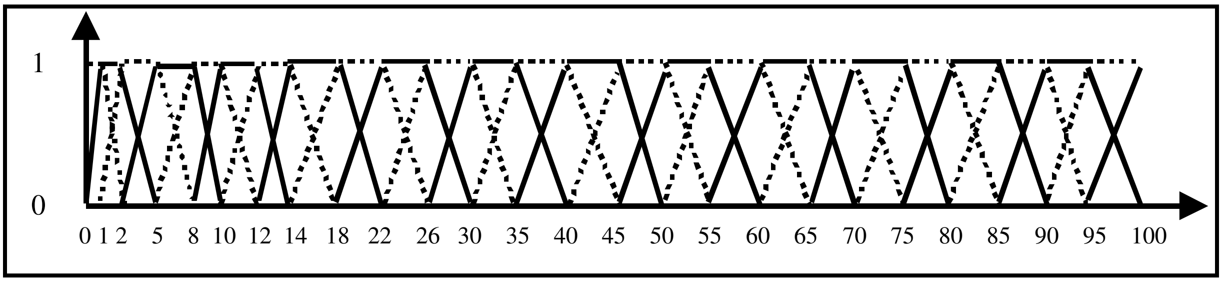


FIG 5.1 Définition des filtres pour la variable « âge ».

Pour le mois

Compte tenu du grand nombre d'observations comptant pour chaque mois et des régularités fortes que nous avions obtenues dans une expérience préliminaire, nous n'avons pas considéré de frontières floues.

Pour la zone géographique

Le découpage par les départements s'étant révélé trop fin dans une étude préliminaire, nous cherchions ici à extraire des informations sur des zones géographiques à l'échelle de la région. Nous avons donc considéré des filtres assez larges. Lorsqu'une observation survient dans un département nous la comptons en totalité pour la case correspondant au barycentre du département, pour 0.75 pour les cases adjacentes à cette case, puis pour 0.5 et pour 0.25 en s'éloignant (figure 5.2).

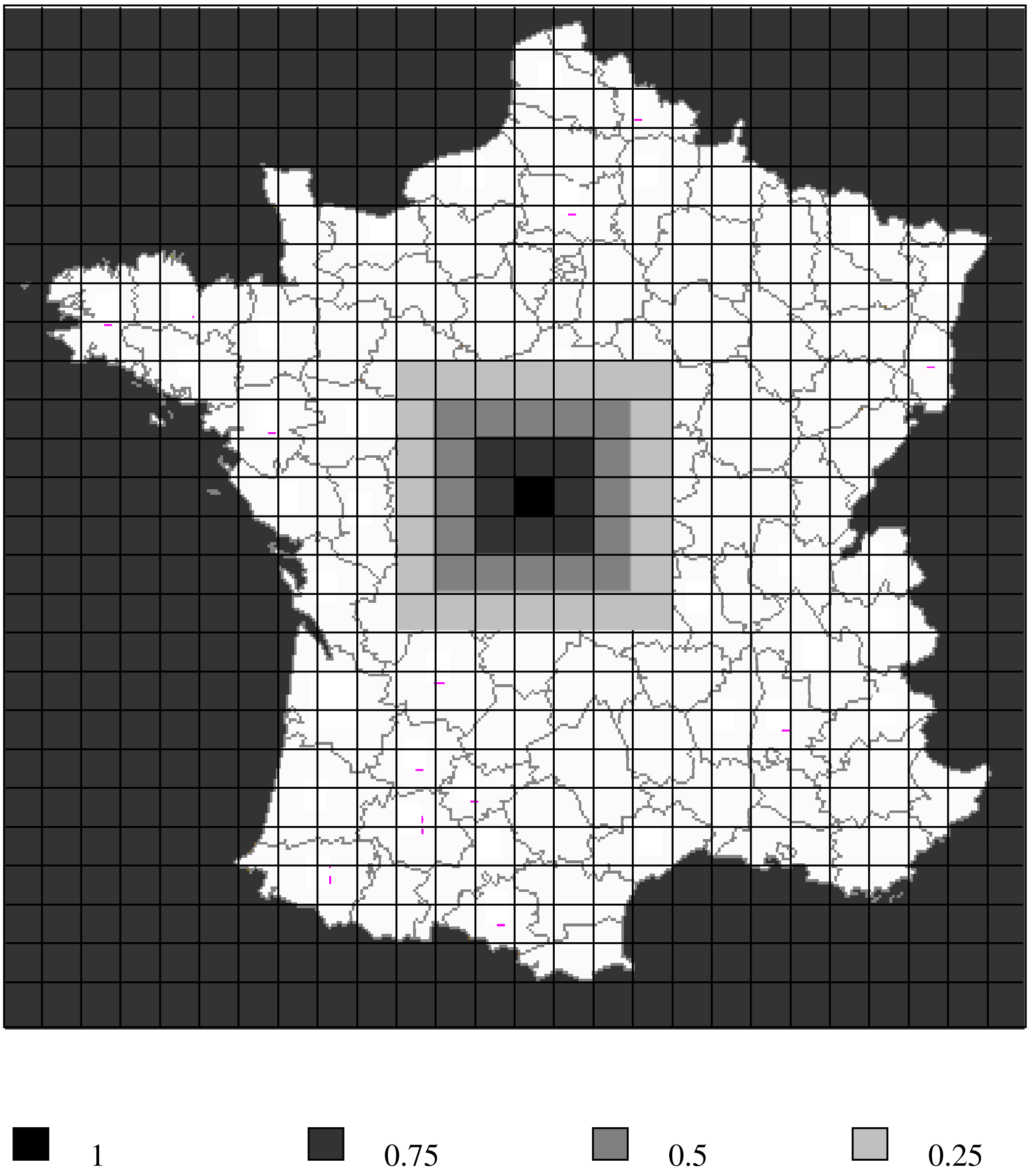


FIG 5.2 Les filtres considérés pour les zones géographiques.

5.1.3 Extraction de l'information pertinente

Cette étude se prolonge actuellement. Nous présentons ici les résultats portant sur la recherche d'associations simples entre propriétés, nous ne considérons donc pas les combinaisons de propriétés.

La variable ajoutée portant sur la présence de la maladie nous a permis dans un premier temps d'extraire ce qui était pertinent au sujet des infections à méningocoque. Compte tenu du fait que nous ne disposions pas d'observation correspondant à des cas où la maladie était absente et donc que toute propriété impliquait la maladie avec un degré 1, le calcul de la spécificité de la distribution des liens entrants ne permettait pas de discerner ce qui était pertinent de ce qui ne l'était pas. Ainsi, compte tenu de l'égalité de tous les liens entrants, nous avons calculé la spécificité de la distribution des liens sortants pour le nœud représentant la présence de la maladie.

De fortes spécificités ont été obtenues pour le sérogroupe (0.9), l'évolution (0.9), l'âge (0.75), les positions en X et Y (respectivement, 0.58 et 0.56), la forme « purpura fulminans » de la maladie (0.68), et enfin le mois (0.33) et l'année (0.32). On observe en effet au regard des distributions que :

- le sérogroupe B est particulièrement dominant,
- la guérison est l'évolution la plus fréquente,
- les enfants et les adolescents sont nettement les plus concernés (figure 5.3)
- la forme « purpura fulminans » de la maladie est relativement rare.
- les mois d'hiver et de printemps (figure 5.3) comptent le plus grand nombre de cas.
- un plus grand nombre de cas a été enregistré à la fin des années 80 et au début des années 90.

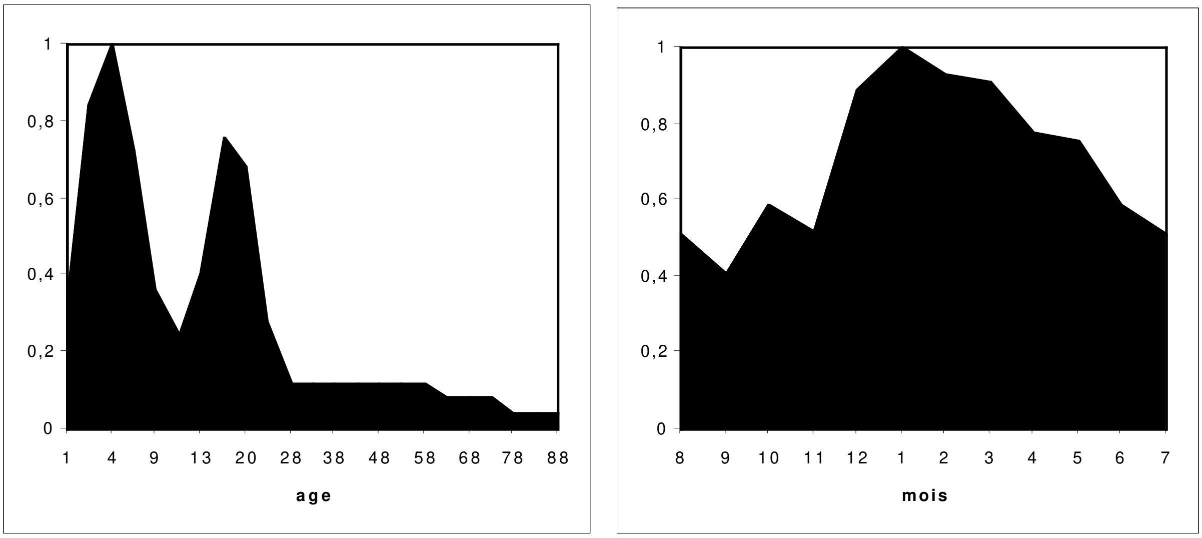


FIG. 5.3 Distribution normalisée des cas d'infection à méningocoques sur l'âge (la spécificité est 0.75) et sur les mois (la spécificité est de 0.33).

La spécificité enregistrée ici pour la position géographique n'apprend rien puisqu'elle suit la densité de la population.

Nous avons ensuite extrait les plus fortes spécificités concernant les liens entrant cette fois-ci pour l'ensemble des autres variables. Nous donnons une petite partie du réseau extrait dans la figure 5.4.

Ces résultats ont été l'objet d'une première analyse par les épidémiologistes de l'InVS. Ils y ont retrouvé des dépendances qu'ils connaissaient (l'ensemble des dépendances définies vis-à-vis des infections à méningocoques ont été confirmées) et ont découvert un certain nombre de dépendances qui n'avaient auparavant pas été détectées. Il s'agit selon une première analyse de données en marge qui concernent des propriétés assez rares (sérogroupes peu fréquents, par exemple), qui de ce fait n'apparaissaient pas lors d'une analyse ne considérant pas la relation de spécificité. Elles correspondent à des hypothèses qui n'ont pas été soulevées et testées par le passé.

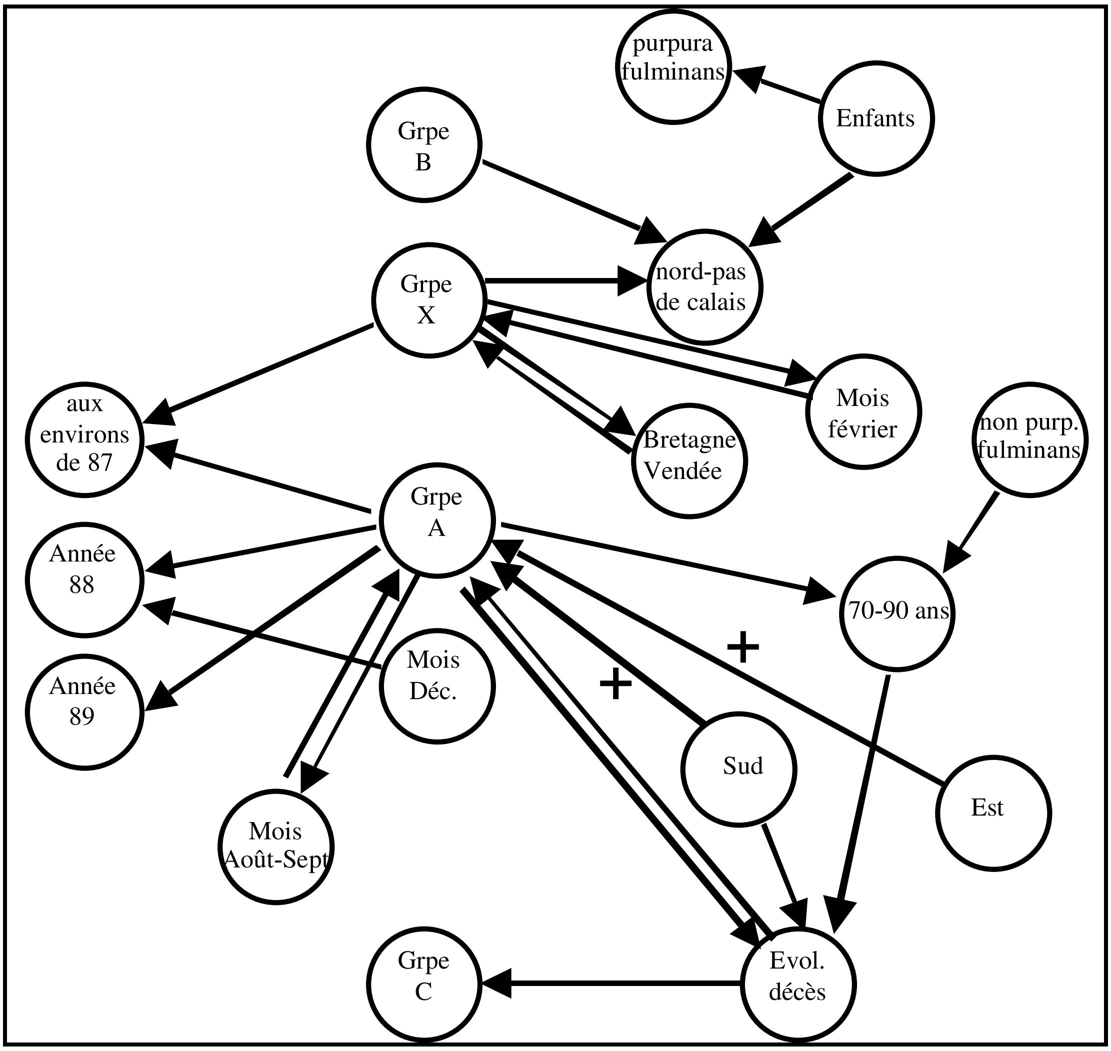


FIG. 5.4 Une partie du réseau extrait par l'analyse des associations simples.

Pour les sérogroupes, par exemple, on observe de fortes dépendances avec la zone géographique et le mois. De plus, certaines dépendances sont graduelles, ainsi plus on va vers l'est et vers le sud plus on retrouve le sérogroupe A. Ces relations graduelles peuvent être extraites automatiquement en considérant la monotonie de la distribution. On remarquera aussi qu'on ne fait pas figurer dans notre réseau les définitions de ce qu'on entend par exemple par « enfant » ou « aux environs de l'année 87 » mais que cette définition est bien construite automatiquement et peut être consultée.

Le système est interrogeable et à la requête portant par exemple sur l'année 88, sa réponse indique que, le sérogroupe A et le mois de décembre sont spécifiquement liés à cette année.

5.1.4 L'intérêt des filtres pour les zones géographiques

Le découpage initialement fait sur les départements ne permettait pas d'avoir une vue globale de la répartition géographique de la maladie. L'importante largeur de filtre que nous avons considérée a permis de mettre en avant un aspect particulièrement intéressant du principe de diffusion. Cet aspect concerne l'idée que deux zones proches comportant un nombre de cas important se renforcent mutuellement, ce qui permet d'extraire des régions qui globalement sont particulièrement touchées et non d'aplanir totalement la distribution comme le principe de diffusion pouvait le laisser supposer. Il pourrait également être intéressant d'automatiser la construction des filtres initiaux sur le principe du maximum de spécificité considérant un compromis entre des filtres trop petits qui ne permettent pas d'opérer des regroupements et des filtres trop importants qui font disparaître tout relief des distributions (figure 5.5).

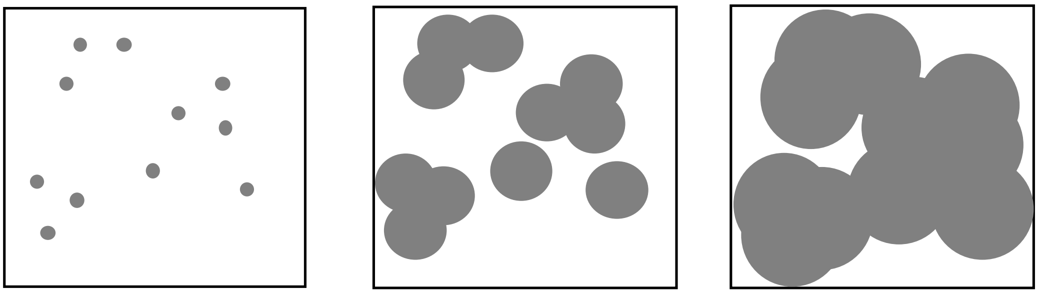


FIG. 5.5 Suivant la largeur des filtres considérée, des régularités apparaissent plus ou moins nettement.

Il a été possible, à partir des associations simples, d'extraire des zones géographiques particulièrement concernées en étudiant la spécificité des latitudes (Y) vis-à-vis des différentes longitudes (X). On a par exemple pu observer une forte spécificité des latitudes correspondant aux zones (6-14) vis-à-vis des longitudes correspondant aux zones (2-8). Ainsi, en suivant les plus fortes spécificités, lorsque l'on se déplace en longitude de l'ouest vers l'est, le déplacement en latitude suit les côtes atlantiques dessinant une zone allant des côtes sud de la Bretagne à la Vendée. On notera que trois autres zones ont été extraites, mais que parmi celles-ci ne figure pas la région parisienne qui enregistre pourtant le plus grand nombre de cas.

5.1.5 Prise en compte de la relation de transitivité

L'heuristique tenant compte de la transitivité des relations a été utilisée et a permis de supprimer un certain nombre de spécificités qui apparaissaient. Par exemple, la spécificité du groupe B vis-à-vis du Nord-Pas-de-calais (qui apparaît sur la figure 5.3) s'explique par la jeunesse de la population du nord celle-ci étant majoritairement touchée par le sérogroupe B. On supprime aussi la spécificité de la guérison vis-à-vis de l'année 85 par l'absence d'infection de forme « purpura fulminans » cette année là. Même si la plupart des suppressions proposées s'interprètent assez bien, certaines ne suivent pas des liens de causalité naturels. Par exemple, la spécificité de la région nord vis-à-vis de la forme « purpura fulminans » s'explique par l'importance du nombre de décès dans cette région qui entraîne la présence de la forme virulente de la maladie. Or, on peut penser que c'est plutôt la forme « purpura fulminans » qui entraîne le décès.

5.2 Une méthode pour l'analyse de données sensorielles basée sur la notion de filtre

L'imprécision est inhérente à la perception et à l'évaluation subjective. Dans cette partie, nous identifions d'abord deux formes d'imprécision présentes lorsque l'on recueille les jugements d'experts chargés d'évaluer des produits alimentaires (crème dessert). Nous montrons ensuite comment la notion de filtre permet de prendre en compte l'imprécision. Cette étude a été menée conjointement avec le laboratoire de psychologie cognitive de l'université Paris 8. Les expériences sur lesquelles s'appuie cette application ont été réalisées par Urdapilleta (1998).

5.2.1 Introduction

Le processus psychologique mis en œuvre dans l'évaluation du stimulus fait intervenir différentes déformations (figure 5.6).

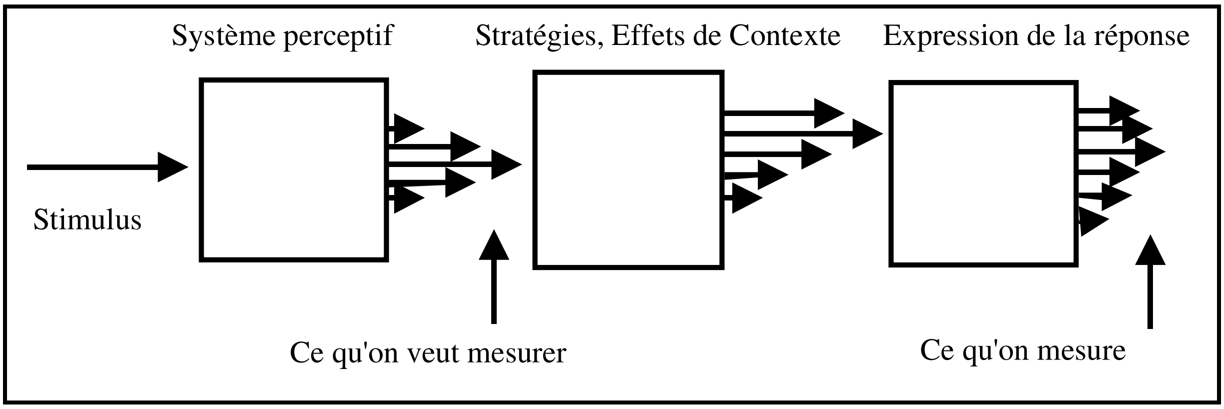


FIG 5.6 La chaîne des différentes déformations intervenant dans le recueil de réponses concernant l'évaluation d'un stimulus perçu.

En plus des biais liés à l'attitude du sujet et aux effets de contexte, il faut aussi considérer l'introduction d'imprécision à deux niveaux. Au niveau de la perception même, il existe des seuils de discrimination en deçà desquels deux stimuli deviennent difficilement distinguables. Ce seuil indique le degré de résolution ou encore la sensibilité du système perceptif vis-à-vis d'une variable particulière. Il existe différentes méthodes pour mesurer ces seuils (Guilford, 1954). La méthode la plus précise consiste à demander plusieurs fois à un sujet de comparer un stimulus S_i donné à un stimulus S_j, en demandant de choisir parmi deux ou trois valeurs, soit plus petit, soit plus grand ou soit, éventuellement, égal. On obtient ainsi une distribution de réponses telles que plus S_j est différent plus la proportion de réponses exactes est importante (figure 5.7).

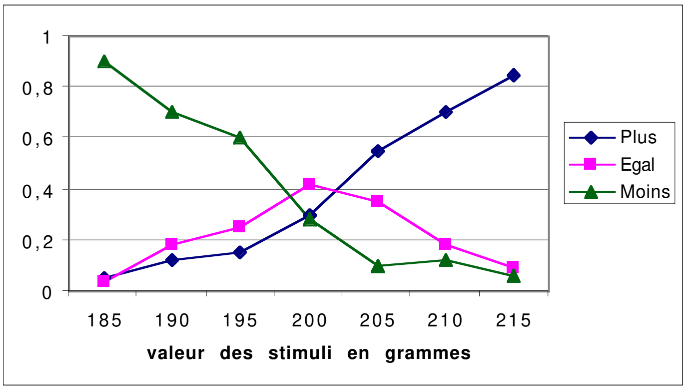


FIG. 5.7 Distribution des réponses « Moins », « Egal » et « Plus » dans une expérience portant sur l'évaluation d'une différence de poids avec un poids de référence de 200 g. Extrait de Guilford (1954, p. 136).

Il n'existe pas de seuil naturel à partir duquel on peut considérer que la différence entre les stimuli est faite, le passage d'une situation à l'autre est progressif. Pourtant, dans le but de définir un seuil différentiel, dans l'objectif plus général de définir une échelle des sensations graduée par ces seuils, la distribution ne peut être utilisée. Elle est remplacée par une valeur précise (correspondant généralement à la valeur pour laquelle la probabilité de donner une réponse correcte est de 0.75) et l'aspect graduel disparaît donc. Ainsi, même si la gradualité est le plus souvent évacuée, la zone de doute correspondant à la réponse « égal » pose cependant de véritables problèmes (Guilford, 1954, pp. 135-150). Une des questions porte sur la façon dont ces valeurs doivent être comptées (plus grand ou plus petit ?). Il n'y a pas de consensus sur la réponse et Guilford (1954, p. 140) préconise d'éviter dans la mesure du possible ce type de réponses.

La première source d'imprécision se situe au niveau de la perception du stimulus. Au niveau de la communication du jugement lorsqu'il correspond au choix d'un item dans une échelle, une autre source d'imprécision réside dans la représentation mentale de l'échelle par le sujet. On le comprend particulièrement bien lorsque ces items sont verbaux, la frontière par exemple entre « légèrement sucré » et « faiblement sucré » n'étant pas clairement définie. De plus, compte tenu du nombre limité d'items, ce choix, lorsqu'il doit être unique, peut ne pas correspondre parfaitement à un item de la liste mais se situer par exemple à la frontière entre deux items. Par ailleurs, on notera aussi que la distance entre les items n'est pas nécessairement constante comme l'ont montré différentes études. Ainsi

Cloninger et al. (1976) ont comparé les résultats de neuf études sensorielles utilisant des échelles de catégories comprenant entre 5 et 15 points. Ils constatent que la distance entre deux items successifs est plus réduite au centre de l'échelle que sur les bords.

Pourtant, comme le soulignent les auteurs, dans le cadre de profils sensoriels , les analyses statistiques de ce type de données sont basées sur l’hypothèse que les différents items sont espacés de façon régulière. Nous ajouterons de plus, qu’elles ne rendent pas compte de l’étalement de la réponse. Bien que la méthode des intervalles successifs (Guilford, 1954, chap. 10) permette de prendre en compte cet aspect variable de la largeur des empans, la réponse est toujours considérée comme précise (une valeur ou un intervalle de valeurs) et la perte de l'imprécision de la réponse et de son étalement peut être considérée comme une perte d'information. Notre objectif est ici de proposer une méthode d’analyse des données qui tienne compte de ces problèmes.

Nous décrivons dans un premier temps les expériences à partir desquelles les données que nous analysons ont été recueillies. Nous montrons ensuite comment la notion de filtre peut être utilisée pour rendre compte de l’imprécision des réponses.

5.2.2 Expériences

Nous résumons ici les expériences menées par Urdapilleta (1998), qui fournissent les données sur lesquelles notre méthode s’applique.

5.2.2.1 Expérience 1

La première expérience correspond à une tâche de cotation de crèmes dessert à l’aide d’attributs sensoriels. Elle met en comparaison deux types d’échelle. L’une est numérique (un ensemble de notes), l’autre est sémantique (un ensemble d’items verbaux).

Participants

Trente-trois personnes ont participé à l’expérience.

Matériel

Il est constitué de huit crèmes dessert au chocolat qui constituent un échantillon contrasté de l'espace produit. Quinze descripteurs ont été définis à partir d'une expérience préliminaire. Les descripteurs retenus sont les suivants :

– Concernant l'aspect des produits : *intensité de la couleur, brillance, uniformité de la couleur, granuleux, points blancs-bulles d'air.*

– concernant la texture : *fluide, collant, farineux.*

– concernant la flaveur : *goût de cacao, goût de crème-de lait, goût sucré, goût noisette, autre goût, goût défaut, persistance du goût de cacao*.

Procédure

Les trente-trois participants évaluent les huit produits avec les deux échelles à l'aide des 15 descripteurs. Deux groupes expérimentaux sont constitués pour contrebalancer les effets d'ordre dans l'utilisation échelles : un premier groupe utilise l'échelle sémantique dans un premier temps et l'échelle de catégories numériques dans un second temps, un second groupe procède de façon inverse. A raison de huit produits à évaluer, de trois répétitions par produit et de deux produits par séance, douze séances d'évaluation ont été réalisées.

L'évaluation des produits pour chacun des descripteurs se fait en plusieurs temps. Dans un premier temps, sur la première page, la consigne suivante est proposée au participant *"Un certain nombre de produits vont vous être proposés tour à tour, évaluez l'intensité perçue à l'aide de chaque descripteur en utilisant les modalités qui vous sont proposées."* Ensuite, sur la page suivante se trouve le numéro du produit, le nom du descripteur, sa définition et éventuellement sa modalité d'évaluation ainsi que l'énoncé suivant : *"l'intensité du descripteur est…"* (par exemple pour la couleur *"marron clair, marron moyen, marron foncé" ou l'échelle des notes*). Dans le cas des échelles sémantiques, sur la page suivante, le participant doit affiner sa réponse (par exemple pour la couleur, si le participant a répondu *"marron clair",* il doit préciser sa réponse : l'intensité de la couleur de la crème est *"très claire, claire"*. Le participant doit cocher l'une des deux modalités).

Notons que l'échelle de catégories numériques comporte huit catégories (de 0 à 7) pour tous les descripteurs, alors que l'échelle sémantique élaborée par les juges comporte, selon les descripteurs, six ou sept catégories.

5.2.2.2 Expérience 2

Cette expérience tente de mettre en correspondance les deux échelles (sémantique et numérique) en demandant aux participants d'apparier directement à une modalité sémantique une ou plusieurs notes.

Participants

Les participants sont vingt-cinq volontaires qui ont participé à l'expérimentation précédente.

Matériel

Il comprend un ensemble de questionnaires sur lesquels figurent, pour chaque descripteur, une modalité d'évaluation et une échelle de notes (de 0 à 7). Les participants peuvent donner plusieurs réponses.

Voici un exemple de questionnaire :

Concernant le *"goût de crème fraîche"* ou de *"lait",* dans cette crème, ce goût est *"moyennement crémeux",* c'est-à-dire plus précisément :

… léger

… ce qui correspond à une note ou une zone de notes de…

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

… marqué

… ce qui correspond à une note ou une zone de notes de…

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Procédure

Après avoir évalué les crèmes dessert avec l'échelle hiérarchique à items verbaux et l'échelle non hiérarchique numérique (cf. expérience précédente), les participants étaient invités à remplir un questionnaire sur lequel il leur était demandé de faire correspondre à une modalité sémantique une ou plusieurs notes.

Les huit modalités concernant chacun des quinze descripteurs étaient tirées au hasard et présentées après chaque séance d'évaluation. Douze séances ont été nécessaires pour que l'ensemble des modalités (soit 95) soit présenté aux participants.

Résultats

Les résultats montrent que tous les participants ont attribué des notes en utilisant l'ensemble des items de l'échelle de notes, mais chaque participant a attribué une plage de notes plus ou moins large pour chaque modalité.

Concernant l'empan, à une modalité sémantique correspondent 1 à 5 notes. Pour un descripteur, l'empan couvert par chaque modalité n'est pas régulier.

Prenons l'exemple du descripteur « collant » ; on observe que :

- pour la modalité « pas collante, fuyante », les notes vont de 0 à 2,
- pour la modalité « pas collante, glissante », les notes vont de 2 à 4,
- pour la modalité « moyennement collante , légèrement glissante », les notes vont de 2 à 5,
- pour la modalité « moyennement collante, légèrement collante » , les notes vont de 2 à 6,
- pour la modalité « collante, collante », les notes vont de 4 à 6,
- pour la modalité « collante, très collante », les notes vont de 6 à 7.

Concernant les frontières floues des modalités, on peut remarquer que pour chaque modalité, il existe une ou deux notes préférentielles mais que les notes voisines figurent aussi dans la distribution des réponses et ceci d'autant plus qu'elles sont proches des notes préférentielles. On notera aussi que les distributions se chevauchent largement (figure 5.8).

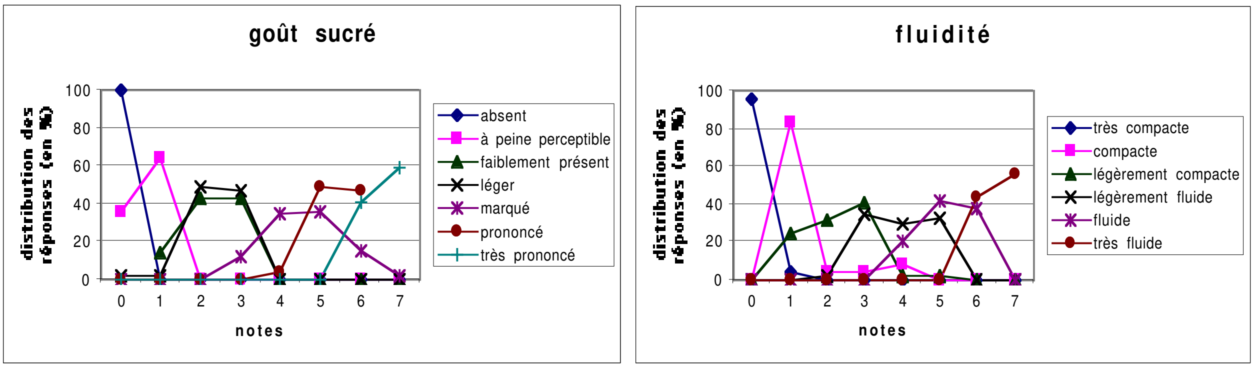


FIG. 5.8 Distributions des réponses des sujets sur l'échelle des notes pour différents items verbaux et pour le descripteur concernant le goût sucré et celui concernant la fluidité.

5.2.3 Application de la méthode de diffusion

5.2.3.1 Construction des filtres

Il est possible pour chaque couple d'items verbaux de mesurer l'intersection de leur distribution sur l'échelle des notes en considérant la proportion de réponses communes. En effectuant cette opération pour l'ensemble des couples d'items pour un même descripteur, on obtient un ensemble de filtres caractéristiques (représentables par des ensembles flous) de l'échelle sémantique particulière utilisée (figure 5.9). Ces filtres associés à chaque item peuvent être identifiés à des ensembles flous définis sur l'ensemble des items. Ainsi l'item « léger » est représenté par l'ensemble flou LEGER = 0,16/ « à peine perceptible » + 0,64/ « faiblement présent »+ 1/ « léger »+ 0,26/ « marqué » + 0,1/ « prononcé ».

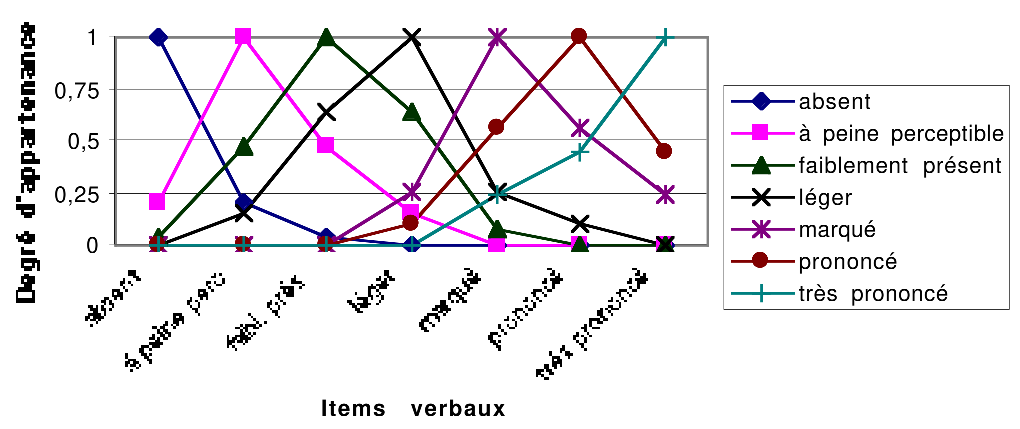


FIG. 5.9 Sous-ensembles flous décrivant les différents items de l'échelle sémantique utilisée pour l'évaluation du goût en fonction des autres items de cette même échelle. On peut par exemple observer que « absent » peut être décrit par « absent » avec un degré 1 par « à peine perceptible » avec un degré 0.2 et par « faiblement présent » avec un degré 0.05.

5.2.3.2 Modification des profils par diffusion

Nous avons appliqué le principe de diffusion aux données recueillies dans la première expérience pour modifier les profils des différents produits sur les différents descripteurs. Ainsi, on ne compte pas seulement l'item choisi comme réponse mais on distribue cette réponse sur les items voisins. Il faut cependant prendre garde à ce que chaque réponse compte de façon égale. Dans ce but, on divise chaque degré par la cardinalité de l'ensemble flou de façon à ce que la cardinalité de l'ensemble flou associé résultant soit égale à 1 (figure 5.10).

Par exemple l'item « léger » est initialement représenté par l'ensemble flou LEGER.

On a $|\text{LEGER}| = 0{,}16 + 0{,}64 + 1 + 0{,}26 + 0{,}1 = 2.16$

D'où la nouvelle définition de LEGER :

LEGER' = 0,07/ « à peine perceptible » + 0,3/ « faiblement présent »+ 0,46/ « léger »+ 0,12/ « marqué » + 0,05/ « prononcé ».

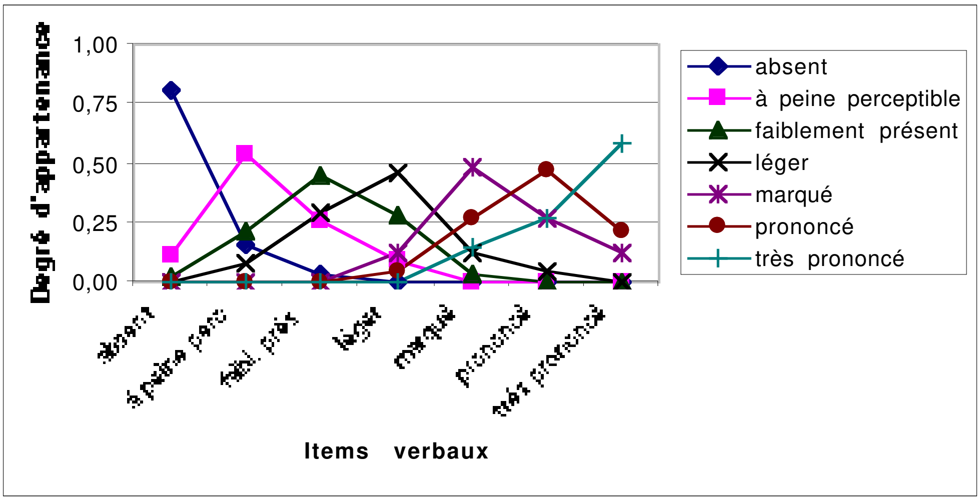


FIG. 5.10 Sous-ensembles flous de cardinalité égale à 1, décrivant les différents items de l'échelle sémantique utilisée pour l'évaluation du goût en fonction des autres items de cette même échelle.

Par principe de diffusion, en utilisant les filtres représentés dans la figure 5.10, la distribution des réponses des sujets est recalculée en tenant compte de l'imprécision des termes. Considérons par exemple, la distribution des réponses concernant le goût crème du produit P_2 (figure 5.11) qui est particulièrement accidentée. La réponse « absent » choisie dans 24% des cas, compte selon le principe de diffusion, pour 81% pour « absent », mais aussi pour 16% pour « à peine perceptible » et pour 3% pour « faiblement présent ».

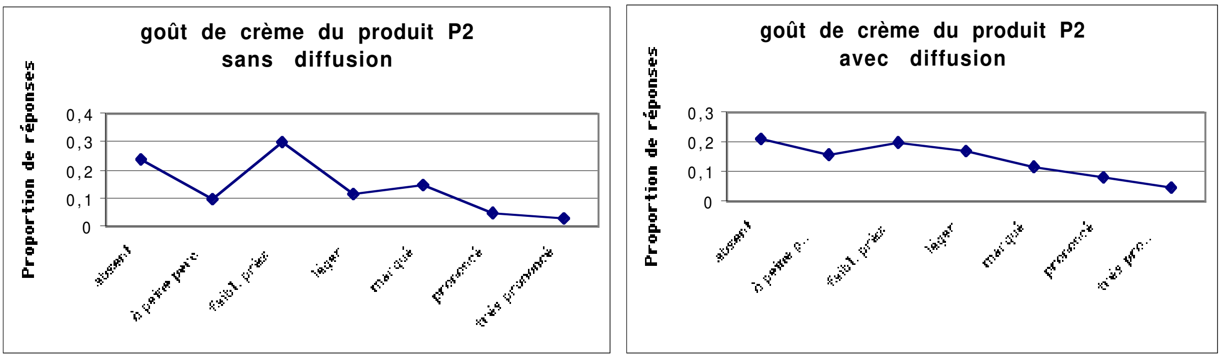


FIG. 5.11 Distributions sans et avec diffusion des réponses concernant le goût de crème du produit P_2.

Notons que la méthode de diffusion a ici pour effet global de lisser la courbe, et de fournir une distribution plus harmonieuse qui tiendra plus facilement lieu de synthèse (figure 5.11). Elle ne gomme cependant pas la bimodalité (« absent »/ « faiblement présent ») de la distribution que l'on peut aussi constater pour les réponses obtenues sur l'échelle des notes.

Sur cet exemple (figure 5.11), on peut en fait constater deux puits de réponses (c'est-à-dire des proportions de réponses inférieures à celles des items voisins à gauche et à droite). Le premier est observé pour l'item « à peine perceptible », le second, pour l'item « léger ». Seul le second est supprimé par diffusion et cela se justifie par le fait que les items « faiblement présent » et « léger » sont sémantiquement très proches, et qu'il est possible que l'item « faiblement présent » soit simplement plus facilement utilisé. Par contre le premier puit ne peut être comblé car l'item « absent » a une intersection très faible avec les items voisins. La méthode qui, rappelons-le, n'est pas artificielle puisqu'elle part de l'imprécision des termes définis par les juges eux-mêmes, permet donc d'avoir une vue plus synthétique et plus harmonieuse sans pour autant gommer toute différence. On peut ajouter qu'elle permet de rendre compte du cumul de réponses d'items voisins. Ainsi, même si individuellement des items voisins ne correspondent pas à des pics dans la distribution des réponses non corrigées, la méthode de diffusion peut permettre de faire ressortir l'item central qui profitera le plus de la diffusion provenant des différents voisins (figure 5.12).

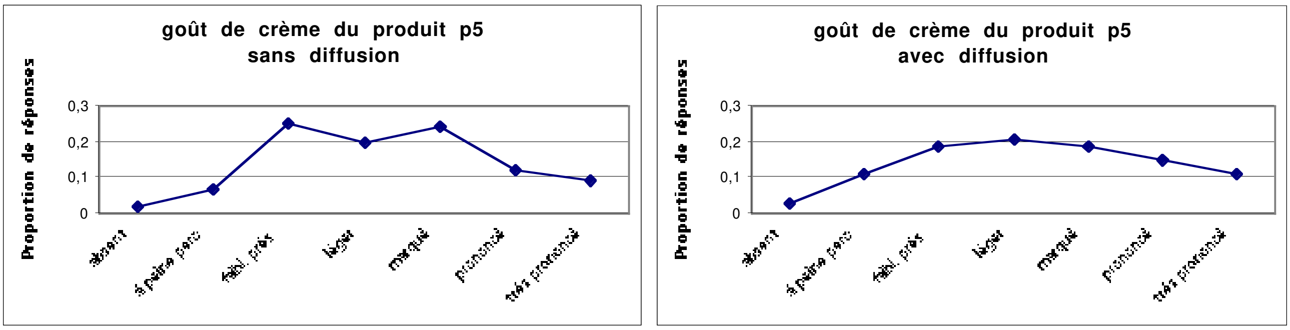


FIG. 5.12 Distributions sans et avec diffusion des réponses concernant le goût de crème du produit P_5.

D'un point de vue formel, les proportions de réponses obtenues avec diffusion correspondent aux calculs de cardinalité d'ensembles flous. Il est en effet possible de décrire chaque item par un ensemble flou défini sur l'univers des évaluations. Considérons trois évaluations une première (e_1) avec goût de crème absent, une seconde (e_2) avec goût de crème à peine perceptible (qui compte donc pour absent avec un degré de 0,16) et une troisième (e_3) avec goût de crème marqué (qui compte pour absent avec un degré 0). On a donc l'ensemble flou ABSENT = $0.81/e_1 + 0{,}16/e_2 + 0/e_3$. D'où $|\text{ABSENT}| = 0.97$. En rapportant cette cardinalité au nombre d'évaluations, on obtient la proportion obtenue par diffusion soit 0.97/3=0.32.

Plus généralement, si l'on désigne par e_i la i-ème évaluation, et par A, la caractérisation floue d'un item sur l'ensemble des évaluations, et que l'on considère n évaluations, la proportion calculée par diffusion est :

$$P(A) = \frac{|A|}{n} = \frac{\sum_{i=1}^{n} \mu_A(e_i)}{n}.$$

5.2.3.3 Réévaluation des implications entre propriétés

Par ailleurs, l'intérêt des échelles sémantiques, en plus de leur commodité d'utilisation, est de permettre à partir de l'analyse des résultats, de dégager des règles du type « SI goût de crème marqué ALORS goût de sucre faiblement présent » (on peut parler d'implication entre propriétés), plus naturellement qu'avec les échelles numériques.

De la même façon que précédemment, on considère l'ensemble flou décrivant chaque item sur l'univers des évaluations. Supposons trois évaluations e_1, e_2, e_3, chacune décrivant le goût de crème (gc) et le goût de sucre (gs) du produit. On considère cette fois-ci, les ensembles flous normalisés (ceux de la figure 5.9) qui n'ont pas nécessairement une cardinalité égale à 1 (la contrainte portant sur une prise en compte équivalente de chaque évaluation ne tient plus, l'évaluation e_2, étant par exemple moins utile à estimation de la règle « SI goût de crème marqué ALORS goût de sucre faiblement présent » que l'évaluation e_1)

e_1 : gc : « marqué » avec un degré 1 (la réponse est « marqué »)
gs : « faiblement présent » avec un degré 0.64 (la réponse est « léger »)

e_2 : gc : « marqué » avec un degré 0 (la réponse est « faiblement présent »)
gs : « faiblement présent » avec un degré 1 (la réponse est « faiblement présent »)

e_3 : gc : « marqué » avec un degré 0.57 (la réponse est « prononcé »)
gs : « faiblement présent » avec un degré 1 (la réponse est « faiblement présent »)

Considérons l'ensemble flou MARQUE = 1/e1 + 0/e2 +0.57/e3. On a donc $|$MARQUE$|$ = 1.57. Considérons l'ensemble flou (MARQUE $\cap$ FAIBLEMENT PRESENT) = 1*0.64/e1+0*1/e2+ 0.57*1/e3. (On notera que l'on prend ici le produit (qui est une t-norme) et non le min pour évaluer l'intersection).
On a donc $|$(MARQUE $\cap$ FAIBLEMENT PRESENT)$|$ = 1.21.

On se propose d'évaluer la règle « SI goût de crème marqué ALORS goût de sucre faiblement présent » par $|$(MARQUE $\cap$ FAIBLEMENT PRESENT)$|$ / $|$MARQUE$|$ soit 1.21/1.57 soit environ 0.77. Si l'on ne considère pas la diffusion, on obtient (1*0+0*1+0*1) / (1+0+0) = 0, ce qui intuitivement ne correspond pas à la réalité puisque deux évaluations sur trois (e_1 et e_3) ont plutôt tendance à confirmer la règle et que e_2 est sans effet. La méthode décrite permet donc d'extraire des implications entre propriétés visiblement valides qui n'auraient pas été détectées sans l'introduction du « flou ».

Plus généralement, si l'on désigne par e_i la i-ème évaluation, et par A et B la caractérisation floue de deux items sur l'ensemble des évaluations, et que l'on considère n évaluations, la mesure de l'implication que nous proposons est :

$$P(A/B) = \frac{|A \cap B|}{|A|} = \frac{\sum_{i=1}^{n} \mu_A(e_i).\mu_B(e_i)}{\sum_{i=1}^{n} \mu_A(e_i)}$$

A partir de l’ensemble des implications, il est ensuite possible d’extraire l’ensemble des informations pertinentes de la même façon que dans le cas des données épidémiologiques.

5.3 Application à l'interprétation de commandes verbales

Nous avons insisté sur le lien privilégié qui existe entre la notion de pertinence et celle de communication. Cette application porte précisément sur ce thème. Une expérience dans laquelle nous plaçons des sujets en situation de communication nous permet d'étudier les différents éléments qui entrent en jeu dans ce processus. En particulier, nous montrons qu'il est possible d'expliquer et de compléter l'information manquante (celle qui n'est pas transmise et qui est pourtant nécessaire à l'exécution de la tâche) en nous ramenant à la notion de pertinence. Nous proposons ensuite un modèle de l'interprétation de commande verbale intégrant un réseau sémantique flou dans le traitement de l'incomplétude et permettant plus globalement d'adapter la réponse du système à la situation. Cette étude a été réalisée conjointement avec le laboratoire de Psychologie de l'Université Paris 8 dans le cadre d'un contrat DRET.

5.3.1 Introduction

Les commandes verbales sont en apparence pauvres, ambiguës et elliptiques. Néanmoins, elles apparaissent très efficaces quand on les compare à la proximité de la réponse de l'opérateur (celui qui exécute la commande) vis-à-vis de l'objectif de l'instructeur (celui qui formule la commande). En résumé, quelques mots suffisent à la réalisation précise d'actions complexes. On peut donc s'interroger sur la puissance des énoncés verbaux.

Une explication partielle repose sur le fait que l'opérateur dispose d'un certain nombre de modèles de situation, de scénarios et de procédures. Cette connaissance générale lui permet de compléter l'information reçue et d'activer d'autres connaissances pour finalement comprendre ce qui lui est demandé et exécuter l'action appropriée. Quand, par exemple, il nous est demandé de poster une lettre, nous savons que cette lettre doit être timbrée, avoir une adresse, être mise dans une boîte aux lettres ou portée à la poste.

Des explications de nature pragmatique peuvent aussi apporter des éléments de réponse à la question portant sur l'efficacité des énoncés verbaux. La théorie de la pertinence de Sperber et Wilson (1989) et les maximes de Grice (1975) sont des explications de cet ordre. Ainsi, dans l'exemple précédent, si aucune information n'est donnée au sujet du type d'envoi dont il s'agit (tarif lent, normal ou express), l'opérateur pourra avec raison supposer que la lettre doit être timbrée au tarif normal ; parce que si elle devait être envoyée au tarif express cette information très pertinente (dans le sens où elle remet en cause nos croyances si l'on considère que le tarif classique correspond à la croyance par défaut) aurait très certainement été fournie.

L'approche « action située » fournit une façon plus circonstancielle d'expliquer la puissance des commandes verbales. Les représentants de ce courant de pensée accordent en effet moins d'importance aux représentations internes et plus à l'action et aux indices présents dans la situation. C'est cette approche que nous avons adoptée ici : la puissance du langage réside dans sa relation avec la situation.

5.3.2 Expérience

5.3.2.1 Objectifs

Le but de cette expérience était de fournir des données empiriques sur le degré de précision avec lequel il est possible d'interpréter des commandes verbales portant sur le dessin de figures géométriques simples. Elle s'intéresse en particulier, à mettre en évidence les conditions dans lesquelles l'information communiquée est incomplète ainsi que la façon dont ces informations sont complétées. Plus généralement, elle a pour objectif d'étudier comment une situation concrète influence la précision avec laquelle la commande est exécutée.

5.3.2.1 Méthode

Participants

Trente cinq étudiants de l'Université Paris 8 ont joué le rôle d'instructeur. Un étudiant provenant de la même population a joué le rôle d'opérateur pour l'ensemble de l'expérience.

Matériel

Un ensemble de 35 dessins (8.2 cm de large et 13.2 cm de haut), un pour chaque instructeur, a été créé. Chaque dessin est composé de trois figures géométriques simples. L'ensemble des dessins a été conçu de façon à fournir une grande variété de combinaison de propriétés pour les figures géométriques. Les différentes propriétés des figures sont : rectangle, cercle et carré, pour la forme; rouge, vert et bleu, pour la couleur; petit, moyen et grand pour la taille (de 1 cm jusqu'à 6.2 cm pour la largeur, de 0.6 cm jusqu'à 6.46 cm pour la hauteur); en haut, au centre et en bas pour la position verticale (de 0.01 cm jusqu'à 11.65 cm pour la coordonnée Y) ; et enfin, gauche, milieu et droite pour la position horizontale (de 0.31 cm jusqu'à 5.98 cm pour la coordonnée X). La complexité des combinaisons, du point de vue du treillis de Galois est maximale dans tous les cas. En d'autres termes, toute paire de figures a des propriétés en commun et des propriétés différentes (figure 5.13). Nous avons retenu une structure maximale dans le but d'avoir les différents ordres de généralité des propriétés représentés à chaque fois, notre hypothèse étant de vérifier si la généralité des propriétés jouait un rôle dans la formulation ou l'interprétation des commandes. Et nous avons fait en sorte de construire un ensemble de planches où la position des propriétés dans le treillis variait.

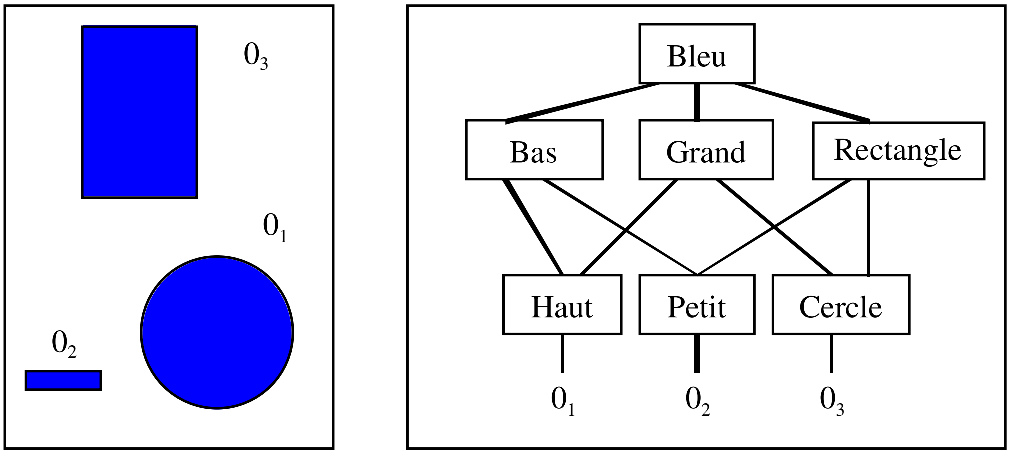


FIG. 5.13 Un exemple de dessin et la structure de propriétés correspondante (treillis de Galois).

Dispositif informatique

Le dispositif informatique comprend deux grands moniteurs placés dos-à-dos sur une longue table (figure 5.14). Ainsi, l'instructeur et l'opérateur, chacun derrière un moniteur sont cachés l'un à l'autre. L'instructeur peut seulement communiquer au moyen de commandes verbales. L'opérateur ne peut pas voir le dessin original que l'instructeur à en main.

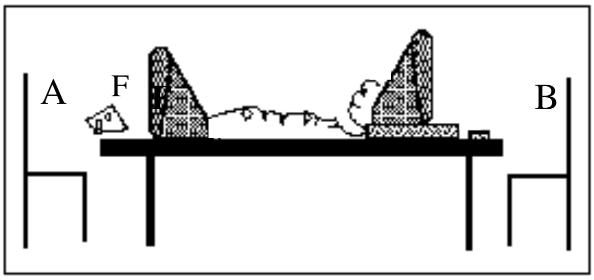


FIG. 5.14 L'instructeur est placé en A et tient dans ses mains un dessin (F). L'opérateur est placé en B.

Procédure Générale

Chacun des trente-cinq dessins est donné à un instructeur. L'instructeur doit faire reproduire ce dessin au moyen de commandes verbales uniquement. L'opérateur auquel il n'est pas permis de parler écrit à l'aide d'un traitement de texte chaque commande qu'il reçoit et l'exécute ensuite. L'interface graphique sur laquelle l'opérateur travaille a la même taille que l'image de l'instructeur. Sur son écran, l'instructeur voit ce que l'opérateur dessine. Quand l'opérateur a fini d'interpréter la commande, l'instructeur peut faire corriger le dessin au moyen d'une nouvelle commande, etc., jusqu'à ce que l'instructeur soit satisfait du dessin réalisé par l'opérateur.

Recueil automatisé des données

Toutes les opérations d'écriture des commandes verbales (sur le traitement de texte) ainsi que celles de dessin (à partir de l'interface graphique) ont été enregistrées avec un logiciel espion, tant en ce qui concerne le moment précis où l'action est réalisée qu'en ce qui concerne le déplacement du pointeur de dessin sur l'écran.

5.3.3 Résultats et discussion

5.3.3.1 Résultats empiriques

En moyenne, 9 commandes ont été nécessaires pour une reproduction satisfaisante du dessin original. Le minimum est 4 et le maximum 18. En résumé, 3 commandes sont requises pour la reproduction de chaque figure géométriques, dont deux qui corrigent le premier essai.

5.3.3.2 Précision des commandes

Pour estimer la précision avec laquelle la commande était satisfaite pour les propriétés continues, nous avons retenu pour les positions la différence en centimètres entre la réalisation du sujet et la réelle position de la figure dans la planche. Pour la taille (Largeur et Hauteur), il était plus pertinent d'un point de vue perceptif selon la loi de Weber (Guilford, 1954) de considérer une différence relative.

X: 0.53 cm Y : 0.72 cm L: 19 % H: 22.2 %

TAB. 5.1 L'erreur moyenne pour les commandes de dessin.

X: 0.20 cm Y : 0.36 cm L: 9.2 % H: 12.4 %

TAB. 5.2 L'erreur moyenne finale.

Il est donc clair que l'opérateur reproduit avec succès le dessin original et ceci assez rapidement. On peut considérer que les différentes précisions du résultat final nous indiquent des différences perceptives très faibles puisque l'instructeur est satisfait d'une telle précision. De ce point de vue, la précision obtenue initialement qui correspond à deux fois cet écart perceptif pour toutes les dimensions peut être considérée comme relativement bonne.

5.3.3.3 Au sujet des commandes incomplètes

Les informations sur la taille, la couleur et la position de la figure n'ont pas été toujours communiquées (figure 5.15). On compte en fait 49 omissions sur l'ensemble des 105 (35*3) commandes de dessin. Plus précisément, 24 concernent la taille, 6 la couleur, 11 la position verticale et 8 la position horizontale. Du point de vue de l'opérateur, cette omission correspond dans 26 cas à une propriété partagée par les figures déjà dessinées (par exemple, toutes les figures déjà dessinées sont grandes) et parmi ces 26 cas, dans précisément 20 cas, il s'agit de la propriété partagée par toutes les figures (déjà dessinées ou non). Dans les autres cas, l'omission correspond dans 10 cas à une valeur moyenne (la figure est placée au milieu) et dans 4 cas à des contraintes de l'environnement (seule une petite figure peut être réalisée à l'endroit spécifiée). Elle n'a pas une origine évidente dans 9 cas (voir tableau 5.3).

On peut considérer ici que la situation expérimentale pousse plutôt l'instructeur à être exhaustif et compte tenu du peu d'omission dont nous disposons, il est difficile de conclure de façon générale quand à l'origine de ces omissions. Néanmoins, les résultats laissent clairement penser que les régularités ont tendance à ne pas être communiquées. Ainsi, 40% des propriétés omises sont des propriétés qui apparaissent au plus haut niveau de généralité dans la structure des propriétés ce qui est largement supérieur au hasard si l'on considère que dans chaque dessin, il n'y a par définition qu'une seule propriété partagée par toutes les figures, alors qu'il y a 3 propriétés partagées par deux figures et plus de 3 qui soient spécifiques.

S28

- Dans le coin en bas à droite, vous avez un grand cercle rouge
- à sa gauche, vous avez un petit rectangle plutôt allongé
- remontez le un petit peu et il est aussi plus à gauche
- il est plus à gauche, plus bas et aussi plus gros
- un peu moins gros et moins vers la gauche
- en haut, plutôt vers la gauche, à mi-hauteur, vous avez un petit cercle rouge
- un peu plus grand et un tout petit peu plus bas

FIG. 5.15 Les sept commandes verbales pour faire dessiner les 3 figures de la planche 28. Pour la deuxième figure à dessiner (un petit rectangle), 4 propositions sont nécessaires. Mais aucune d'entre elles n'indique la couleur du petit rectangle, auquel l'opérateur a attribué la couleur rouge. Ce qui est le cas dans la planche originale.

Par ailleurs, il y a autant d'omissions qu'il s'agisse de la première figure réalisée ou de la dernière. Cependant on constate pour les secondes et dernières figures, que l'explication de l'omission est une régularité dans 62% des cas alors qu'elle n'est une régularité que dans 33% des cas pour la première (tableau 5.4). Ce qui peut laisser penser que les omissions liées à des régularités surviennent principalement dans un environnement partagé par l'instructeur et l'opérateur. D'autre part, on ne constate véritablement qu'une seule propriété omise apparaissant au niveau spécialisé (et qui ne soit pas une valeur moyenne).

	R	M	E	_	TOTAUX
TAILLE :	14	4	1	5	24
COULEUR :	5	0	0	1	6
POSITION VERTICALE :	4	5	1	1	11
POSITION HORIZONTALE :	3	1	2	2	8
TOTAUX	26	10	4	9	49

TAB. 5.3 Répartition des omissions en fonction de la variable considérée (taille, couleur, position) et du type d'omission (R : Régularité, M: valeur moyenne, E : contraintes d'environnement, _ : Pas d'explication).

	R	M	E	_	TOTAUX
FIGURE 1	5	4	0	6	15
FIGURE 2	14	3	3	1	21
FIGURE 3	7	3	1	2	13

TAB. 5.4 Répartition des omissions par figure (1, 2 et 3) et par type d'omission (R : Régularité, M: valeur moyenne, E : contraintes d'environnement, _ : Pas d'explication).

On peut constater que les valeurs moyennes constituent des valeurs par défaut. Les explications R et E ont un sens commun : ce qui peut se déduire de la situation est plus facilement omis. Les explications R, M et E ont un sens commun : ce qui ne constitue pas une remise en cause des croyances (valeurs par défaut, ou valeur des autres figures ou plus globalement ce qui peut se déduire de la situation) n'est pas communiqué. On rejoint ici la notion de pertinence définie par Sperber et Wilson (1989) selon laquelle plus une information comporte d'effet plus cette information est pertinente (à effort égal) et plus cette information a tendance à être communiquée. Et donc, inversement, moins l'information comporte d'effet moins l'information est pertinente, plus elle a tendance à être omise.

Or, il est intéressant de vérifier qu'ici les informations omises donc censées être les informations les moins pertinentes selon le point de vue de Sperber et Wilson (1989) sont, dans le cas des explications de type R et M, les informations les moins spécifiques puisque partagées par l'ensemble des figures. La pertinence s'identifie donc une nouvelle fois naturellement avec une relation de spécificité. Plus précisément, il s'agit de la spécificité de la propriété vis-à-vis de la figure à dessiner.

Du point de vue de l'opérateur, les omissions sont globalement bien complétées, bien que la précision soit moins grande comme en témoignent les résultats reportés dans le tableau 5.5 auxquels il faut ajouter que 5 fois sur les six omissions de couleur, cette dernière a été bien complétée.

X: 0.69 cm	Y : 1.22 cm	L: 23.4 %	H: 26.3 %

TAB. 5.5 Précision dans le cas des commandes incomplètes

Il n'y a pas de cas dans lesquels la réalisation est à l'opposé de ce que l'opérateur souhaitait et la tendance est toujours la bonne. Les plus grands écarts entre la figure réalisée et la figure à réaliser apparaissent dans les cas où l'omission n'a pas d'explication évidente ou lorsque les contraintes liées à l'environnement ne sont pas assez fortes. Il semble aussi que l'opérateur hésite entre une valeur moyenne et une valeur inférée à partir de la structure des propriétés des figures déjà réalisées.

Cette expérience fonde les principes d'un modèle pour l'interprétation de commandes incomplètes et imprécises, en suggérant une utilisation particulière du contexte, c'est-à-dire la prise en compte de la structure de propriétés des objets pour l'incomplétude, les contraintes de l'environnement (espace disponible) et les précédentes commandes pour la précision.

5.3.4 Un modèle de l'interprétation de commandes verbales : le système SIROCO

5.3.4.1 Description générale

Pour SIROCO (Système d'Interprétation de Requête par Optimisation de COntraintes), l'interprétation de commandes verbales consiste à construire le sens d'un énoncé à partir de l'énoncé proprement dit mais aussi et surtout à partir de la situation sur laquelle il s'appuie, cette dernière permettant de lever ses ambiguïtés et ses imprécisions. Ainsi, quand une ou plusieurs propriétés ne sont pas indiquées explicitement, celles-ci sont inférées à partir du réseau de propriétés des figures qui ont déjà été réalisées. Par ailleurs, souvent, les catégories sont spécifiées au moyen de caractérisations simples comme « grand », « en haut » mais parfois, il s'agit d'énoncés composés du type « plutôt grand » ou « plus haut ». Dans ce cas, le sens doit être construit à partir du sens des adjectifs simples ou de la valeur à modifier. Les indications et corrections données dans les précédentes commandes (l'historique) doivent aussi être prises en compte. Enfin toute cette information peut être représentée sous la forme d'un ensemble de contraintes floues et une procédure d'optimisation permet de déterminer la ou les meilleures solutions en tenant compte des contraintes de l'environnement par rapport à l'action (ici la place disponible pour réaliser le dessin) (figure 5.16).

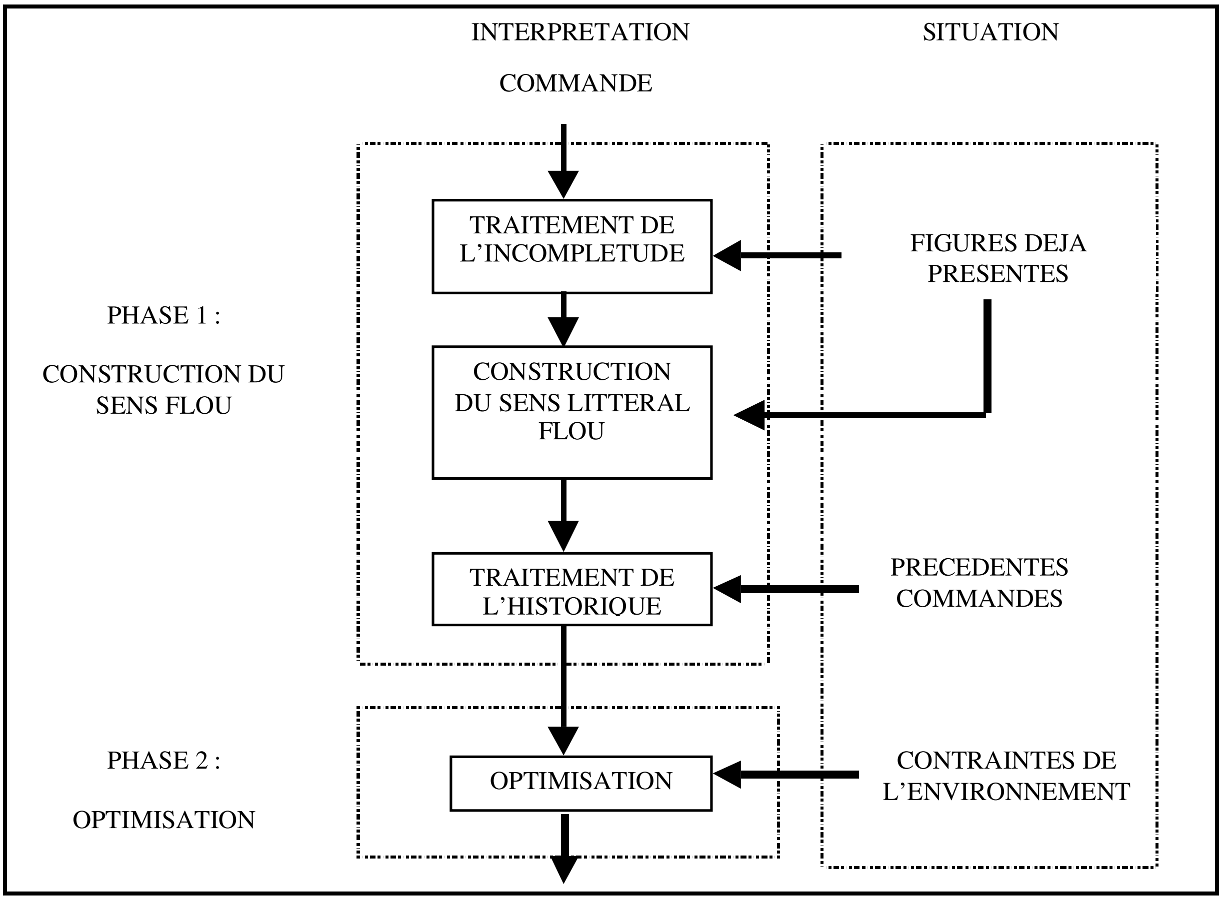


FIG. 5.16 Architecture générale du système SIROCO. La situation intervient à différents stades de l'interprétation.

5.3.4.2 Analyse des différentes composantes

5.3.4.2.1 Traitement de l'incomplétude à partir de réseaux sémantiques

La construction du réseau sémantique est réalisée au fur et à mesure de l'arrivée des nouveaux objets. L'exemple suivant reprend la construction du réseau sémantique pour les deux premières figures de la planche 11 (respectivement les figures 5.17 et 5.18). Pour des raisons de lisibilité, nous ne faisons figurer qu'une partie du réseau pour la prise en compte de la seconde figure. Seuls les liens valués à 1 apparaissent et les nœuds du réseau correspondant à des conjonctions de propriétés ne sont pas représentés.

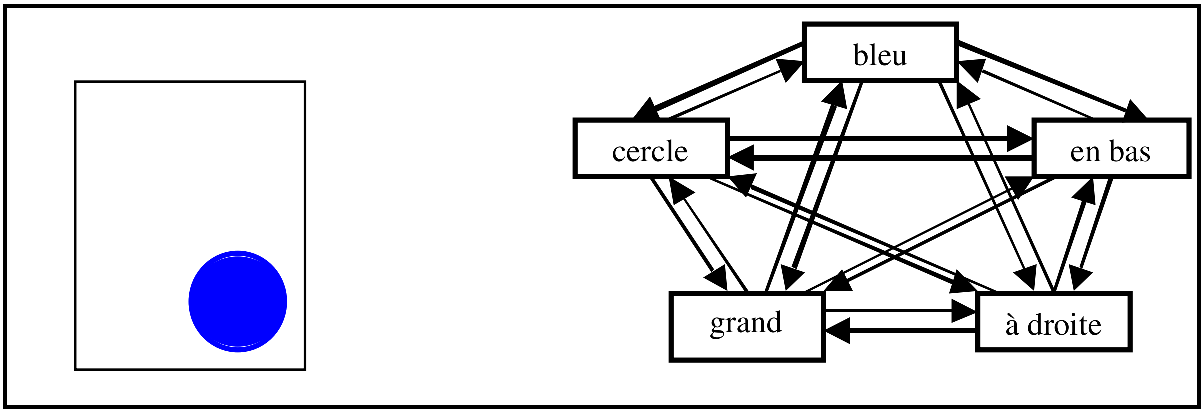


FIG. 5.17 Dessin et réseau correspondant à la commande « un grand cercle bleu en bas à droite ». Comme on peut le constater le graphe est complet. Il n' y a qu'une observation "un grand cercle bleu en bas à droite" et par conséquent tous les cercles sont bleus, tous les bleus sont cercles, tous les cercles sont grands, etc.

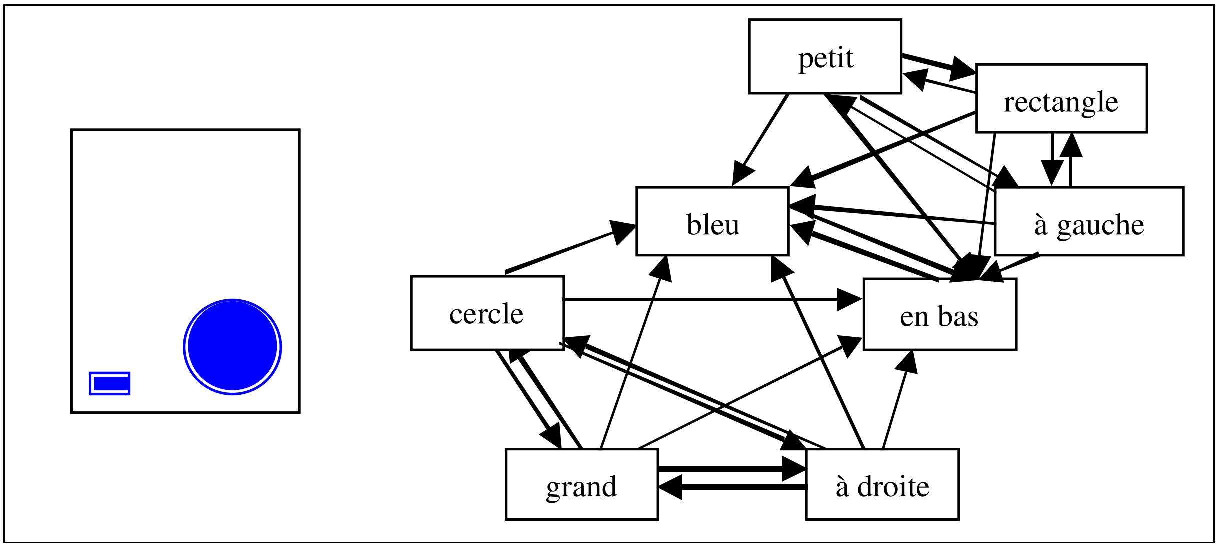


FIG. 5.18 Dessin et réseau correspondant à la commande « un grand cercle en bas à droite » suivie de la commande « un petit rectangle bleu en bas à gauche droite ».

Conformément aux résultats de l'expérience, les régularités extraites sont utilisées pour compléter l'information manquante. Si dans cette situation un nouveau cercle doit être réalisé sans aucune précision quant à sa couleur, on choisira de le réaliser de la même couleur que le précédent, considérant que l'information partagée (moins pertinente) a été omise. Le réseau est donc exploité en considérant des implications simples ("cercle→bleu").

5.3.4.2.2 Représenter une commande par des sous-ensembles flous

Une commande de dessin spécifie une taille, une forme, une couleur et une position. Mise à part la couleur qui est précisément définie (il y a juste un type de bleu, de vert et de rouge), les autres catégories (par exemple, grand, rectangle) ont des frontières imprécises. Ainsi, un élément (comme une valeur correspondant à une surface en centimètres carrés) peut avoir un degré d'appartenance intermédiaire à une catégorie. Nous avons donc choisi de représenter ces catégories par un sous-ensemble flou.

Un problème important réside dans le choix de la variable de référence sur laquelle on définit les sous-ensembles flous. Pour la taille, par exemple, on peut naturellement penser à la surface de la figure en centimètres carrés. Ce choix doit être fait de telle façon que la variable soit bien adaptée pour déterminer l'appartenance d'un élément à une catégorie. Idéalement, cette variable correspond à une dimension perceptive psychologiquement pertinente. La psychophysique qui étudie les relations entre les dimensions physiques et perceptives fournit ce type de variable. Ici, nous focalisant sur des principes généraux, nous avons choisi d'utiliser des variables simples (comme la surface pour la taille et l'abscisse pour la position horizontale) et des sous-ensembles flous trapézoïdaux.

Nous représentons donc une commande en associant un ensemble flou à chaque dimension de la description. Zadeh (1975) a introduit le concept de variable linguistique qui consiste en un triplet composé d'une variable (la taille représentée par la surface), un univers (l'ensemble des réels par exemple) et un ensemble de sous-ensembles flous représentant les différentes caractérisations de cette variable (par exemple petit, grand, moyen pour la taille).

Ici, nous utilisons quatre variables linguistiques pour représenter une commande : (i) la taille qui est l'aire de la figure et qui est caractérisée par « petit », « moyen » et « grand », (ii) l'aspect vertical ou horizontal de la figure qui est le rapport largeur/hauteur et qui est caractérisé par « debout », « égal » et « couché », (iii) la position horizontale sur l'abscisse qui est caractérisée par « gauche », « milieu » et « droite » et finalement, (iv) la position verticale qui est l'ordonnée et qui est caractérisée par « en haut », « au centre » et « en bas ». Deux variables discrètes complètent cette représentation : la couleur qui peut être bleue, verte ou rouge et la forme qui peut être rectangle ou ellipsoïdale.

Parce qu'il y a un nombre impair de catégories pour chaque variable (exactement trois), il y a toujours une catégorie centrale. De plus, les sous-ensembles flous qui correspondent aux différentes caractérisations constituent une partition floue de l'univers. Ce qui signifie que pour chaque élément, la somme de ses degrés d'appartenance pour toutes les différentes caractérisations pour une variable donnée est égale à 1. Ainsi, les côtés des trapèzes se coupent en leur milieu.

5.3.4.2.3 Appliquer des modificateurs linguistiques

Notre but est ici de représenter des énoncés comme « très grand » ou « plutôt à gauche » c'est-à-dire des versions modifiées des catégories.

Depuis les premiers travaux de Zadeh (1972), de nombreux modificateurs ont été introduits. Ici, nous utilisons des modificateurs (Bouchon & Yao, 1992) qui exploitent la distribution des catégories pour une variable donnée. La transformation mathématique correspond à une translation (qui correspond à ce que Hersh et Caramazza (1976) ont observé dans leurs expériences) dont l'amplitude et la direction peut être déduite automatiquement. Nous les avons choisis pour leur plausibilité psychologique et pour leur adaptabilité aux différentes catégories (figure 5.19).

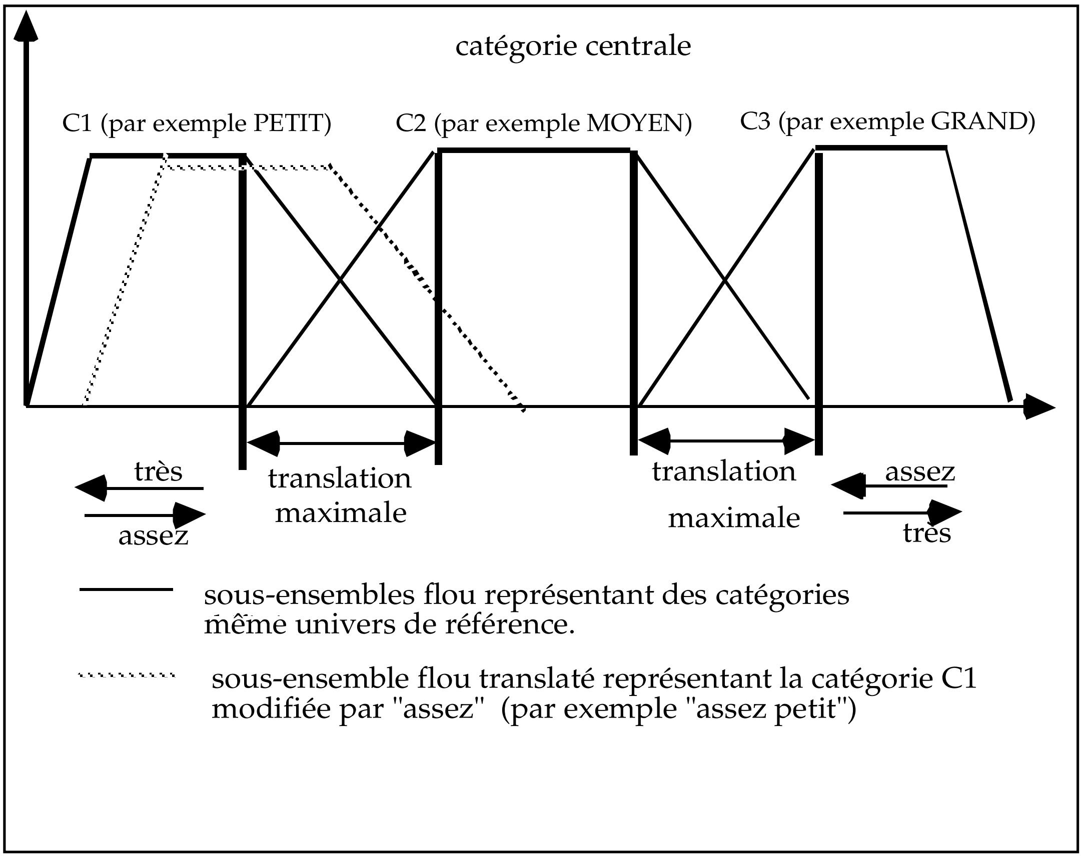


FIG. 5.19. Illustration des mécanismes associés aux modificateurs linguistiques.

Pour une caractérisation et un modificateur donnés, des mécanismes simples permettent de déterminer la translation qui doit être appliquée. Ainsi, pour les modificateurs comme « très », la translation doit se faire vers le bord où se situe la catégorie à modifier pour des modificateurs comme « plutôt ou assez» la direction est celle du centre (figure 5.19). L'amplitude de la translation est définie comme une proportion de la translation maximale qui correspond à la distance du noyau de la première catégorie voisine. Ainsi, une catégorie modifiée n’empiète jamais sur une catégorie voisine. De plus, la translation maximale définit l’amplitude de la translation automatiquement pour chaque variable. Enfin, il est possible d’utiliser des modificateurs de différentes forces. Ainsi, « très très » est un modificateur du même type que « très » mais l’amplitude associée à la translation est plus grande. Pour être plus précis, le coefficient associé à « très très » qui déterminera la proportion de la translation maximale, est plus grand que celui associé à « très ».

5.3.4.2.4 Appliquer une relation floue

Des énoncés comme « plus grand » ou « un petit peu plus à gauche » peuvent être représentés par des relations floues. Le concept de relation floue (Zadeh, 1971) est une généralisation du concept de relation puisqu’il autorise des degrés intermédiaires de relation entre les éléments. Il correspond encore à un sous-ensemble flou. Contrairement au cas des modificateurs, cet ensemble flou ne sera pas construit à partir d’un autre ensemble flou mais à partir d’une valeur précise (l’aire de la figure si la commande est « plus grand »). On peut décomposer ce type de commande en deux parties : la relation proprement dite, qui est par exemple « beaucoup plus », « moins » ou « même » et la catégorie qui est, par exemple, « sur la gauche » ou « grand ».

Il est possible de définir des mécanismes tels qu’à partir de la relation et de la catégorie, le sous-ensemble flou représentant la catégorie puisse être construit. Dans un premier temps, après avoir défini un signe pour chaque relation (les relations de type « moins » sont négatives et les relations de type « plus » sont positives) et pour chaque catégorie (les catégories positives sont à droite de la catégorie centrale et les catégories négatives à gauche), la direction (augmentation/diminution) indiqué par la commande (par exemple « moins grand ») est calculée par le produit des signes de la relation et de la catégorie. Quand la catégorie est centrale, la direction dépend de la position de la valeur à modifier comparativement à la valeur centrale de la catégorie (quand la commande est par exemple « plus rond », la question qui doit être posée est de savoir si la figure est déformée horizontalement ou verticalement). Comme pour les modificateurs, différents coefficients sont associés avec chaque relation exprimant différentes forces. « Beaucoup plus » indique une plus grande variation que « plus ».

5.3.4.2.5 L'historique de la communication

A n'importe quel moment de la communication, ce qui a déjà été dit et fait qui constitue l'historique de la communication, est un élément déterminant dans l'interprétation. Par exemple, ce que la commande « plus grand » signifie peut varier suivant qu'il s'agit d'une première correction visant à approcher la véritable taille de la figure ou qu'il s'agit d'une ultime précision. L'historique permet ainsi d'interpréter la commande avec plus de précision. Sans historique, l'enchaînement des commandes « un peu plus grand » et « un peu plus petit » ne consisterait par exemple, en rien de plus qu'un aller et retour sur la taille initiale de la figure.

Construction de l'historique

Durant la communication, différentes indications et corrections sont données. Ceci peut être représenté par la liste des pentes des différents trapèzes construits pour les précédentes commandes. Pour chaque variable, il y a un historique. Fixer une taille maximale pour cette liste permet de tenir compte des capacités mnémoniques limitées de l'opérateur.

Prise en compte de l'historique

Seulement deux côtés sont utiles pour chaque variable. Ils correspondent aux contraintes les plus restrictives à droite et à gauche (par exemple, beaucoup plus petit que 10.3 cm2 et plus grand que 5.4 cm2) et permettent de construire un sous-ensemble flou. La prise en compte de cet historique est ensuite réalisée en considérant l'intersection de ce sous-ensemble flou avec celui représentant l‘énoncé de la commande courante. Quand l'intersection est trop petite (inférieure à un seuil donné), on peut décider de ne pas tenir compte de l'historique pour donner une réponse appropriée à la commande malgré la contradiction avec les précédentes commandes.

5.3.4.2.6 Choisir une solution appropriée

Choisir un point pertinent

Le point pertinent est défini relativement à la position spécifiée dans la commande de dessin. Par exemple, si la commande demande de dessiner une figure dans le coin supérieur gauche, le point de référence de la figure sera le coin supérieur gauche de la figure. Ce qui signifie que c'est ce point qui sera pris en compte pour caractériser la position de la figure (considérant le même exemple, plus le coin supérieur gauche de la figure est dans le coin supérieur gauche de la zone de dessin plus la position de la figure est acceptable). Si la commande est « à gauche du carré dessiner un … », le milieu du côté droit de la figure est pertinent.

Considérant les trois caractérisations possibles de la position verticale et les trois caractérisations possibles de la position horizontale, nous définissons 9 points pertinents (figure 5.20). La définition d'un point pertinent permet de simplifier la procédure de décision. Nous aurions pu choisir une variable plus sophistiquée peut-être plus psychologiquement pertinente, mais, nous focalisant sur les principes généraux, nous nous en tenons à cette solution.

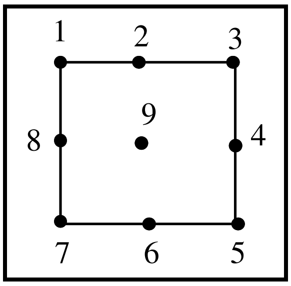


FIG. 5.20 Les neufs point pertinents possibles d'une figure.

Définition d'un degré d'acceptabilité pour tous les points de la zone de dessin

Pour un point p de la zone de dessin, le degré d'acceptabilité est calculé par agrégation de deux degrés intermédiaires $d_1(p)$ and $d_2(p)$ qui correspondent respectivement à deux conditions que doit remplir le point p. Puisqu'il s'agit d'exprimer une conjonction nous avons retenu l'opérateur min, pour l'agrégation.

$$d(p) = \min (d_1(p), d_2(p)),$$

Le degré $d_1(p)$ indique le degré avec lequel le point p de la zone de dessin est un point satisfaisant en tant que point pertinent de la figure et pour la position spécifiée par la commande ("dans le coin supérieur gauche" or "près du cercle"). Le degré $d_2(p)$ indique le degré avec lequel il est possible de placer au point p une figure de la taille et de la forme spécifiée dans la commande. Ce degré est calculé de la façon suivante :

$$d_2(p) = \sup\{(\min (\mu_{taille}(l), \mu_{forme}(l,h)) \,/\, 0 < l < l_{max}, 0 < h < h_{max}(l)\},$$

où l_{max} et h_{max} sont respectivement la plus grande largeur et la plus grande longueur possible en tenant compte de l'espace laissé disponible par les autres figures et où μ_{taille} et μ_{forme} sont respectivement les fonctions d'appartenance des sous-ensembles flous caractérisant la taille et la forme spécifiées par la commande (figure 5.21).

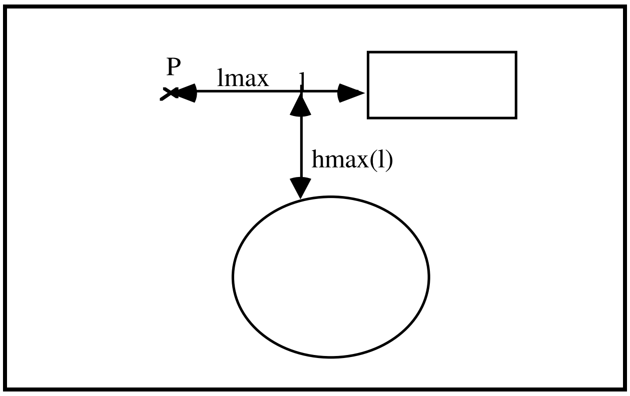


FIG. 5.21 Illustration du calcul de d2(p) quand p est le coin supérieur gauche de la figure.

Le calcul de d(p) pour tous les points de la zone de dessin permet de définir des zones favorables au dessin (figure. 5.22).

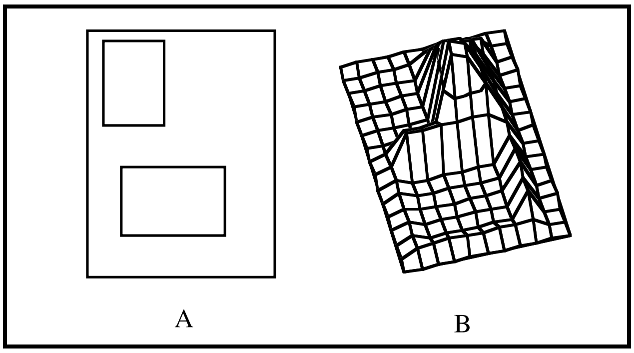


FIG. 5.22 A) Figures déjà présentes. B) Visualisation des zones plus ou moins favorables à la réalisation "d'un grand cercle au centre".

Une solution adaptée à la situation

Le principe général d'une procédure d'optimisation permet de choisir une solution adaptée à la situation sans avoir à la décrire explicitement (figure 5.23).

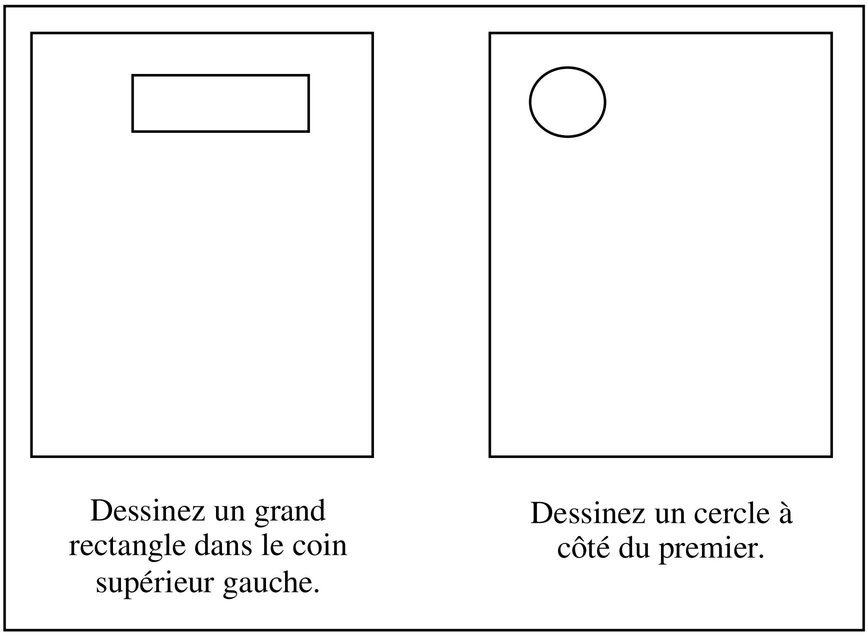


FIG. 5.23 dans ces deux situations, la procédure d'optimisation choisit de dessiner un rectangle vertical et un cercle à droite du premier.

5.3.5 Validation

Ce modèle de l'interprétation de commande a été implémenté, donnant naissance au système SIROCO. Ce système permet de simuler l'opérateur et de valider le modèle en comparant les réponses du système aux réponses de l'opérateur.

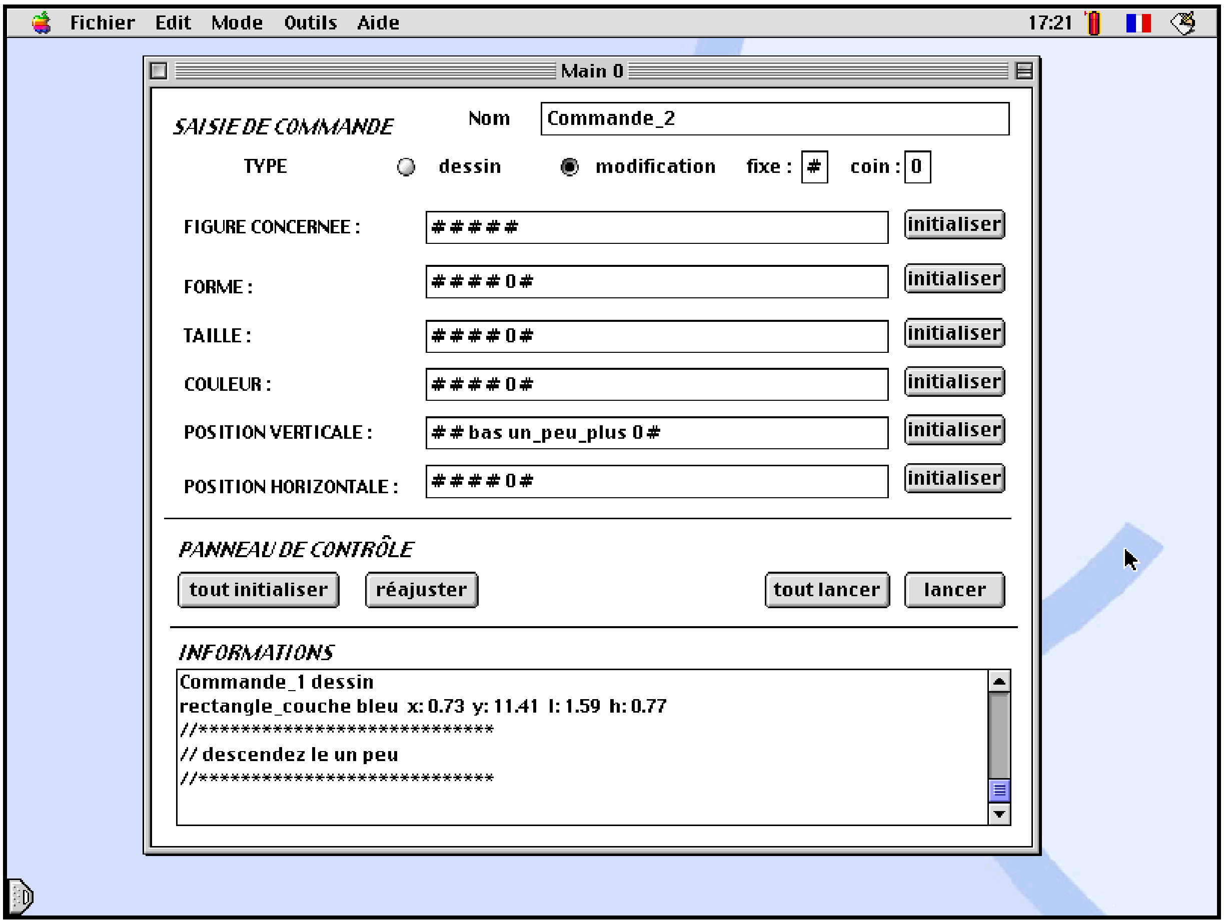
5.3.5.1 Les paramètres du modèle

L'expérience a fourni trente-cinq enregistrements de communication. Dix enregistrements ont été gardés dans le but de tester le modèle. Les autres ont été utilisées pour apprendre les différentes caractérisations floues et les coefficients attachés aux modificateurs et aux relations. Plus précisément, le premier dessin de chaque communication (qui correspond à un contexte minimal et qui par conséquent est censé constituer une représentation fiable de la caractérisation pour l'opérateur) permet de définir les noyaux de toutes les caractérisations. Les supports ont été définis de façon à construire une partition floue pour chaque variable. Etant donné ces caractérisations, l'analyse des résultats expérimentaux a permis de définir les coefficients pour les modificateurs et pour les relations.

5.3.5.2 Simulation

5.3.5.2.1 Description de l'environnement

Développé en C++, SIROCO inclut une interface graphique qui permet de visualiser les réponses du système et celles de l'opérateur. Il permet aussi d'exécuter des fichiers de commandes, d'en saisir de façon interactive et de réajuster les réponses du système sur celles de l'opérateur. Enfin, il permet de visualiser les zones favorables de dessin (en créant un fichier matlab).



SAISIE DE COMMANDE

Nom: Commande_2

TYPE ○ dessin ● modification fixe : # coin : 0

FIGURE CONCERNEE : #####

FORME : ####0#

TAILLE : ####0#

COULEUR : ####0#

POSITION VERTICALE : ##bas un_peu_plus 0#

POSITION HORIZONTALE : ####0#

INFORMATIONS

```
Commande_1 dessin
rectangle_couche bleu x: 0.73 y: 11.41 l: 1.59 h: 0.77
//*****************************
// descendez le un peu
//*****************************
```

FIG. 5.24 La fenêtre principale de l'interface de SIROCO.

5.3.5.2.2 Définition d'un langage minimal

Les commandes qui ont été gardées pour les tests ont été traduites dans le langage minimal qui comprend un nombre limité de mots et une structure stricte. La plupart de ces mots sont des adjectifs caractérisant les figures géométriques (grand, en haut, rectangle, carré, bleu), et aussi, les modificateurs (« très », « assez », « un peu »,…) et les relations (« plus », « beaucoup plus »,…) souvent utilisés dans les commandes. Ce langage se doit de représenter les commandes sans les interpréter. Par exemple, "plus près du bord" n'est pas traduit par "plus vers la gauche" (si la figure est près du bord gauche) mais par « plus extrême » ("extrême" est automatiquement remplacé par la catégorie de l'objet, dans ce cas, la gauche). De même, "faites le rectangle plus long" est traduit par « plus extrême ». Ainsi, si le rectangle est vertical, la hauteur sera augmentée et s'il est horizontal la largeur est augmentée. Nous utilisons aussi le connecteur OU pour représenter des termes comme « rectangle » (un rectangle vertical OU un rectangle horizontal) ou encore "à côté de" (à gauche OU à droite).

5.3.5.2.3 Simulation avec réajustement

Pour cette simulation, la comparaison avec l'opérateur a été faite commande par commande. Chacune des réponses du système a été automatiquement réajustée aux réponses de l'opérateur avant que la commande suivante ne soit interprétée. L'historique de la communication a également été réajusté. Ainsi, pour chaque nouvelle commande, le système a été placé précisément dans la même situation d'interprétation que l'opérateur.

5.3.5.3 Résultats

Dans le but d'évaluer la validité du modèle, nous avons comparé les figures dessinées par le système et celles dessinées par l'opérateur. Plus précisément, nous avons comparé les coordonnées x, y du coin gauche de la figure, la largeur et la hauteur (tableau 5.6).

X:	Y:	L:	H:
0.25 cm	0.35 cm	12.5%	12%

TAB. 5.3.6 L'erreur moyenne du système en comparaison avec la réalisation du sujet.

L'erreur moyenne est tout à fait comparable à la précision finale avec laquelle l'opérateur réalise le dessin qui correspond à une parfaite satisfaction de l'instructeur (rappelons que cette précision est X: 0.20 cm, Y: 0.36 cm, L: 9.2 %, H: 12.4 %). Les dessins réalisés par le système et ceux réalisés par le sujet sont d'ailleurs souvent difficilement distinguables. Le résultat est donc très satisfaisant.

5.3.6 Conclusions

En résumé, nous avons montré ici un ensemble de mécanismes pour construire le sens des commandes en tenant compte de la situation et nous avons montré l'adéquation des réseaux sémantiques et des sous-ensembles flous pour traiter l'incomplétude et l'imprécision inhérente au langage.

Ce type de méthode pourrait être utilisé dans des situations plus riches comportant un nombre plus grand d'objets et de propriétés. En effet, ici l'opérateur change à chaque nouveau dessin et le réseau sémantique représente donc seulement la structure des propriétés des trois figures du dessin. Par contre, une situation dans laquelle, un même opérateur utiliserait le système de façon prolongée dans un processus de création, pour construire des objets d'une grande diversité, serait plus à même de montrer l'intérêt des réseaux sémantiques. Le réseau créé par le système au fur et à mesure de son interaction avec un utilisateur particulier se rapprocherait petit à petit des relations implicites établies par l'utilisateur entre les propriétés, le système enregistrant d'une certaine façon des connaissances du type "pour cet utilisateur les grands carrés ont une couleur bleue" permettant ainsi à l'instructeur de ne plus être aussi exhaustif dans ses commandes. Cette remarque va dans le sens d'une plus grande interaction entre les systèmes de requête et les utilisateurs.

Enfin, la méthode générale que nous avons suivie, consistant dans un premier temps à déterminer un ensemble de contraintes floues et dans un second temps, à trouver les solutions satisfaisant le mieux possible l'ensemble de ces contraintes en tenant compte de l'environnement, semble bien adaptée à la modélisation de l'action. En comparaison avec un système à base de règles où les règles doivent couvrir toutes les situations et être explicitées, cette méthode semble plus adaptative et plus simple à implémenter.

Chapitre 6

6 CONCLUSIONS ET PERSPECTIVES

6.1 Résumé

6.1.1 Un modèle de la pertinence

Nous basant sur une approche pluridisciplinaire de la pertinence, nous avons proposé une vue synthétique de cette notion renvoyant à celle de spécificité. Nous avons étayé cette proposition par de nombreux arguments et nous avons montré dans une expérience portant sur la communication que cette proposition s'accordait bien avec celle faite par Sperber et Wilson (1989).

6.1.2 Un nouveau système de représentation des connaissances

En accord avec cette vue de la notion de pertinence, nous avons développé un nouveau système de représentation des connaissances ainsi qu'une méthode d'apprentissage et une méthode d'exploitation associée viable dans un environnement imparfait (imprécision, incomplétude).

Ce système a l'intérêt de reposer sur 4 principes simples :

- Un principe d'association pour l'évaluation des implications
- Un principe d'économie pour la limitation de la taille du réseau
- Un principe de diffusion pour la construction automatique des sous-ensembles flous
- Un principe de spécificité maximale pour la sélection des informations pertinentes

6.1.3 Faisabilité et plausibilité psychologique

Nous avons montré l'applicabilité du système RELIEFS en tant qu'outil d'extraction d'information pertinente permettant de sélectionner des informations ayant une véritable relation avec la requête (tenant compte des distributions marginales et des effets de transitivité). Nous avons aussi mis en évidence sa plausibilité psychologique en établissant une correspondance entre RELIEFS et le fonctionnement du système cognitif.

6.2 Perspectives

6.2.1 Divers développements et améliorations du système

Le système extrait pour l’instant des implications dont la prémisse peut être définie par un sous-ensemble flou. Nous étudions actuellement une solution dans le cadre de l’exploitation pour que le système puisse gérer des requêtes floues, considérant alors aussi une définition des conclusions par des sous-ensembles flous. De plus, le lien entre deux variables continues pourrait apparaître plus automatiquement (sans requête) en mettant en correspondance des sous-ensembles flous des deux variables, et en étudiant cette correspondance de façon à y reconnaître éventuellement une relation connue. Ceci permettrait par exemple la découverte de lois mathématiques liant les données par extraction d’une forme de relation et comparaison de cette forme avec des formes connues.

Nous souhaitons aussi considérer des mécanismes plus locaux et dynamiques pour l’exploitation, nous ramenant à l'idée de résonance sous une forme proche de celle que l'on retrouve dans les réseaux de Carpenter et Grossberg (1987). Nous souhaitons dans cette perspective, tenter de mieux expliciter le rôle respectif des deux implications qui reste pour l’instant mal défini.

Le système est encore dans une phase de développement et son évaluation précise n’a pas été réalisée. Nous pensons en collaboration avec l’Institut de Veille Sanitaire, pouvoir corroborer les différentes dépendances extraites par le système par des tests statistiques. Le système est encore à l’état de maquette, sa véritable mise en service dans le cadre de la surveillance épidémiologique permettrait de le valider. Cette application a de plus l’intérêt de mettre en avant la dimension temporelle (prise en compte des observations les plus récentes, découvertes de nouveautés) qui est un aspect présent dans le système (principe d’oubli) que nous n’avons pas étudié.

Enfin, la bonne interaction du système avec l’utilisateur est particulièrement importante puisque comme nous l’avons vu, elle est au cœur du processus d’extraction d’information pertinente. De ce point de vue, nous souhaitons poursuivre l’étude initiée avec le système SIROCO dans un cadre plus large pour intégrer SIROCO et RELIEFS en un seul système.

6.2.2 Application à la Recherche d’Information

Nous futures recherches porteront aussi sur l’application de ce système de représentation des connaissances dédié à l’extraction d’information pertinente dans le cadre particulièrement adapté de la Recherche d’Information.

Ainsi, dans le cadre de la recherche documentaire, si l’on considère chaque document comme une observation et chaque mot-clef comme une propriété, ce type de réseau peut être identifié à un thésaurus flou.

Nie et Brisebois (1996) ont résumé les différents problèmes posés par la construction automatique de thésaurus. Dans ces thésaurus, l’évaluation du lien entre deux mots est le plus souvent établi sur la base de la co-occurrence des deux mots dans les documents. Or, d'après ces auteurs :

> - La co-occurrence ne représente pas nécessairement la proximité sémantique des mots. Par exemple, les synonymes ont peu de chance d’être utilisés pour décrire les mêmes documents (on utilise l’un ou l’autre).
>
> - La prise en compte de la co-occurrence pour étendre une requête, ne constitue pas un gain d’information très important puisque justement, les termes choisis pour l’extension sont présents dans les mêmes documents que les termes utilisés dans la requête.
>
> - De plus, on peut ajouter qu’un même mot n’a pas le même sens suivant le contexte dans lequel il est utilisé.

Or, notre système fonctionne sur un principe d’association et mesure donc des co-occurrences. Nous souhaitons néanmoins explorer les nouvelles solutions que semble pouvoir offrir ce nouveau type de représentation, en proposant de :

- Calculer une co-occurrence basée non pas sur la présence réelle des mots dans le document, mais également sur celle des mots liés aux mots présents. Considérant donc de façon analogue à ce que nous avons fait avec les filtres, un voisinage pour chaque mot. Cette solution peut permettre en effet de créer des liens entre des termes synonymes, même s’ils n’apparaissent pas dans les mêmes documents puisqu’ils partagent le même voisinage.

- Ne pas considérer une propagation d’activation suivant les implications ayant pour origine des termes présents dans la requête, mais comme nous l’avons proposé, considérer la spécificité du terme impliqué c'est-à-dire de l'implication inverse. Cette solution permet d'une part de limiter la propagation d'activation et de ne pas retenir les termes trop généraux et d'autre part, de sélectionner des termes qui ajoutent véritablement une information, puisque les termes liés par la relation de spécificité ne sont alors pas nécessairement présents dans les mêmes documents que les termes de la requête.

- Considérer des conjonctions de termes de façon à tenir compte du contexte. En effet si un mot peut survenir dans différents contextes, ceci est beaucoup moins probable pour une conjonction. Cette solution n'est à notre connaissance pas retenue et la complexité du réseau construit en est certainement la cause. Mais comme nous l'avons vu, il est possible de limiter cette complexité et compte tenu de la diversité des mots, le principe d'économie que nous avons proposé semble bien adapté à la recherche d'information puisqu'il permet d'autant plus d'économie que les propriétés (ou mots) sont différentes à chaque observation (ou document).

BIBLIOGRAPHIE

Agrawal, R., Imielsinski, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases*. Proceedings of the ACM-SIGMOD, Conference on management of data, 207-216.

Agrawal, R., Manilla, H., Srikant, R., Toivonen, H., & Verkamo, A. (1995). Fast discovery of association rules, *Advance in knowledge discovery and data mining*, pp 307-328.

Aladenise, N., & Bouchon-Meunier, B. (1997). Acquisition de connaissances imparfaites : mise en évidence d'une fonction d'appartenance. *Revue Internationale de Systémique, 11(1)*, 109-127.

Almuallim, H., & Dietterich, T. G. (1991). *Learning with many irrelevant features*. Proceedings of the AAAI-91, 547-552, Anaheim, CA.

Anderson, A. R., & Belnap, N. D. (1975). *Entailment. The logic of relevance and necessity*. Princeton: Princeton University Press.

Anderson, J. R. (1983a). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1983b). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261-295.

Baddeley, A. (1992). *La mémoire humaine. Théorie et pratique*. Grenoble: PUG.

Bagot, J. D. (1996). *Information, sensation et perception*. Paris: Armand Colin.

Bandler, W., & Kohout, L. J. (1980). Semantics of implication operators and fuzzy relational products. *International Journal of Man-Machine Studies, 12*, 89-116.

Barbut, M., & Monjardet, B. (1970). *Ordre et classification: algèbre et combinatoire*. Paris: Hachette.

Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science, 45(3)*, 149-159.

Baumgartner, E., & Ménard, P. (1996). *Dictionnaire étymologique et historique de la langue française*: Librairie générale française.

Bernard, J. -M. (1996). L'Analyse implicative Bayésienne : Une méthode pour l'étude des dépendances orientées. I: données binaires. *Mathématiques Informatique et Sciences humaines, 134*, 5-38.

Bernard, J.-M., & Poitrenaud, S. (1999). L'analyse implicative bayésienne multivariée d'un questionnaire binaire : Quasi-implications et treillis de Galois simplifié. *Mathématiques Informatique et Sciences humaines, 37(147)*, 25-46.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97*, 245-271.

Bosc, P., Liétard, L., & Pivert, O. (1998). *Extended functional dependencies as a basis for linguistic summaries*. Proceedings of the Second European Symposium on Principles of Data Mining and Knowledge Discovery, 255-263, Nantes.

Bouchon, B., & Yao, J. (1992). Linguistic modifiers and gradual membership to a category. *International Journal on Intelligent Systems, 7*, 25-36.

Bouchon-Meunier, B. (1995). *La logique floue et ses applications*. Paris: Addison-Wesley.

Bournaud, I. (1996). *Regroupement conceptuel pour l'organisation de connaissances*. Thèse de doctorat, Université Paris 6.

Brachman, R. J. (1977). What's in a concept : Structural foundations for semantic networks. *International Journal of Man-Machines Studies, 9*, 127-152.

Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science, 9(2)*, 171-216.

Bruce, H. B. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science, 45(3)*, 142-148.

Camus, J.-F. (1996). *La psychologie cognitive de l'attention*. Paris: Armand Colin.

Carpenter, A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37*, 54-115.

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., & Freeman, D. (1990). AutoClass : A Bayesian Classification System. In J. W. Shavlik & T. G. Dietterich (Eds.), *Readings in Machine Learning*, (pp. 296-306). San Mateo, CA: Morgan Kauffman.

Chen, H., & Ng, T. (1995). An algorithmic approach to concept exploration in a large knowledge network (automatic thesaurus consultation): symbolic branch-and-bound search vs. connectionist Hopfield net activation. *Journal of the American Society for Information Science, 46(5)*, 348-369.

Chen, S., & Horng, T. J. (1999). Fuzzy query processing for document retrieval based on extended fuzzy concept networks. *IEEE Transactions on Systems, Man and Cybernetics, 29(1)*, 96-104.

Cheung, D. W., Han, J., Ng, V. T., & Wang, C. Y. (1996a). *Maintenance of discovered association rules in larges databases : An incremental updating technique.* Proceedings of the International Conference on Data Engineering, 106-114.

Cheung, D. W., Ng, V. T., & Tan, B. W. (1996b). *Maintenance of discovered knowledge : A multi-level association rules.* Proceedings of the Second International Conference on Knowledge Discovery and Data-Mining, 307-310.

Chevallet, J. P. (1992). *Un modèle logique de recherche d'informations appliqué au formalisme des graphes conceptuels. Le prototype elen et son expérimentation sur un corpus de composants logiciels*. Thèse de doctorat, Université Grenoble 1.

Clancey, W. (1997). *Situated Cognition*. Cambridge: Cambridge University Press.

Clark, P., & Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning, 3*, 261-283.

Cloninger, M. R., Baldwin, R. E., & Krause, G. F. (1976). Analysis of sensory rating scales. *Journal of Food Science, 41*, 1225-1228.

Cohen, P. R., & Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management, 23(4)*, 255-268.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval, 7*, 19-37.

Crestani, F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review, 11*, 453-498.

Cuadra, C. A., & Katter, R. V. (1967). *Experimental studies of relevance judgments* (TM-3520. 001-002): System Developpement Corporation, Santa Monica, californie.

Damasio, A., & Damasio, H. (1992). Le cerveau et le langage. *Pour la science, 181*, 80-87.

Darwiche, A. (1997). A logical notion of conditional independence: properties and applications. *Artificial Intelligence*, *97*, 45-82.

Denos, N. (1997). *Modélisation de la pertinence en recherche d'information : modèle conceptuel, formalisation et application*. Thèse de doctorat, Université Grenoble I.

Devlin, K. (1991). *Logic and information*. Cambridge: Cambridge University Press.

Diday, E. (1971). La méthode des nuées dynamiques. *Revue de Statistique Appliquée, XIX*, 19-34.

Dubois, D., Martin-Clouaire, R., & Prade, H. (1988). Practical Computing in Fuzzy Logic. In M. M. Gupta et T. Yamakawa (Eds.), *Fuzzy Computing*, (pp. 11-34): Elsevier.

Dubois, D., & Prade, H. (1991). Conditioning, non-monotonic logic and non-standard uncertainty models. In I. R. Goodman, M. M. Gupta, H. T. Nguyen, & G. S. Rogers (Eds.), *Conditional Logic in expert systems*, (pp. 115-158). North-Holland: Elsevier Science Publishers.

Dunn, J. M. (1976). Intuitive semantics for first-degree entailments and coupled trees. *Philosophical Studies, 29*, 149-168.

Fahlman, S. E. (1979). *NETL. A system for representing and using real-world knowledge*. Cambridge, MA: MIT Press.

Faÿ, A., & Jaffray, J.Y. (1997). *Une Introduction aux réseaux probabilistes*. Rapport interne du Laboratoire d'Informatique de Paris 6.

Fayyad, U. M., & Irani, K. (1993). *Multi-interval discretization of continuous-valued attributes for classification learning*. Proceedings of the the 13th International Joint Conference on Artificial Intelligence, 1022-1027, Chambéry.

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning, 2*, 139-172.

Ford, K., & Hayes, P. (1998). L'intelligence artificielle. *Pour la science, 254*, 158-163.

Foskett, D. J. (1972). A note on the concept of relevance. *Information Storage and Retrieval, 8(2)*, 77-78.

Froehlich, T. J. (1994). Relevance reconsidered-toward an agenda for the 21st century: Introduction to special issue on relevance research. *Journal of the American Society for Information Science, 45(3)*, 124-134.

Fuhr, N. (1992). Probabilistic models in information retrieval. *The Computer Journal, 35(3)*, 243-255.

Galles, D., & Pearl, J. (1997). Axioms of causal relevance. *Artificial Intelligence*, *97*, 9-43.

Ganascia, J. G. (1987). *AGAPE et CHARADE, deux techniques d'apprentissage symbolique appliquées à la construction de bases de connaissances*. Thèse d'Etat, Université Paris 11.

Gaudin, R., Missaoui, R., & Alaoui, H. (1995). Incremental concept formation algorithm based on galois (concept) lattices. *Computational Intelligence, 11(2)*, 246-267.

Girard, R., & Ralambondrainy, H. (1996). *Conceptual classification from imprecise data.* Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'96), 247-252, Granada, Spain.

Gluck, M., & Corter, J. (1985). *Information uncertainty and tue utility of categories.* Proceedings of the Seventh Annual Conference of the Cognitive Science Society, 283-287.

Goodman, L. A., & Kruskal, W. H. (1959). Measures of association for cross classifications. II: Further discussion and References. *Journal of the American Statistical Association, 54*, 123-163.

Gras, R., & Laher, A. (1993). L'implication statistique : une nouvelle méthode d'analyse de données. *Mathématiques Informatique et Sciences Humaines, 120*, 5-31.

Green, R. (1995). Topical relevance relationships. I. Why topic matching fails. *Journal of the American Society for Information Science, 46(9)*, 646-653.

Green, R., & Bean, C. A. (1995). Topological relevance relationships. II. An exploratory study and preliminary typology. *Journal of the American Society for Information Science, 46(9)*, 654-662.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics*, (Vol. 3: Speech and acts). New York: Academic Press.

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurence. *American Psychologists, 39(12)*, 1372-1388.

Haykin, S. (1999). *Neural networks. A comprehensive foundation*. Upper Saddle River, New Jersey: Prentice Hall International Inc.

Hebb, D. O. (1949). *The organisation of behaviour*. New York: John Wiley and Sons Inc.

Hersh, H. M., & Caramazza, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General, 105(3)*, 254-276.

Holsheimer, M., & Siebes, A. (1994). *Data mining, the search for knowledge in databases* (CS-R9406), Amsterdam.

Hong, T. P., & Lee, C. Y. (1996). Induction of fuzzy rules and membership function from training examples. *Fuzzy Sets and Systems, 84(1)*, 33-47.

Hopfield, J. J. (1982). *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the National Academy of Sciences, 2554-2558, USA.

Horng, Y. J., & Chen, S. M. (1999). Finding inheritance hierarchies in fuzzy-valued concept networks. *IEEE Transactions on Systems, Man, and Cybernetics, 29*, 126-135.

Howard, D. L. (1994). Pertinence as reflected in personal constructs. *Journal of the American Society for Information Science, 45(3)*, 172-185.

Huang, C. (1997). Principle of information diffusion. *Fuzzy Sets and Systems, 91*, 69-90.

Ichikawa, R., Nishimura, K., Kunugi, M., Shimada, K., Shimakura, Y., Fujisawa, Y., & Matsunoki, Y. (1992). *Auto-tuning method for fuzzy membership functions using neural network learning algorithm*. Proceedings of the Second International Conference on Fuzzy Logic & Neural Networks, 345-348, Lizuka, Japan.

Itzkovich, I., & Hawkes, L. W. (1994). Fuzzy extension of inheritance hierarchies. *Fuzzy Sets and Systems, 62(2)*, 143-153.

John, G. H., Kohavi, R., & Pfleger, K. (1994). *Irrelevant features and subset selection problem*. Proceedings of the 11th International Conference on Machine Learning, 121-129, New Brunswick.

Jones, W. P., & Furnas, G. W. (1987). Pictures of relevance : A geometric analysis of similarity measures. *Journal of the American Society for Information Science, 38(6)*, 420-442.

Kacprzyk, J., & Iwanski, C. (1992). *Fuzzy quantifiers in inductive learning with perception of errors in data*. Proceedings of the IEEE International Conference on Fuzzy Systems, 477-484.

Kayser, D. (1997). *La représentation des connaissances*. Paris: Hermes.

Kerber, R. (1992). *ChiMerge: Discretisation of numeric attributes*. Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92), 123-127, San Jose, CA.

Kim, J. H., & Pearl, J. (1985). *A computational model for combined causal and diagnostic reasoning in inference systems*. Proceedings of the 8th International Conference on Artificial Intelligence, 190-193.

Kintsch, W. (1988). The role of knowledge in discourse comprehension : A construction-integration model. *Psychological Review, 95(2)*, 163-182.

Kira, K., & Rendell, L. (1992). *A practical approach to feature selection*. Proceedings of the 9th International Conference on Machine Learning, 249-256, Aberdeen, Scotland.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59-69.

Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man Machine Studies, 24*, 65-75.

Laberge, D. (1995). *Attentional Processing*. Cambridge, Mass: Harvard University Press.

Langley, P., & Sage, S. (1994). *Oblivious decision trees and abstract cases*. Proceedings of the Working Notes of the AAAI-94 Workshop on Case-Based Reasoning, 113-117, Seattle.

Lauritzen, S. L., & Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of The Royal Statistical Society (B)*, *50*, 157-224.

Le Ny, J. F. (1972). *Le conditionnement et l'apprentissage*. Paris: PUF.

Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: a new linear threshold algorithm. *Machine Learning, 2*, 285-318.

Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of scale analysis and factor Analysis. *Psychological Bulletin, 45*, 507-530.

Lucarella, D. (1990). *A model for hypertext-based Information retrieval*. Proceedings of the ECHT’90, Paris.

Lucarella, D., & Morara, R. (1991). FIRST: Fuzzy information retrieval system. *Journal of Information Sciences, 17(1)*, 81-91.

Machado, R. J., & Rocha, A. F. (1990). *The combinatorial neural network: a connexionnist model for knowledge based systems*. Proceedings of the 3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-bases systems, 9-11, Paris.

Marsala, C. (1998). *Apprentissage inductif en présence de données imprécises : Construction et utilisation d'arbres de décision flous*. Thèse de doctorat, Université Paris 6.

Massaro, D. W., Cohen, M. M., & Gesi, A. T. (1993). Long-term training, transfer, and retention in learning to lipread. *Perception & Psychophysics, 5*, 549-562.

McAdams, S. (1994). Audition : Psychologie, perception et cognition. In M. Richelle, J. Requin, & M. Robert (Eds.), *Traité de Psychologie Expérimentale*, (Vol. 1, pp. 283-344). Paris: Presses Universitaires de France.

Meghini, C., & Straccia, U. (1996). *A relevance terminological logic for information retrieval*. Proceedings of the 19th Annual ACM Conference on Research and Development in Information Retrieval, Zurich.

Michalski, R. S. (1990). A theory and methodology of inductive learning. In J. W. Shavlik & T. G. Dietterich (Eds.), *Readings in machine learning*, (pp. 70-95). San Mateo, CA: Morgan Kauffman.

Michalski, R. S., & Stepp, R. E. (1983). Learning from observation : Conceptual clustering, *Machine Learning, an Artificial Intelligence Approach*, (Vol. 1, pp. 331-363): Morgan Kauffmann.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mineau, G. (1990). *Structuration des base de connaissances par généralisation.* Thèse de doctorat, Université de Montréal. Canada.

Miyamoto, S. (1990). Information retrieval based on fuzzy associations. *Fuzzy Sets and Systems, 38*, 191-205.

Nagel, T. (1974). What is like to be a bat. *Philosophical Review, 83.*

Neely, J. H. (1991). Semantic priming effects in visual words recognition : A selective review of current findings and theories. In D. B. G. W. Humpreys (Ed.), *Basic processes in reading : visual word recognition*, (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nie, J. Y. (1988). *An outline of a general model for information retrieval.* Proceedings of the 11th Annual ACM Conference on Research and Development in Information Retrieval, Grenoble.

Nie, J. Y., & Brisebois, M. (1996). An inferential approach to information retrieval and its implementation using a manual thesaurus. *Artificial Intelligence Review, 10*, 409-439.

Norman, D. A. (1993). Cognition in the head and in the world : an introduction to the special issue on situated action. *Cognitive Science, 17(1)*, 1-6.

Omri, N. (1994). *Système Interactif Flou d'Aide à l'utilisation de Dispositifs techniques : Le système SIFADE* .Thèse de doctorat, Université Paris 6.

Pavlov, I. P. (1927). *Conditioned reflexes*. Londres: Oxford University Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan-Kaufmann.

Piatetsky-Shapiro, G. (1991). Discovery, analysis and presentation of strong rules. In G. Piatetsky-Shapiro & W. J. Frawley (Eds.), *Knowledge discovery in datatbases*, (pp. 229-248). Menlo Park, CA: AAAI Press.

Pitrat, J. (1977). A chess combination program which uses plans. *Artificial Intelligence, 8(3)*, 275-321.

Pitrat, J. (1990). *Métaconnaissances*. Paris: Hermès.

Poitrenaud, S. (1995). The Procope semantic network : an alternative to action grammars. *International Journal of Human-Computer Studies, 42*, 31-69.

Poitrenaud, S. (1998). *La représentation des PROCédures chez l'OPErateur : Description et mise en oeuvre des savoir-faire.* Thèse de doctorat, Université Paris 8.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing*, (pp. 216-270). Cambridge, MA: MIT Press.

Quinlan, J. R. (1983). Learning efficient classification procedures and their application to chess en games. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning : An artificial intelligence approach*. San Mateo: Morgan Kauffman.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81-106.

Radecki, T. (1976). Mathematical model of information retrieval system based on the concept of fuzzy thesaurus. *Information Processing and Management, 12*, 313-318.

Ragel, A., & Cremilleux, B. (1998). *Treatment of missing values for association rules*. Proceedings of the Second Pacific Asia Conference, PAKDD'98., Melbourne.

Richard, J. F., Poitrenaud, S., & Tijus, C. A. (1993). Problem-solving restructuration: Elimination of implicit constraints. *Cognitive Science, 17*, 497-529.

Rifqi, M. (1996). *Mesures de comparaison, typicalité et classification d'objets flous : théorie et pratique*. Thèse de doctorat, Université Paris 6.

Rosch, E. (1976). Classification d'objets du monde réel : Origines et représentations dans la cognition. *Bulletin de Psychologie, Numéro spécial. La mémoire sémantique*, 242-250.

Rossazza, J. P. (1992). *Une représentation centrée objet*. Actes du Deuxième Congrès National sur les Applications des Ensembles Flous, Nîmes.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGrawHill.

Saracevic, T. (1975). Relevance: a review of the literature and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science, 6*, 321-343.

Savoy, S. (1992). Bayesian inference networks and spreading activation in hypertext systems. *Information Processing and Management, 28(3)*, 389-406.

Schaüble, P. (1997). *Multimedia Information Retrieval. Content-Based Information Retrieval from Large Text and Audio Databases*. Boston: Kluwer Academic Publishers.

Schneider, M., Shnaider, E., Kandel, A., & Chew, G. (1998). Automatic construction of FCMs. *Fuzzy Sets and Systems, 93*, 161-172.

Shortliffe, E. H. (1976). *Computer-Based medical consultations : MYCIN*. New York: Elsevier.

Sowa, J. F. (1984). *Conceptual structures: Information processing in mind and machine*. Reading, MA: Addison-Wesley.

Sowa, J. F. (1991). *Principles of semantic networks*. San Mateo: Morgan Kauffman Publishers.

Sperber, D., & Wilson, D. (1989). *La pertinence: Communication et cognition*. Paris: Les éditions de minuit.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31-95.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*. Oxford.

Subramanian, D., Greiner, R., & Pearl, J. (1997). The relevance of relevance. *Artificial Intelligence, 97*, 1-5.

Tijus, C. A., & Moulin, F. (1997). L'assignation de signification d'histoires drôles. *L'Année Psychologique*, 33-75.

Treisman, A. (1992). L'attention, les traits et la perception des objets. In D. Andler (Ed.), *Introduction aux sciences cognitives*, (pp. 152-191). Paris: Gallimard.

Turtle, H., & Croft, W. B. (1990). *Inference networks for document retrieval*. Proceedings of the 13th International Conference on Research and Development in Information Retrieval, 1-24, New York.

Urdapilleta, I. (1998). *Catégorisation et implication de propriétés*. Thèse de doctorat, Université Paris 8.

van Rijsbergen, C. J. (1986). A non-classical logic for information retrieval. *The Computer Journal, 29(6)*, 481-485.

van Rijsbergen, C. J. (1989). *Towards an information logic*. Proceedings of the Twelfth Annual ACM Conference on Research and Development in Information Retrieval, 77-86, New York.

Varela, F. (1989). *Autonomie et connaissance. Essai sur le vivant*. Paris: Seuil.

Wallach, D., & Lebière, C. (1998). *Introduction to ACT-R.* Tutorial of The European Conference on Cognitive Modelling. Nottingham.

Wang, L. X., & Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics, 2*, 1414-1427.

Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval, 9(8)*, 457-471.

Woods, W. A. (1975). What's a link: foundations for semantic networks. In D. a. C. Bobrow, A. (Ed.), *Representation and Understanding*, (pp. 35-82). New York: Academic Press.

Wu, T. P., & Chen, S. M. (1999). A new method for constructing membership functions and fuzzy rules from training examples. *IEEE Transactions on Systems, Man and Cybernetics, 29*, 25-40.

Yager, R. R. (1982). Measuring tranquility and anxiety in decision making: an application of fuzzy sets. *International Journal of General Systems, 8*, 139-146.

Yager, R. R. (1992). Default knowledge and measure of specificity. *Information Sciences, 61*, 1-44.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338-353.

Zadeh, L. A. (1968). Probability measures of fuzzy events. *J.Math. Analysis and Appl., 10*, 421-427.

Zadeh, L. A. (1971). Similarity Relations and Fuzzy Orderings. *Information Sciences, 3*, 177-200.

Zadeh, L. A. (1972). A fuzzy-set theoretic interpretation of linguistic hedges. *Journal of Cybernetics, 2(2)*, 4-34.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences, 8*, 199-249.

Zadeh, L. A. (1976). The Concept of a Linguistic Variable and its Application to Approximate Reasoning, part 3. *Information Sciences, 9*, 43-80.

Zadeh, L. A. (1978). A meaning representation language for natural languages. *International Journal of Man-Machines Studies, 10*, 395-460.

Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural language. *Comp. and Maths with Appls, 9*, 149-184.

ANNEXE

Cet extrait de programme commenté a pour objectif de donner une vue concrète et globale du système RELIEFS. Le langage choisi pour l'implémentation est le C++. Fournissant une vue précise et concise des opérations réalisées et mettant en avant l'adéquation d'un langage objet pour l'implémentation de nos méthodes, le code a été laissé dans sa forme originale. Néanmoins, des descriptifs généraux des étapes de traitement réalisées apparaissent en gras.

```
                        //**********************************************************
                                  // PROCEDURE PRINCIPALE
                        //**********************************************************
void main()
{

// CONSTRUCTION DU RESEAU A PARTIR DES EXEMPLES

        /*******************************************************/
        /* Ouverture des fichiers de données            */
        /*******************************************************/
        // fichier décrivant les différents exemples
        VG_fichier_des_exemples = new ifstream("exemples",ios::in);
        assert(VG_fichier_des_exemples !=0);

        // fichier décrivant les différentes variables
        VG_fichier_des_variables = new ifstream("variables",ios::in);
        assert(VG_fichier_des_variables !=0);

        // fichier décrivant les différentes propriétés que l'on cherche à prédire
        VG_fichier_des_objectifs = new ifstream("objectifs",ios::in);
        assert(VG_fichier_des_objectifs !=0);

        combinaison_brute* exemple_courant;

        /********************************
        // Boucle de lecture des exemples
        //*******************************
        while (c != '*')
        {
                //******************************
                // Lecture de l'exemple courant à partir du fichier des exemples
                //******************************
                exemple_courant = new combinaison_brute();
                (*VG_fichier_des_exemples)>>exemple_courant;

                //******************************
                // Prise en compte de l'exemple courant pour la construction du réseau
                //******************************
                VG_structure->prendre_en_compte(exemple_courant);
                VG_structure->prise_en_compte_somme();


                //******************************
                // Réinitialisation du réseau pour la prise en compte de l'exemple suivant
                //******************************
                VG_structure->desactivation();
        }
```

```
// EXTRACTION DES SPECIFITES LES PLUS FORTES

        //******************************
        // Une fois le réseau construit,  les spécificités sont calculées
        // pour l'ensemble des noeuds du réseau et imprimées dans un fichier.
        //******************************
        VG_structure->imprimer_dependances(VG_fichier_des_resultats);
}
```

```
                              //*************************************
                  // PRISE EN COMPTE D'UN EXEMPLE POUR LA CONSTRUCTION DU RESEAU
                              //*************************************


void structure::prendre_en_compte(combinaison_brute* cb)
{
        //*************************************
        // Génération des nouveaux nœuds et connections
        // nécessaires à la prise en compte de l'exemple
        //*************************************
        generer_noeuds_et_connections(cb);

        //*************************************
        // Mise à jour des connections du réseau à partir de l'exemple
        //*************************************
        mise_a_jour_connections();
}


//*********************************************
// Génération des nœuds et connexions.
// Les nœuds s simples sont systématiquement créés lorsqu'une propriété est portée par l'exemple.
// Concernant les nœuds complexes (représentant des conjonctions de propriétés), on applique le
// principe d'économie implémenté dans la procédure prediction_creation.
//*********************************************
void structure::generer_noeuds_et_connections(combinaison_brute* cb)
{
        combinaison_pondere* liste_ref;
        liste_noeud* actives;
        actives = new liste_noeud();

        //********************************
        // Traitement des nœuds simples (pas de conjonction, une seule propriété)
        //********************************
        liste_ref = traitement_noeuds_simples(cb);
        liste_des_noeuds->activation(liste_ref, actives);
        // remarque : a chaque fois qu'on active un noeud,
        // on en profite pour l'ajouter a la liste des noeuds actives (actives)
        // construction des connections
        actives->construire_connections();

        //********************************
        // Traitement des nœuds complexes (conjonction de propriétés)
        //********************************
        traitement_noeuds_complexes(liste_ref, actives);
        actives->construire_connections();
}
```

```
//***************************************
// Traitement des nœuds complexes.
// La liste des nœuds activés (c'est-à-dire les nœuds dont les propriétés
// sont portées par l'exemple) est parcourue.
// Pour chacun de ces nœuds, on appelle la procédure prediction_creation.
// Du fait que chacun des nouveaux  nœuds créés peut donner lieu à son tour
// à la création de nouveaux nœuds, on définit un système de marquage et la
// procédure ne se termine que lorsque tous les nœuds sont marqués.
//***************************************
void structure::traitement_noeuds_complexes(combinaison_pondere* liste_ref,liste_noeud* actives)
{
//boucler sur la liste des noeuds actives
        maillon_noeud* nc = actives->retourne_tete();
        boolean pas_termine = VRAI;
        while (pas_termine)
        {
                nc = actives->retourne_tete();
                while (nc != 0)
                {
                        if (!(nc->retourne_valeur()->est_marque()))
                        {
                                nc->retourne_valeur()->prediction_creation(liste_ref, actives);
                                nc->retourne_valeur()->marquer();
                        }
                        nc = nc->retourne_suivant();
                }

                if (actives->tous_marque()) pas_termine= FAUX;
        }
}


//*****************************************
// Procédure prediction_creation qui limite le nombre de nœuds créés.
// Le principe : la catégorie des carrés blancs n'a de raison d'exister
// que s'il existe au moins un carré d'une autre couleur ou une forme
// blanche qui n'est pas carrée.
//*****************************************
void noeud::prediction_creation(combinaison_pondere* liste_ref,liste_noeud* actives)
{
        combinaison_pondere* predictions;
        maillon_attribut_valeur_pondere* mavp;
        maillon_attribut_valeur_pondere* mavp1;
        maillon_attribut_valeur_pondere* mavp2;
        maillon_attribut_valeur* nmav;

        mavp1 = liste_ref->retourne_tete();
        int prop_trouve, var_trouve;
        mon_int v1, m1, v2, m2, v_memo, m_memo;
        float p_memo;
```

```
noeud* nn;
combinaison* id;
combinaison* id1;
float le_poids;

// ****************************
// Boucle sur les propriétés portées par l'exemple
// mavp1 :  propriété portée par l'exemple.
//*****************************
while (mavp1 != 0)
{
        v1 = mavp1->retourne_valeur()->retourne_variable();
        m1 = mavp1->retourne_valeur()->retourne_modalite();
        id = this->retourne_identite();

        //***************************************************
        // Traitement du cas où le nœud existe déjà
        //**************************************************
        if ((actives->rechercher_spe(id,v1,m1)==1))
        {
                // ne rien faire
        }//if

        //******************************
        // Dans les autres cas
        //******************************
        else
        {
                // les prédictions simples suffisent
                predictions = retourne_predictions();
                prop_trouve = 0;
                var_trouve = 0;
                if (predictions != 0)
                {
                mavp2 = predictions->retourne_tete();

                //********************************
                // Boucle sur les propriétés prédites par le réseau à partir des propriétés
                // du nœud pour lequel la procédure est appelée.
                // Suivant le résultat de la comparaison entre ce qui est prédit et
                // les propriétés portées par l'exemple,  on construit ou non
                // de nouvelle conjonctions.
                //********************************
                // mavp2: propriété prédite par le reseau
                while ((mavp2 != 0))
                {
                        v2 = mavp2->retourne_valeur()->retourne_variable();
                        m2 = mavp2->retourne_valeur()->retourne_modalite();
```

```
        if (v1 == v2)
        {
                //memorisation pour le cas ou une seule variable
                p_memo = mavp2->retourne_valeur()->retourne_poids();
                v_memo = v2;
                m_memo = m2;

                var_trouve++;

                if (m1 == m2)
                {
                        prop_trouve = 1;
                }

        }
mavp2 = mavp2->retourne_suivant();
}// while
}// predictions!=0

if (var_trouve!=0)
{
//*************************************************
// (var_trouve == 0) La propriété n'est pas prédite et aucune
// propriété n'est en contradiction : ne rien faire.
// Ce cas est ajouté pour traiter les cas initiaux et les données incomplètes.
// Exemple: C'est la première figure, ou bien les figures précédentes
// ne comportaient pas d'indications de couleur, survient un carré blanc,
// il n'est pas utile de construire la catégorie des carrés blancs.
//*************************************************

if (var_trouve==1)
{
        if (prop_trouve==0)
        {
        //*********************************************
        // La propriété n'est pas prédite mais une propriété et une seule
        // est prédite pour la même variable.
        // La prédiction est donc en contradiction avec la propriété portée
        // par l'exemple.
        // Il faut créer deux nœuds.
        // Exemple:  Les figures précédentes étaient des carrés noirs
        // survient un carré blanc, il faut construire le nœud représentant
        // les carrés blancs ainsi que celui représentant les carrés noirs.
        // C'est précisément ici que ce fait l'économie :
        // On ne construit un nœud réprésentant une conjonction que si un nœud
        // représentant la même prémisse mais une conclusion différente est présent.
        // Il faut aussi recopier toutes les connections du nœud représentant
        // la prémisse.
        //*********************************************
```

```
    // ***************************
    // Création du premier nœud
    // On ajoute prémisse+conclusion (prédiction)
    // **************************
    id1 = new combinaison();
    id1->copie(identite);
    nmav = new maillon_attribut_valeur(v_memo,m_memo);
    id1 ->inserer(nmav);
    // l'activation est le produit de l'activation du noeud courant (premisse)
    // avec l'activation de la conclusion (mavp)
    le_poids = activation * p_memo;
    nn = VG_structure->rechercher_placer(id1, le_poids);

    if (nn!=0)
    {
        if (le_poids>0) actives->placer(nn);

        //*****************************************
        // Création des connections identiques à celles partant
        // du nœud représentant la prémisse.
        //*******************************************
        nn->copie_connections(this);
        nn->definir_somme(somme);
    }
    //**************************


    // ***************************
    // Création du second nœud
    // On ajoute prémisse+conclusion (propriété portée par l'exemple)
    // **************************
    id1 = new combinaison();
    id1->copie(identite);
    nmav = new maillon_attribut_valeur(v1,m1);
    id1 ->inserer(nmav);
    // l'activation est le produit de l'activation du noeud courant (premisse)
    // avec l'activation de la conclusion (mavp)
    le_poids = activation * mavp1->retourne_valeur()->retourne_poids();
    nn = VG_structure->rechercher_placer(id1, le_poids);
    if (nn!=0)
    {
        actives->placer(nn);
    }
    }// if (prop_trouve == 0)
}//if (var_trouve == 1)
```

```
            else // var_trouve>1
            {
            // ***************************
            // Il y a plusieurs propriétés prédites en contradiction
            // avec la propriété portée par l'exemple. Seul le nœud
            // représentant la conjonction est créé.
            // Création du nœud :
            // On ajoute prémisse + conclusion (propriété portée par l'exemple)
            // ***************************

            id1 = new combinaison();
            id1->copie(identite);
            nmav = new maillon_attribut_valeur(v1,m1);
            id1->inserer(nmav);
            // l'activation est le produit de l'activation du noeud courant (premisse)
            // avec l'activation de la conclusion (mavp)
            le_poids = activation * mavp1->retourne_valeur()->retourne_poids();
            nn = VG_structure->rechercher_placer(id1, le_poids);

            if (nn!=0)
            {
                actives->placer(nn);
            }
        }// else
        }//if  (var_trouve!=0))
        }// else

        // ***************************
        // On passe au nœud activé suivant
        // ***************************
        mavp1 = mavp1->retourne_suivant();
    }//while
    delete predictions;
}//void
```

```
                    //**************************************
                    // MISE A JOUR DES CONNECTIONS
                    //**************************************

//********************************************************
// La mise à jour est appelée pour l'ensemble des nœuds du réseaux
//********************************************************
void structure::mise_a_jour_connections()
{
        maillon_noeud* mn = liste_des_noeuds->tete;
        while (mn != 0)
        {
                mn->retourne_valeur()->mise_a_jour_connections();
                mn = mn->retourne_suivant();
        }
}

//*****************
// Mise à jour de toutes les connections partant du nœud
// pour lequel est appelé la méthode.
//*******************
void noeud::mise_a_jour_connections()
{
        sorties->mise_a_jour();
}

//*******************************************
// La connection est mise à jour à partir de son ancienne valeur (force),
// de l'activation du noeud de départ, de l'activation du nœud d'arrivée
// et de la somme des activations du nœud de départ sur les précédentes observations.
//*******************************************
void connection::mise_a_jour()
{
        if (depart->retourne_activation()!=0)
        {
                force =
                fonction_modification(    force,
                                          depart->retourne_activation(),
                                          arrivee->retourne_activation(),
                                          depart->retourne_somme());
        }
}

//************************************
// La mise à jour est de type associatif (dd*da)
//************************************
float fonction_modification(float f, float dd, float da, float s)
{
        return( ((s * f) + dd*da  / (s+dd) );
}
```

```
                    //***********************************
                              // CALCUL DES SPECIFICITES
                       //*****************************

//********************************************
// Le calcul des spécificités relatives est réalisé pour tous les nœuds du réseau
//************************************

void structure::imprimer_dependances(ofstream* fichier)
{
        maillon_noeud* mn = liste_des_noeuds->tete;

        while (mn != 0)
        {
                mn->retourne_valeur()->imprimer_dependances(fichier);
                mn = mn->retourne_suivant();
        }
        *fichier<<"------------------------"<<"\n"<<flush;
}


//************************************
// Détail du calcul pour un nœud
//************************************

void noeud::imprimer_dependances(ofstream* fichier)
{
        maillon_connection* mc = entrees->retourne_tete();
        maillon_connection* mc2;
        maillon_connection* mc3;
        float f;
        noeud* depart;
        noeud* depart2;
        float t[C_nombre_modalites];
        int trouve;
        float max, Spe,;
        liste_connections* liste_resultat;

        //********************************************
        // Boucle sur les connections qui correspondent à des entrées
        // (le nœud courant correspond au nœud "arrivée" de la connection).
        //********************************************
        while (mc != 0)
        {
                depart = mc->retourne_valeur()->retourne_depart();
                if (!(depart->est_marque()))
                {
                        depart->marquer();
                        liste_resultat = new liste_connections();
                        mc2 = mc->retourne_suivant();
```

```
t[0]=1;
t[1]=mc->retourne_valeur()->retourne_force();
max = mc->retourne_valeur()->retourne_force();
liste_resultat->ajouter3(mc);

//*************************************
// Boucle sur les connections  qui correspondent à des entrées
// et dont le nœud de départ représente des propriétés
// qualifiant la même variable que celle trouvée juste ci-dessus.
// Ceci permet de calculer la spécificité par variable.
//*************************************
while (mc2 != 0)
{
        depart2 = mc2->retourne_valeur()->retourne_depart();
        if (!(depart2->est_marque()))
        {
                if ((depart->retourne_identite())->egal_var(
                                                depart2->retourne_identite())
                  )
                {
                        depart2->marquer();
                        f = mc2->retourne_valeur()->retourne_force();
                        if (f>max) max=f;
                        liste_resultat->ajouter3(mc2);
                        trouve = 0;
                        float temp1,temp2;

                        //***************************
                        // La valeur de la connection est rangée
                        // dans un tableau ordonné.
                        //***************************
                        for(int i=1;i<=t[0];i++)
                        {
                                if ((f>t[i]) && (trouve==0))
                                {
                                        trouve = 1;
                                        temp1 = t[i];
                                        //on decale tout a droite et on le place
                                        for(int j=i;j<=t[0];j++)
                                        {
                                                temp2 = t[j+1];
                                                t[j+1] = temp1;
                                                temp1 = temp2;
                                                }
                                        t[i]=f;
                                }
                        }
                        t[0] = t[0]+1;
                        // On ne l'a pas trouve
                        if (trouve == 0) t[(int)t[0]]=f;
```

```
                        }
                    }
                }
                mc2 = mc2->retourne_suivant();
            }//WHILE

        //********************
        // Calcul de la spécificité
        //********************
        Spe = 0;
        for(int i=1;i<=t[0];i++)
        {
            Spe = Spe -(t[i]/max);
        }

        Spe = 1 + ((Spe +1)/(t[0]-1));
        //***********************

        //***********************
        // Impression
        if (Spe > 0.2)
        {
            *fichier<<"-----------------------------"<<"\n"<<flush;
            *fichier<<" SPECIFICITE: "<<Spe<<"\n"<<flush;
            liste_resultat->imprimer(fichier);
            *fichier<<"-----------------------------"<<"\n"<<flush;
        }
        //***********************

        liste_resultat->detruire_coquille();


        }//IF MARQUE
        mc = mc->retourne_suivant();
    }//WHILE GEN

    mc = entrees->retourne_tete();
    while (mc != 0)
    {
        mc->retourne_valeur()->retourne_depart()->demarquer();
        mc = mc->retourne_suivant();
    }
    delete liste_resultat;
}
```