



## **Thèse**

**présentée pour obtenir le titre de docteur  
de l'École Nationale Supérieure des Télécommunications**

**Spécialité : Informatique et Réseaux**

**Jean-Louis Dessalles**

**Modèle cognitif de la communication spontanée,  
appliqué à l'apprentissage des concepts**

TELECOM Paris 93-E-022

Soutenue le 22 novembre 1993 devant le jury composé de

Jean-Pierre Desclés	Président
Paul Bourguine	Rapporteurs
Daniel Kayser	
Pierre Falzon	Examineurs
Alain Grumbach	
Martial Vivet	

**École Nationale Supérieure des Télécommunications**



**© Copyright  
Jean-Louis Dessalles  
1993**



*Ce travail est dédié à tous  
ceux qui voudront bien le critiquer.*



## Remerciements

Je remercie MM. les membres du Jury : Jean-Pierre Desclés, Professeur à l'Université Paris IV, qui a accepté de le présider ; Paul Bourguine, Directeur de Recherche au CEMAGREF et Daniel Kayser, Professeur à l'Université Paris XIII, qui ont bien voulu rapporter cette thèse ; Pierre Falzon, Professeur au CNAM et Martial Vivet, Professeur à l'Université du Maine (Le Mans) qui m'ont fait l'honneur d'être examinateurs. Je les remercie d'avoir considéré mon travail, de m'avoir donné leur avis détaillé et de continuer à m'accorder leur confiance.

Je remercie Alain Grumbach, Professeur à TELECOM-Paris, pour avoir accepté d'être le directeur de cette thèse, pour sa patience et sa grande tolérance lors des nombreuses discussions qui m'ont obligé à préciser mes idées.

Je veux remercier ici toutes les personnes qui, par leur aide directe ou indirecte, ont contribué à l'aboutissement de ce travail. En premier lieu Lamia Choukair, qui a relu chaque détail de cette thèse et qui, par ses corrections, ses remarques, ses conseils et ses encouragements, m'a donné une confiance que je saurai garder.

Merci à toutes les personnes que je suis allé déranger pendant les différentes phases de ce travail. Olivier Hudry, Irène Charon, Gérard Cohen, Cédric Chappelier et bien sûr Eric Bonabeau. Merci aussi à Georges Sallé, à Nathalie Le Ruyet et à Francette Delbor pour leur bienveillance à mon égard. Je remercie enfin toutes les personnes à qui j'ai parlé de ma recherche et qui m'ont écouté.

Je veux me souvenir ici de mes débuts à TELECOM-Paris, de la confiance que m'a accordée Daniel Cadé, et de tout ce que j'ai appris en travaillant avec Jean-Pierre Bach et l'équipe du Centre de Calcul lors du développement du système SAVANT.

Je repense aussi aux heures de travail prises sur le temps que j'aurais dû consacrer à ma famille.

Il est d'usage, au début d'une thèse, de remercier ceux qui ont donné le jour à son auteur. Mes parents m'ont donné bien davantage. Je leur dois mon goût pour la science et la recherche, et bien des idées qui se retrouvent indirectement dans le travail que je présente ici.



# **Introduction : appliquer les règles de la conversation à l'apprentissage**



# Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Comprendre les aspects supérieurs du langage

Les êtres humains éprouvent un besoin fondamental de communiquer, et y consacrent une part importante de leur temps. Grâce à la technologie, ils peuvent communiquer davantage d'informations, à travers l'espace, à travers le temps, à tel point qu'il est banal de dire que nous entrons dans l'ère de la Communication.

Pourtant, la communication, poussée à ce niveau de complexité et d'importance, semble aussi ancienne que notre espèce. La communication par un langage appris est caractéristique du genre *homo*. Nous suivons Jaques Monod [1970] en estimant que nous devons au développement du langage le processus même selon lequel l'intelligence humaine a été façonnée par la sélection naturelle.

Si le langage est aussi fondamental pour nous, comment ne pas s'étonner que la communication reste aussi mystérieuse ?

Certes, les progrès des sciences du langage ont commencé à démonter les mécanismes de la phonologie, de la syntaxe, de certaines règles qui nous permettent d'attribuer un sens à des énoncés. Mais même si une partie des mécanismes du langage commence à être mieux comprise, nous ne savons toujours pas expliquer *pourquoi* les humains communiquent, nous ne savons rien des critères qui leur dictent le *choix des informations* qu'ils vont transmettre. Pour oser une analogie, les sciences du langage sont dans la position du chimiste qui analyse un vin en termes de molécules, qui sait vaguement que ce vin sera bu et procurera du plaisir, mais qui serait totalement incapable d'établir un lien entre les molécules qu'il trouve et la complexité de l'arôme.

Notre ignorance face au phénomène du langage est doublement gênante. Il est gênant, tout d'abord, d'être incapable d'analyser un trait constitutif de notre humanité. Certes ce n'est pas le seul qui résiste à notre compréhension, et beaucoup considèrent que c'est un problème complexe réservé à quelques scientifiques spécialisés.

Mais notre ignorance concernant la fonction du langage a un deuxième effet gênant, qui va devenir de plus en plus flagrant dans les prochaines années. Nous sommes maintenant en mesure de construire des systèmes capables de produire des raisonnements. Nous dépendrons de plus en plus de ces systèmes, car ils disposeront de connaissances que nous n'aurons pas et qui seront essentielles pour notre vie quotidienne. Or ces systèmes seront inexploitables s'ils ne savent pas communiquer leurs connaissances et leurs raisonnements sous une forme qui soit humainement admissible. Ce problème était passé inaperçu à l'époque du développement des premiers systèmes experts. On pensait que communiquer la trace d'un raisonnement suffirait à le faire accepter. On comprend maintenant que cet aspect de la communication humain / machine

est fort complexe, qu'il passe nécessairement par le langage, et que les humains sont fort peu adaptables pour ce qui touche à la communication des connaissances et des raisonnements.

L'ordinateur de la dernière décennie de ce siècle est avant tout un outil sophistiqué, mais qui ne réagit qu'à des ordres simples, donnés par exemple au moyen d'une souris. L'ordinateur du siècle prochain sera certainement capable de se comporter comme un collaborateur. Il prendra en compte des consignes complexes données verbalement, et il sera capable de les commenter, de les critiquer, bref, d'*argumenter*.

A l'heure actuelle, très peu de chercheurs se sont posé le problème de savoir comment fonctionne l'argumentation humaine, et comment nous pourrions en reproduire les mécanismes dans des machines. Certains sous-estiment le problème, d'autres sont effrayés par sa complexité. Nous nous sommes engagés sur ce domaine de recherche, et le document que nous livrons ici présente nos premiers résultats.

## 2. Etudier la communication spontanée

Nous sommes partis de l'étude de la communication spontanée. La plus grande partie de notre comportement communicant s'exerce à l'occasion des conversations. Nous avons estimé que la conversation "décontractée", non ciblée vers l'exécution d'une quelconque tâche, était la situation par excellence dans laquelle nous pourrions analyser les mécanismes argumentatifs. C'est dans ces situations spontanées que l'on peut le plus facilement comprendre pourquoi les interlocuteurs se mettent à parler et pourquoi ils choisissent de dire ce qu'ils disent.

Notre modèle de la conversation, forgé à partir de l'analyse minutieuse de nombreux extraits réels, s'exprime en un certain nombre de contraintes. Nous lui consacrons la première partie de ce document. La plus spectaculaire de ces contraintes est sans doute celle qui oblige le premier locuteur, celui qui introduit un nouveau sujet, à présenter ce sujet comme problématique (bizarre, incroyable ou indésirable). Il est remarquable que tout autre mode de présentation sera conversationnellement sanctionné (réaction du type "Et alors?" ou "Pourquoi dis-tu cela?"). La suite de la conversation est elle aussi fortement contrainte. Nous analysons toutes ces contraintes dans le chapitre I.1. Dans le chapitre I.2, nous montrons que ces contraintes sont absentes des principales théories pragmatiques, et que ces dernières sont donc très insuffisantes pour décrire la communication spontanée.

Bien entendu, ces contraintes que nous avons pu mettre en évidence laissent encore la place à une certaine liberté d'expression. Elles sont loin de pouvoir réglementer à elles seules le comportement langagier. Pourtant, dans certaines situations favorables où la connaissance des interlocuteurs est connue avec précision, ces contraintes sont suffisamment limitatives pour permettre une reconstruction de la conversation. Notre programme PARADISE est ainsi en mesure de valider notre modèle en synthétisant une conversation qui ressemble à la conversation réelle (chap. I.3). PARADISE utilise pour cela une connaissance totalement déclarative, constituée par un ensemble non ordonné de relations logiques. Il est capable de calculer le "bon" argument qui sera pertinent à un moment donné de l'échange, en suivant quelques stratégies simples et systématiques.

## 3. Apprendre les connaissances qui peuvent être communiquées

Ce modèle de la communication spontanée rend très bien compte des données recueillies et explique une partie importante du phénomène conversationnel. Ce faisant, il ouvre un certain nombre de perspectives. Nous avons choisi d'explorer un domaine d'application particulièrement important à nos yeux : celui de l'apprentissage. Notre objectif final est d'enseigner des connaissances grâce à une conversation avec une machine. Mais il est essentiel, pour atteindre cet objectif, de comprendre auparavant la *nature* des connaissances qui peuvent être ainsi communiquées, ainsi que le processus par lequel de telles connaissances sont *apprises*.

Il se trouve que la compréhension de certains mécanismes à l'oeuvre dans les conversations nous permet d'analyser un type d'apprentissage qui a été très largement délaissé par les théories psychologiques, bien qu'il soit d'une importance majeure : *l'apprentissage des connaissances par le langage*. Les conversations constituent en tant que telles une situation d'apprentissage. Nous avons acquis la majeure partie de nos connaissances sur notre environnement matériel et humain à l'occasion de conversations. Notre thèse est que les processus conversationnels et les processus d'apprentissage des connaissances sont intimement liés.

Nous consacrons le chapitre II.1 à une revue et à une classification des principaux modèles de l'apprentissage. Ceci nous permet, dans le chapitre II.2, de caractériser le mode de fonctionnement préconisé par la plupart de ces modèles. Ces modèles, qui incluent l'apprentissage par essais-erreurs ou les systèmes connexionnistes courants, décrivent des mécanismes d'*extraction* qui sont de type "indifférent" : le comportement d'un tel système apprenant est indifférent à tout changement de repère dans lequel les entrées du système sont représentées. Nous montrons que de tels systèmes indifférents sont incapables de rendre compte de l'acquisition des connaissances dont les individus montrent la maîtrise dans leurs conversations.

Nous sommes donc amenés à proposer un autre mécanisme, un mécanisme non indifférent, pour rendre compte de l'apprentissage des connaissances. Nous suggérons dans le chapitre II.3 qu'un mécanisme d'*appariement* explique correctement la formation de nouvelles connaissances qui pourront être manipulées par le langage.

Notre modèle de l'apprentissage des connaissances qui seront manipulées par le langage a d'importantes implications cognitives. Nous sommes conduits, toujours dans le chapitre II.3, à postuler l'existence d'un appareil spécifique au traitement des concepts. La situation conversationnelle, caractérisée par la perception d'une relation problématique entre des concepts (incohérence, improbabilité, indésirabilité) fournit un contexte adéquat à la formation de nouveaux concepts et de nouvelles relations par cet appareil conceptuel.

L'existence de cet appareil conceptuel est centrale dans notre modélisation. Il rend possible la détection des situations problématiques telle que l'effectuent les interlocuteurs. Il permet d'expliquer l'acquisition de connaissances nouvelles "en un coup" telle qu'on l'observe dans les conversations. Pourtant, de nombreux auteurs refusent d'envisager l'existence d'un tel appareil parce qu'il leur semble non plausible d'un point de vue neuronal. Nous consacrons le chapitre II.4 à montrer qu'au contraire, il est possible de suggérer une implantation neuronale de l'appareil conceptuel qui repose sur des principes simples comme la synchronisation de l'activité neuronale et sur un codage spatio-ensembliste des concepts.

## 4. Enseigner les concepts par une conversation avec un ordinateur

Après avoir ainsi justifié, sur le plan théorique, l'intérêt de la situation conversationnelle pour l'apprentissage des connaissances, nous proposons un système d'Enseignement Assisté par Ordinateur qui fonctionne sur les principes argumentatifs que nous avons analysés.

Nous consacrons le chapitre III.1 à caractériser ce qu'un tel système peut enseigner. Alors que la plupart des systèmes d'enseignement ont été conçus pour faire acquérir des *savoir-faire*, nous montrons que l'acquisition des connaissances *conceptuelles* réclame des systèmes reposant sur des principes différents. Alors que la mesure de la performance et la rétroaction peuvent être utiles pour l'apprentissage de procédures, elles sont sans effet direct sur l'apprentissage des connaissances conceptuelles. En revanche, ce dernier peut être grandement facilité par des mécanismes argumentatifs comme ceux qui opèrent dans les conversations.

Notre système SAVANT3, décrit dans le chapitre III.2, argumente avec l'apprenant sur des sujets techniques. Son intérêt pédagogique est de ne pas faire appel aux notions de fausseté ou de vérité des réponses. La seule chose que doit éviter l'apprenant, c'est l'incohérence. Comme dans certains types de

conversations, l'un des interlocuteurs (ici c'est SAVANT3) marque son étonnement, et l'autre (l'élève) s'attache à montrer que la situation qui résulte de ses choix est cohérente.

L'intérêt technique de SAVANT3 vient du fait qu'il gère lui-même le dialogue à partir d'une connaissance assez simple. L'auteur du cours n'a pas à se soucier de l'ordonnement de questions. Il se contente d'indiquer au système la connaissance relative au contexte du dialogue, et ceci d'une manière totalement déclarative. Notre espoir est de pouvoir amener l'auteur du cours à fournir cette connaissance à la faveur d'une interaction qui serait elle aussi de type conversationnel ! Nous décrivons nos tentatives dans ce sens dans le chapitre III.3.

L'application pédagogique des principes conversationnels représente à elle seule un enjeu significatif. C'est pourquoi nous l'avons explorée en priorité, en développant SAVANT3. Toutefois, dans le chapitre III.4, nous montrons comment les mêmes principes peuvent s'appliquer à un domaine lui aussi très important: celui de l'*explication* dans les systèmes à base de connaissances. Nous tentons de montrer qu'une explication ne peut être pertinente et satisfaisante que si la requête (réelle ou anticipée) qui l'a provoquée est correctement analysée comme un signal d'incohérence logique. Chaque élément d'explication doit donc être calculé et présenté comme le sont les arguments d'une conversation. L'explication émerge au total d'une interaction, et cette interaction doit obéir selon nous aux lois de l'argumentation conversationnelle.

Dans une dernière partie, nous envisageons les perspectives techniques et théoriques ouvertes par l'analyse des conversations. Nous montrons tour à tour l'intérêt d'appliquer les principes conversationnels aux domaines suivants:

- la génération de texte
- les systèmes critiques
- l'acquisition des connaissances

Nous donnons de plus des indications qui montrent comment des mécanismes de type conversationnel peuvent être à la base du processus de *conception* en résolution de problème. Nous suggérons aussi, dans un cadre *épistémologique*, qu'ils puissent donner une bonne description du processus de modélisation en sciences. Enfin, nous indiquons comment la connaissance explicite des contraintes conversationnelles peut permettre d'analyser les discours pathologiques.

Nous avons étudié un domaine, celui de la communication spontanée, qui était resté paradoxalement très peu exploré. L'objectif de ce document est de montrer que le processus humain de communication peut être analysé, et que son analyse a de nombreuses conséquences théoriques et techniques. Notre modélisation ne peut être bien entendu que partielle, mais nous espérons susciter des recherches qui viendront largement compléter le travail que nous présentons ici.

# **I- Modélisation des conversations naturelles**

<b>I.1</b>	<b>Les contraintes logiques des conversations spontanées</b>	<b>11</b>
<b>I.2</b>	<b>Aspects de la pragmatique conversationnelle</b>	<b>57</b>
<b>I.3</b>	<b>Un modèle computationnel de l'argumentation</b>	<b>67</b>



# Introduction à la partie I

## Modèle de la communication spontanée

*Les conversations seraient donc logiques...*

Notre étude révèle un phénomène étonnant : c'est dans les situations les plus "décontractées" que la communication humaine serait la plus contrainte. L'analyse des conversations quotidiennes montre en effet que les interlocuteurs s'imposent mutuellement des règles très strictes sur la *logique* de ce qui est dit.

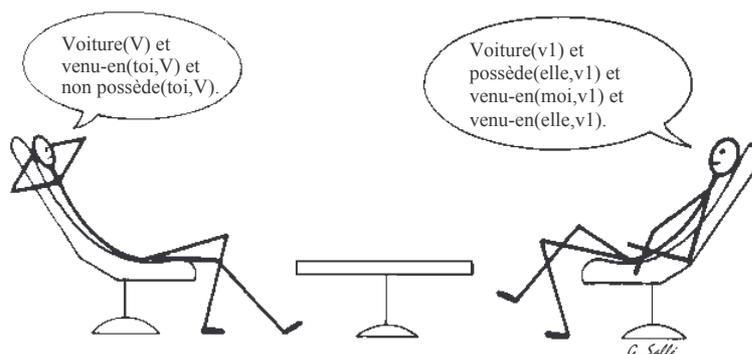
La plus spectaculaire de ces contraintes concerne la première intervention, celle qui introduit un nouveau sujet de conversation. Cette intervention doit obligatoirement être relative à un fait problématique. Les interlocuteurs doivent être capables de relier l'intervention à une situation

- ◆ apparemment incohérente, ou
- ◆ hautement improbable, ou
- ◆ hautement indésirable (ou bien, de manière duale, hautement désirable)

sinon le premier locuteur se voit "sanctionné" par un "Et alors ?", un "Pourquoi dis-tu ça", voire par une réaction agressive.

Les contraintes logiques que nous décrivons ne portent pas seulement sur la première réplique. Elles portent sur l'ensemble de la conversation. Elles sont tellement limitatives qu'elles permettent dans certains cas de prévoir ce qui va être dit par un locuteur à partir du moment où l'on a accès à ce qu'il sait.

- Le chapitre I.1 expose le modèle. Les trois modes conversationnels y sont détaillés avec de nombreux extraits de conversations réelles.



- *Le chapitre I.2 situe ce modèle par rapport aux théories les plus connues de la pragmatique. Nous montrons comment notre modèle vient combler, pour les conversations, certaines lacunes de ces théories et comment il s'oppose à ces théories sur certains points importants.*
- *Dans le chapitre I.3, nous cherchons à valider notre modèle de la communication spontanée en montrant qu'il est possible de reconstruire la dynamique d'un échange conversationnel, à partir des connaissances des interlocuteurs relatives au sujet de la conversation. Nous décrivons le fonctionnement de notre programme PARADISE qui réalise cette reconstruction*

# **Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage**

## **1. Introduction**

Parmi les études qui portent sur l'usage naturel du langage, il y en a peu qui font appel à la logique pour décrire le contenu des répliques successives et pour expliquer pourquoi tous les énoncés ne sont pas admissibles à un moment donné de la conversation. L'étude de conversations spontanées au niveau logique révèle que les interlocuteurs respectent de manière non consciente des contraintes très spécifiques au moment d'introduire un nouveau sujet. Lorsque ces contraintes sont violées, soit expérimentalement soit à la suite d'un malentendu, on observe des réactions d'incompréhension du type "Pourquoi dis-tu ça ?" ou "Et alors ?".

Le modèle de conversation, que nous allons décrire ici à l'aide d'extraits réels, n'autorise que trois manières différentes d'introduire un nouveau sujet au cours d'une conversation libre. De plus, selon ce modèle, toute réplique doit avoir un effet logique sur la connaissance supposée de l'interlocuteur, et la liste de ces effets est extraordinairement limitée.

Cette description de la conversation comme une séquence d'actions logiques est suffisamment précise pour permettre à un programme informatique de reproduire le chaînage dynamique des arguments en utilisant simplement la connaissance que les interlocuteurs possèdent sur le sujet (chap. I.3).

## **2. L'usage naturel du langage naturel**

Chacun participe maintes fois chaque jour à des conversations. Lorsque nous nous arrêtons de parler et tendons l'oreille, nous pouvons parfois remarquer l'étonnant niveau sonore produit par les émissions vocales humaines. Nous consacrons beaucoup de temps à cette activité de conversation qui semble être caractéristique de notre espèce. Il est donc surprenant d'observer un nombre limité d'études consacrées aux conversations spontanées, si l'on compare avec l'énorme quantité d'articles portant sur d'autres aspects du langage naturel (par ex. la syntaxe ou la sémantique). Plusieurs auteurs considèrent de plus que la plupart des modèles portant sur des aspects "supérieurs" du langage sont loin d'être satisfaisants (par ex. [Trognon & Brassac 1988:212]). Les caractéristiques des conversations sont-elles toutes si complexes au point d'apparaître totalement non contraintes ?

Goodwin & Heritage [1990] expliquent l'intérêt tardif des scientifiques pour l'étude des conversations par le fait que

"both sociology and linguistics defined the scope of their subject matter in such a way that the relevance of talk-in-interaction fell between disciplinary boundaries".

H. Bunt [1991] montre la difficulté de la tâche de modélisation des conversations :

"dialogue theories need other devices than, say, constituent structure diagrams and truth-conditional semantic rules. Instruments are needed to model such things as what each of the partners knows and believes and what communicative intentions they pursue. The representation of such things and how they can be used by an intelligent linguistic agent to perform successfully in a dialogue is far beyond the means of traditional linguistic theories".

Les chercheurs ont commencé à se préoccuper des conversations en tant que telles il y a deux décennies, lorsqu'il est devenu plus aisé de les enregistrer et donc de travailler sur des données fiables. Ce point est fondamental et inhabituel car, lorsqu'ils étudient d'autres niveaux du langage comme la syntaxe, les linguistes tirent la plupart du temps leurs exemples de leur propre capacité de production langagière. Ils travaillent sur des phrases qui n'ont sans doute jamais été énoncées par quiconque dans une situation de la vie courante. Une telle approche peut être parfaitement justifiée pour une étude du niveau syntaxique. Toutefois, l'absence de corpus peut rendre impossible la découverte de certaines régularités.

Certains chercheurs étudient les aspects supérieurs du langage sans faire appel à des corpus, comme l'Ecole de l'Analyse du Discours (cf. [Coulthard 1977]) et de nombreuses approches qui utilisent le paradigme des Actes de Langage [Searle 1969]. Compte tenu de nos objectifs, une telle limitation nous a semblé constituer un obstacle.

Notre approche se rattache davantage au paradigme de l'*Analyse des Conversations* (Conversational Analysis) pour au moins une raison : nous considérons la conversation libre comme un comportement naturel qui doit être étudié sous sa forme spontanée, et donc nous prenons le parti de ne travailler que sur des données recueillies en situation réelle. Tous les extraits donnés ici correspondent à des conversations qui ont effectivement eu lieu (pour des discussions sur les différences entre ces diverses écoles : Analyse du Discours, Actes de langage, Analyse des conversations, voir [Reichman 1989], [Goodwin & Heritage 1990], [Moeschler 1990], [Norman & Thomas 1991]).

Les conversations peuvent être étudiées depuis des points de vue très différents. Beaucoup d'auteurs se préoccupent de leurs aspects sociologiques : les règles régissant les prises de parole (les indices qui nous permettent de savoir que nous pouvons ou même devons prendre la parole à un moment donné), les styles conversationnels (qui décrivent des différences systématiques entre participants), l'expression des relations sociales pendant la conversation (par ex. la dominance, l'agressivité, la politesse, la coopération), les types d'actes sociaux qui peuvent être accomplis pendant la conversation (actes perlocutoires), etc.

D'autres auteurs sont davantage concernés par des caractéristiques situées à des niveaux de détail plus fins : interruptions, intonation, hésitations, etc.

Certains auteurs étudient les relations entre structure et fonction dans les conversations : la façon dont les explications, les plaintes, etc. sont exprimées à travers les répliques, la façon dont les décisions ou la vérité sont négociées au cours de l'interaction, et ainsi de suite. Par exemple Drew & Holt [1990] montrent comment des expressions idiomatiques sont utilisées dans certaines circonstances pour exprimer et résumer une plainte au cours de la conversation; Reichman [1985] se propose d'identifier une "structure profonde" des conversations en termes de relations structurelles entre les éléments du discours; Heritage [1990] montre comment les explications sont exprimées en fonction de la situation sociale.

Notre approche s'inscrit plutôt dans ce dernier courant. Dans la présente étude, nous nous préoccuons du contenu des énoncés, en essayant de décrire comment ils sont fonctionnellement reliés les uns aux autres et avec le sujet de la conversation. Notre but n'est pas de décrire des aspects sociaux<sup>(1)</sup>, et les détails de la formulation des énoncés ne nous intéressent ici que dans la mesure où ils peuvent nous aider à comprendre sans ambiguïté la signification des répliques et de leur contexte. Nous voulons parvenir à une modélisation de certains aspects de l'argumentation qui fasse apparaître leurs contraintes logiques. Ce modèle nous

<sup>1</sup> Notre préoccupation ici porte, comme dit A. Trognon, sur les conversations, non sur les personnes engagées dans la conversation. Ce choix va avoir de multiples conséquences pour ce travail. Par exemple, il est souvent dit que les interlocuteurs coopèrent au cours de la conversation. Mais nous montrerons que les répliques ont le plus souvent un effet de "destruction" sur d'autres répliques, même parfois lorsque les unes et les autres émanent du même locuteur!

permettra de faire des prédictions sur ce qu'un énoncé peut ou ne peut pas exprimer à un moment donné d'une conversation.

Il nous faut préciser le type de conversation que nous considérons ici. Une position extrême serait de considérer que tout emploi spontané du langage naturel doit être considéré comme une conversation. Notre ambition est bien entendu beaucoup plus limitée. Nous cherchons les contraintes que les interlocuteurs semblent respecter dans les conditions suivantes :

- interlocuteurs ayant le même statut
- interlocuteurs décontractés
- interlocuteurs se connaissant bien
- interlocuteurs se rencontrant fréquemment
- interlocuteurs n'étant pas engagés dans un dialogue finalisé (*task-oriented*)
- interlocuteurs étant (au moins momentanément) sérieux

On peut résumer ceci en disant que nous considérons les conversations "banales". Comme nous le verrons, elles semblent soumises à davantage de contraintes que les dialogues plus formels (cf. [Dessalles 1992b]). Parmi les contre-exemples que nous éviterons ici, citons une dispute de couple (comme dans [Schank & Lehnert 1979]), des dialogues professionnels (comme les dialogues opératifs étudiés par exemple par Falzon [1991]), ou encore des interactions très courtes comme "Quelle heure est-il s'il vous plaît ?".

Nos propres données sont d'une nature différente : elles consistent en plus de 30 heures de conversations familiales réunissant le plus souvent plus de trois locuteurs adultes. Cette approche de la conversation se veut "écologique" dans la mesure où elle part de l'observation désintéressée d'humains en train de se comporter naturellement. Certains enregistrements ont été effectués à l'insu des protagonistes, mais la plupart du temps l'appareil d'enregistrement était placé sur la table et était rapidement oublié, y compris par l'observateur. Il était présenté comme un moyen de prendre une "photographie sonore".

J'étais toujours présent et j'ai souvent pris part aux conversations ainsi enregistrées. On peut s'étonner de ce manque de distanciation. Mais cela était rendu nécessaire pour au moins deux raisons : premièrement, cette attitude est compréhensible d'un point de vue éthique, et deuxièmement, comme nous le verrons, c'est le meilleur moyen d'avoir *une connaissance parfaite du contexte*, et c'était à mes yeux la seule façon de parvenir à saisir certaines des règles qui gouvernent les conversations. Mais je prendrai aussi des exemples dans d'autres corpus lorsque cela sera possible.

Il faut insister sur ce dernier point, la connaissance du contexte: de nombreux auteurs (par ex. [Sperber & Wilson 1986]) considèrent que le principal rôle de la Pragmatique est de trouver un moyen objectif de déterminer ce qu'est le contexte pertinent. Nous souscrivons en partie à ce point de vue, mais notre travail prétend aller plus loin que la reconnaissance du contexte (tout en s'inscrivant clairement dans la recherche en Pragmatique, cf. chap. I.2) : notre étude des aspects logiques de l'argumentation dans les conversations spontanées commence à partir du moment où le contexte est connu. Comme il n'existe pas actuellement de moyen d'extraire la connaissance contextuelle, nous étudions exclusivement des conversations pour lesquelles le contexte est connu sans ambiguïté.

*Résumé* : la conversation est un comportement naturel, mais peu d'auteurs l'étudient comme telle, en essayant de trouver une structure dans le contenu des échanges. Nous travaillons sur des données réelles pour lesquelles le contexte est parfaitement connu.

### 3. Objectif de l'étude des conversations : découvrir des contraintes

L'un des objectifs principaux de la recherche en Pragmatique, selon nous, est de déterminer l'ensemble des contraintes qui limitent les choix des interlocuteurs dans leurs interactions verbales. Si tout peut être dit, il n'y a pas de recherche pragmatique possible. Or à un moment donné d'une interaction, il apparaît que

l'éventail des possibilités qui s'offre à un interlocuteur, quoique large, est extraordinairement limité par rapport à tout ce qui est imaginable. Un excellent indicateur de cette limitation est l'existence des *cassures* (comme "Pourquoi dis-tu ça ?", "Et alors ?"). Ces cassures, qui restent relativement rares dans les interactions spontanées que nous avons étudiées, sont révélatrices d'une contrainte qui, pour une raison ou une autre (par exemple un malentendu) a été violée.

Notre objectif dans cette étude est de proposer un modèle de certaines contraintes qui limitent logiquement ce qui peut être dit pendant une conversation. L'observation du type de conversation que nous avons choisi de considérer (cf. ci-dessus), et pour lesquelles le contexte nous est parfaitement connu, va nous permettre de montrer que, dès leur début, lors de l'introduction d'un nouveau sujet, les conversations sont logiquement contraintes. Un nouveau sujet doit apparaître comme problématique : soit paradoxal, soit improbable, soit hautement désirable ou indésirable.

Nous allons donner une description logique de ce que nous appellerons le "contexte logique", cette partie de la connaissance partagée qui rend le nouveau sujet problématique; puis nous montrerons que les répliques qui suivent ont un effet logique sur ce contexte. Le modèle n'autorisera qu'un petit nombre de possibilités : invalidations logiques, "banalisation" et réactions "antagonistes".

Nous donnerons quelques indications sur la façon dont les répliques ultérieures peuvent logiquement s'enchaîner en fonction de leur type au cours de la conversation. Dans le chapitre I.3, nous testerons le modèle en donnant une rapide description d'un programme, PARADISE, qui est capable de reconstituer cet enchaînement de répliques, comblant ainsi le fossé qui sépare une simple connaissance logique statique de l'argumentation.

#### 4. Contraintes et conversation spontanée

Notre objectif ici est de décrire certaines des contraintes qui restreignent, au niveau logique, la liberté des interlocuteurs au cours d'une conversation spontanée, sans qu'ils en soient nécessairement conscients. Cela peut surprendre, car il est inhabituel de considérer les conversations quotidiennes comme soumises à des contraintes de ce type. Certes on admet volontiers que tout n'est pas bon à dire au cours d'une conversation. Chacun a expérimenté la gêne causée par un changement abrupt de sujet, accompagnée de réactions comme : "Qu'est-ce que ça a à voir avec ce que nous sommes en train de dire ?" ou de réactions d'incompréhension ("remedial responses"), comme cela a été montré dans les expériences de Vuchinich [1980]. Mais on oublie souvent qu'il existe bien d'autres contraintes, comme le montre par exemple l'extrait suivant :

[ex\_repas]

contexte : A a bénéficié d'un repas amélioré sur son lieu de travail, comme le veut la coutume juste avant Noël. Il pensait qu'il en était de même pour B.

A1- *Et toi, ça va ? Tu as bien mangé, à midi ?*

B1- *Pourquoi tu me demandes ça ?*

La réplique de B fut émise comme une protestation. Une question comme A1 semble être inadmissible hors contexte, comme c'est le cas ici en ce début de conversation. Chaque fois que nous observons une réaction comme "Pourquoi tu me demandes ça ?" ou quelque réaction agressive montrant une incapacité à comprendre la motivation de la réplique précédente, nous pouvons conclure qu'une règle conversationnelle a été violée. Dans l'extrait suivant, la protestation est exprimée tout d'abord à travers une grimace, puis par une question sarcastique :

[ex\_train]

contexte : A (l'auteur) rentre chez lui en train chaque jour. Il énonce A1 intentionnellement comme une expérience, pour observer la réaction de B.

A1- *J'ai pris le train.*

B1- *[grimace]... c'est un exploit ?...*

Dans ces exemples, le problème conversationnel qui apparaît entre les deux interlocuteurs ne peut pas être la conséquence d'un quelconque malentendu. Dans le premier extrait, il aurait été facile pour B de simplement donner l'information requise. B ne l'a finalement jamais fait pendant le reste de la conversation. Dans le deuxième, B aurait pu simplement "accuser réception" de l'assertion toute simple A1. Ceci indique, si nécessaire, que la conversation *n'est pas un simple échange d'informations*, mais une autre sorte de jeu bien plus complexe qui possède ses règles.

Il est surprenant que de telles contraintes, qui affectent le *contenu* de ce qui peut être dit à un moment donné d'une conversation, indépendamment des conventions sociales, n'aient pas davantage retenu l'attention des chercheurs. Or il est permis d'espérer que des résultats dans l'étude de ces contraintes pourront conduire à une meilleure compréhension des processus du langage naturel, ainsi qu'à la conception d'interfaces humain/machine plus efficaces.

Dans son célèbre papier "Logic and Conversation" [Grice 1975], H.P. Grice mentionne plusieurs "maximes" que les interlocuteurs se doivent normalement d'observer (à moins qu'ils cherchent à "impliciter", c'est-à-dire à insinuer, un fait ou une remarque). L'une d'entre elles se résume simplement à "Soyez pertinents" (cf. chap. I.2). Le principal objectif de ce travail est de tenter de définir plus précisément certains aspects logiques de ce concept de pertinence, et d'explorer toutes les manières d'être logiquement pertinent.

## 5. Les conversations spontanées sont-elles logiques ?

Notre tentative pour modéliser les conversations nous a conduits naturellement à faire usage de la logique. Mais les conversations humaines ne sont pas habituellement considérées comme logiques : "Si les gens se comportaient logiquement durant les conversations, tout désaccord serait impossible !". On entend souvent ce jugement simpliste, ou d'autres du même type : "les gens sont illogiques", ou "il raisonne avec sa propre logique". Les sujets de conversations ont peu de chose à voir avec la vérité. Lorsque nous parlons de beauté, d'amour, d'art ou de vin, la logique semble souvent absente. Moeschler, considérant des exemples comme "Je suis en retard, mais je prends un café", conclut que le langage, contrairement à la logique classique, ne se préoccupe pas des contradictions [Moeschler 1985:48] !

Mais nous percevons autrement la relation entre logique et langage : notre objectif n'est pas de considérer la Logique comme un système de maintien de la vérité. Nous ne chercherons pas à appliquer la logique aux conversations de la façon dont Russel, Carnap, Quine, Wittgenstein et beaucoup d'autres l'ont appliquée au langage. Nous allons simplement l'utiliser comme un moyen de représenter la signification conversationnelle de chaque réplique. La question ne sera jamais de savoir si une réplique donnée exprime une vérité, mais de traduire cette réplique en utilisant une syntaxe logique, comme nous le faisons pour l'extrait suivant (pris dans [Tannen 1984:62]) :

[ex\_Goffman] [Tannen 1984]

contexte : A, B et C parlent de sociologie, et il apparaît que B a une bonne connaissance des livres du sociologue Erving Goffman. A et C sont surpris, dans la mesure où ils pensaient que cet auteur n'était connu que des spécialistes.

A1- *But anyway. ... How do you happen to know his stuff ?*

B1- *Cause I read it.*

C1- *What do you do ?*

A2- [ ? ? ] are you in ... sociology or anything ?  
 B2- Yeah I read a little bit of it. [pronounced reed]  
 A3- Hm ?  
 B3- I read a little bit of it. [pronounced red]  
 A4- I mean were you... uh studying sociology ?  
 B4- No.  
 A5- You just heard about it, huh ?  
 B5- Yeah. No. I heard about it from a friend who was a sociologist, and he said read this book, it's a good book and I read that book 'n  
 A6- I had never heard about him before I started studying linguistics.  
 B6- Really ?  
 A7- Yeah.

Lors de la conversation qui précédait, B avait fait preuve d'une excellente connaissance de l'oeuvre d'Erving Goffman, ce qui paraissait étonnant dans la mesure où celle-ci est destinée à des sociologues. Pour citer D. Tannen, qui est A dans cet extrait :

"Both C and I expected B to tell how his life - and more likely his work or education - led him to Goffman's books".

Nous en savons assez pour représenter la "connaissance" sous-jacente, d'abord en français, puis en utilisant un formalisme logique :

réplique	contexte
<b>A1, A2, A4 :</b>	<i>Si quelqu'un connaît les livres d'E.Goffman, alors il doit être sociologue. B connaît les livres d'E.Goffman. B est-il sociologue ?</i>
<b>B1, B2, B3 :</b>	<i>B a lu les livres d'E.Goffman et donc il les connaît.</i>
<b>B5 :</b>	<i>B a un ami, cet ami est sociologue, il lui a recommandé les livres d'E.Goffman, et donc B connaît ces livres.</i>
<b>A6 :</b>	<i>Tant que A était dans les sciences humaines, A ne connaissait pas les livres d'E.Goffman</i>

Cette version simplifiée de l'extrait peut être représentée à l'aide d'un formalisme logique<sup>(2)</sup> :

contexte de <b>A1</b> :	<i>connaît(X, livres_de_Goffman) ⇒ sociologue(X)</i>
<b>B1, B2, B3</b> :	<i>lu(B, livres_de_Goffman) ⇒ connaît(B, livres_de_Goffman)</i>
<b>B5</b> :	<i>connaît(ami, livres_de_Goffman) &amp; sociologue(ami) &amp; recommande(ami, B, livres_de_Goffman)</i>
<b>A6</b> :	<b>non sociologue(A) ⇒ non connaît(A, livres_de_Goffman)</b>

Ce genre de traduction logique suggère d'emblée trois remarques :

- elle n'est pas unique
- elle ne représente pas la totalité de ce qui est exprimé
- certains de ses éléments ne sont pas présents dans les répliques telles qu'elles sont exprimées

<sup>2</sup> Nous utilisons ici la logique du premier ordre pour représenter cet extrait. Les prédicats sont en italique, et les variables en gras. *connaît(X, livres\_de\_Goffman)* représente le fait que X connaît les livres de Goffman. Voir annexe.

Par exemple, la représentation proposée ci-dessous ne fait pas la différence entre l'oeuvre de Goffman ("his stuff") et le livre unique mentionné en B5; les trois répliques B1, B2 et B3 sont considérées comme équivalentes; sociologie et linguistique sont confondues dans la traduction de A6, ainsi que *knows* et *read* dans le contexte de B5, etc. Le contexte de A1, tel qu'il est indiqué ici, n'est jamais exprimé pendant la conversation (noter toutefois que A6 en constitue une instanciation). Comment donc peut-on prétendre que cette représentation soit objective ?

Nous fournissons une seule réponse à ces trois objections : comme nous ne disposons pas actuellement d'un moyen de réaliser cette représentation logique algorithmiquement, nous parlons ici d'une sorte de *traduction*, faite à la main. Comme toute traduction, la représentation logique ne peut être unique, elle est partielle et fait intervenir des éléments sous-entendus. Utilisons l'analogie de la traduction entre langues différentes. Il est encore impossible de réaliser de telles traductions automatiquement (pour un sujet quelconque). Toutefois, nous utilisons les traductions des interprètes, et nous avons confiance en elles, parce que nous savons que des experts dans les deux langues tomberaient d'accord pour considérer qu'elles retiennent la majeure partie de la signification, et parce qu'elles sont toujours perfectibles lorsque la précision est requise. Nous ferons de même avec les traductions logiques. Mais la conséquence de cela est que nous devons vérifier que toute interprétation basée sur une traduction logique restera inchangée avec une autre traduction valide.

Techniquement, le formalisme que nous employons pour traduire les extraits sous une forme logique est parfaitement conforme à la logique des propositions ou à la logique des prédicats, selon les cas. La signification des connecteurs  $\&$ ,  $\Rightarrow$ , *non* est exactement celle qui leur est conférée habituellement, par exemple en mathématique. Le nom des symboles utilisés est un raccourci mnémotechnique qui définit en fait une interprétation des symboles (prédicats, univers des variables, constantes) dans le monde réel qui est celui de l'extrait conversationnel.

Par exemple, si  $[p \Rightarrow q]$  a la valeur "vraie" dans l'interprétation relative à la situation analysée, alors  $[p \& \text{non } q]$  sera considéré comme "faux" par les interlocuteurs, alors que  $[\text{non } q \Rightarrow \text{non } p]$  sera acceptée comme vraie. Si de plus  $p$  est "vraie", alors  $q$  sera considérée comme "vraie". Si en revanche c'est  $q$  qui est connue comme vraie, alors  $p$  peut être vraie ou fausse selon les cas (penser à  $p = \text{"Pierre boit de l'alcool"}$  et  $q = \text{"Pierre a plus de 18 ans"}$ ). La sémantique de l'implication est donc rigoureusement celle de la logique.

Le choix des symboles dépend de la finesse de la traduction recherchée. Par exemple, dans l'extrait ci-dessus, "*B connaît les livres de Goffman*" peut être rendu par :

*B connaît les livres de Goffman*

*connaît*( B, livres\_de\_Goffman)

*connaît*( B, L)  $\&$  *livres*(L)  $\&$  *auteur*(Goffman, L)

etc.

Le nom des symboles renvoie à leur signification. Par exemple, *connaît*(X, Y) est syntaxiquement un foncteur à deux places, et sémantiquement une relation binaire opérant sur le produit cartésien de deux ensembles. Notre utilisation de la logique est donc parfaitement "standard".

D'une certaine manière, la traduction logique est incontournable. Il n'y a par exemple aucun autre moyen de représenter l'étonnement de A et de C dans l'extrait précédent. Il est donc essentiel d'obtenir une traduction logique fiable.

Mais comment pourrait-on se fier à une traduction logique si les locuteurs commettent des fautes de logique ? L'un des résultats que confirme cette étude est que les fautes de logique caractérisées sont tout à fait rares. Je n'ai pu en noter que trois ou quatre. En voici une :

[ex\_animiste]

contexte : A et B parlent de religion. A, à titre de provocation, identifie religion et animisme, et s'étonne que de telles croyances puissent être compatibles avec le fait d'avoir effectué des études supérieures.

A1- *Je comprends pas qu'on puisse être animiste en ayant fait des études supérieures !*

B1- *Ca n'a rien à voir. Moi, j'ai pas fait d'études, et je suis pas animiste.*

Nous pouvons traduire ceci à l'aide de deux formules :

[A1] :  $études\_supérieures(X) \Rightarrow \text{non animiste}(X)$

[B1] :  $\text{non études\_supérieures}(\text{moi-même}) \& \text{non animiste}(\text{moi-même})$

En voyant cet extrait, on ne peut pas imaginer que B a voulu exprimer autre chose en B1 que la négation de A1. Ceci est fortement suggéré par le préambule de B1 : "Ca n'a rien à voir", qui est un moyen d'exprimer que  $q$  n'est pas lié à  $p$  en réponse à une affirmation de  $p \Rightarrow q$ . Toutefois la négation de A1 aurait dû être :

[B1'] :  $études\_supérieures(\text{quelqu'un}) \& \text{animiste}(\text{quelqu'un})$

Il est donc tout à fait probable que B a parlé trop vite. C'est d'ailleurs ce qu'il a reconnu quelques secondes après. Il est important de noter que cette façon de traduire les conversations sous forme logique n'a pas été corrompue par une faute logique commise par un locuteur. Au contraire, la traduction logique semble être le seul moyen de mettre cette faute en évidence.

Il est maintenant possible d'étudier les conversations spontanées d'un point de vue logique et d'observer que même la toute première intervention est hautement contrainte.

*Résumé* : sans suggérer aucunement que les interlocuteurs, en énonçant leur répliques, émettent des vérités qui seraient logiques par référence à une "Vérité absolue", nous pouvons utiliser la logique pour représenter localement le contenu des échanges.

## 6. Décrire les conversations au moyen de la logique

### 6.1. La connaissance implicite

Il n'est pas bon de tout dire lorsque l'on débute une conversation. Comme l'explique Coulthard [1977:79] :

"if one's sister becomes engaged, some relatives must be told immediately, others on a first meeting after the event, whereas some of one's friends might not know the sister or even that one has a sister, and for them the event has no importance or even interest".

En analysant une conversation entre personnes apparentées à propos de cette soeur, un observateur extérieur doit reconstituer le fait qu'elle n'est pas encore mariée, ou qu'elle a récemment changé de fiancé. De tels faits font partie de la connaissance partagée, et c'est là que réside la difficulté pour l'observateur. Cette connaissance est partagée par les interlocuteurs, c'est d'ailleurs ce qui leur permet de parler du sujet. Mais en même temps, c'est justement ce partage qui fait qu'une grande partie de la connaissance est laissée implicite !

La connaissance partagée a tellement d'importance dans la communication que des gestes à peine perceptibles ou des regards entendus peuvent tenir lieu de conversation dans certains contextes, alors qu'il aurait fallu de nombreux mots pour expliquer la situation à un étranger. L'analyse des conversations sera-t-elle jamais possible dans de telles conditions ? Il doit être clair que l'information contenue dans les mots échangés ne peut en aucun cas être suffisante pour permettre une reconstitution de la connaissance implicite. Si quelqu'un dit simplement "J'ai faim", le contexte peut inclure le fait que le restaurant est sur le point de

fermer, ou que l'interlocuteur est supposé préparer le dîner, ou que le locuteur n'a pas eu le temps de manger, etc. Cette connaissance implicite est nécessaire pour que l'analyse de l'énoncé "J'ai faim" soit possible, mais aussi simplement pour comprendre son sens qui peut être "dépêchons-nous, ça va fermer", ou "cette fois c'est à toi de faire à manger", ou "désolée je dois m'absenter pour aller manger un morceau".

Cette connaissance partagée inclut aussi, bien sûr, le sens commun et les connaissances culturelles, et aucun système artificiel n'est actuellement capable d'isoler le contexte de cette immense quantité d'informations, c'est-à-dire les éléments pertinents qui sont nécessaires pour conférer son sens à un extrait conversationnel. La pragmatique commence juste à dégager des principes et des définitions sur ce que doit inclure un contexte pertinent. Mais notre situation ici est bien plus confortable, puisque nous n'étudions que les conversations pour lesquelles nous connaissons déjà le contexte. Le problème est alors simplement d'extraire de ce contexte les quelques éléments qui suffiront pour nos besoins de modélisation.

## 6.2. Isoler le contexte logique de la connaissance partagée

Chacun est parfaitement capable et a l'habitude d'isoler la partie pertinente de la connaissance implicite pendant les conversations. Nous le faisons spontanément, mais comme pour beaucoup d'autres processus psychologiques, nous ne savons pas comment. Mais pouvons-nous être sûrs que cette connaissance implicite est partagée ? Les gens comprennent peut-être ce qu'ils souhaitent entendre, comme lors de bien des débats politiques. Bien entendu ils attribuent aux autres des capacités d'inférence qui leur permettent de deviner ce qu'ils ne leur disent pas. Mais comment peut-on être sûr qu'aucun malentendu ne subsiste ?

L'observation nous donne des indications sur ce point : lorsque des malentendus sont dissipés au cours d'une conversation spontanée, la rectification se produit très tôt après l'origine de la divergence. L'extrait qui suit donne un exemple de malentendu. Il arrive à la suite d'une conversation sur la longévité qui commença par une anecdote à propos d'un homme de 78 ans qui avait encore sa mère.

[ex\_Abraham]

contexte : la conversation précédente portait sur des gens exceptionnellement âgés qui sont en bonne santé.

A1- Hé, quel âge il avait, Abraham, quand il a procréé ?

B1- Oui, mais tu sais, on est pas très sûr ...

A2- L'état-civil n'était pas très sûr.

C1- Je sais pas si on a enregistré un record. Ca doit être assez élevé. C'est sûrement supérieur à 80 ans.

B2- Mais, de quoi tu parles ?

C2- comme ...

B3- comme géniteur ?

Lorsqu'on écoute l'enregistrement avec attention, on est convaincu que B croit que C est en train de parler en C1 d'un record de longévité. C'est B qui avait débuté la conversation sur ce sujet, et le nom d'Abraham est généralement associé avec l'idée d'un homme exceptionnellement âgé. Elle est donc tout à fait surprise d'entendre C1 de la façon suivante: "le record de longévité est certainement supérieur à 80 ans", et elle réagit avec B2. Mais elle comprend ce que C avait en tête (B3) avant même que celui-ci n'ait eu le temps de l'expliquer. Le malentendu détecté par B ne se situe ni au niveau sémantique (ce qu'elle a compris initialement avait un sens), ni au niveau du contexte pragmatique (elle se situait par rapport à un contexte, même si ce n'était pas le bon). Ce que notre modèle suggère, c'est que B détecte un malentendu logique en B2. De telles difficultés logiques sont rares, rapidement détectées et vécues comme intolérables dans les conversations usuelles.

Ceci prouve notre capacité à extraire la logique qui sous-tend les répliques de nos interlocuteurs et qui reste souvent implicite. Nous utiliserons les mots *contexte logique* pour désigner cette *partie implicite ou explicite de la connaissance partagée qui donne sa signification logique à chaque réplique*.

Le contexte logique doit apparaître de manière évidente aux interlocuteurs, sinon ils expriment leur gêne et une cassure se produit. Il est relativement aisé de provoquer ce genre de problème sciemment en commençant volontairement une conversation avec une remarque sans intérêt, comme ce fut fait dans [ex\_train] :

[ex\_train2]

contexte : A rentre chez lui en train chaque jour. Il énonce A1 intentionnellement comme une expérience, pour observer la réaction de B.

A1- *J'ai pris le train.*

B1- [grimace]... *c'est un exploit ?... Ah ! C'est parce qu'il y a des grèves.*

A1 ne présente aucun intérêt parce que A prend le train chaque jour. Il est frappant de constater que l'interlocutrice parvient à élaborer un contexte logique plausible, quoique sans fondement (il n'y avait pas de grève ce jour là), qui donne à A1 une pertinence logique : *si* il y a des grèves, *alors* les trains sont pris d'assaut, *et donc* prendre un train bondé est hautement indésirable. Mais avant cela, elle réagit de manière agressive, du fait de l'absence de tout contexte logique lui permettant de trouver une pertinence logique à A1.

Ce concept de contexte logique pourrait sembler non opératoire. Comment décider ce qui, dans la connaissance partagée, doit ou ne doit pas faire partie du contexte logique, quand ce contexte n'est pas explicité par les interlocuteurs ? Sperber et Wilson [1986] abordent cette question d'une manière qui ne peut convenir ici, car elle ne permet pas de prédire les cassures (cf. chap. I.2). L'obtention du contexte logique peut sembler constituer un problème délicat. Heureusement on peut imaginer une procédure qui donne un moyen expérimental pour l'exhiber. Dites simplement : "Et alors ?", "Pourquoi me demandez-vous ça ?", "Pourquoi dites-vous ça", ou "Où voulez-vous en venir ?", et le premier locuteur explicitera spontanément certains éléments du contexte logique de son intervention.

### 6.3. *L'expérience du "Et alors ?"*

L'expérience du "Et alors ?" peut parfois se produire spontanément, comme ici :

[ex\_Mercedes]

contexte : Sur l'autoroute de Paris vers l'Allemagne. Les grosses Mercedes sont chose courante. B ne remarque pas que la Mercedes qui vient de doubler était immatriculée en Hongrie. A cette époque, la Hongrie était communiste et les individus étaient supposés ne pas pouvoir être riches.

A1- *T'as vu la Mercedes ?*

B1- *Hé bé quoi ?*

A2- *Tu savais que les hongrois, ils avaient des voitures comme ça ?*

Les éléments donnés en A2 (les hongrois ne sont pas supposés posséder de grosses voitures) nous permettent d'isoler sans ambiguïté le contexte logique de la connaissance partagée :

*hongrois*( X )  $\Rightarrow$  **non riche**( X )

[*possède*( X, Voiture ) & *grosse*( Voiture )]  $\Rightarrow$  *riche*( X )

Parfois, le contexte logique est spontanément explicité par le premier locuteur sans réelle demande des autres participants. Nous avons pu observer ce phénomène dans [ex\_Goffman, p.15], où la représentation logique de A6 est exactement une instanciation de A1 (avec X instancié en A). Dans l'extrait qui suit, nous pouvons observer une réplique ironique provoquée par un malentendu sur le contexte, qui va être ensuite partiellement explicité par le premier locuteur.

[ex\_antenne]

contexte : en 1987 près de Paris, les antennes TV paraboliques sont rares sur les maisons privées, alors que chaque maison a son antenne "râteau".

A1- Tu sais, j'ai vu une antenne sur une maison de la rue des Roissys

B1- Dis donc, quel événement !

A2- mais non, une antenne comme ... une antenne parabolique

Il se produit quelque chose d'intéressant ici. Des deux contextes : "A a vu une antenne râteau dans le voisinage", et "A a vu une antenne parabolique dans le voisinage", seul le second rend A1 pertinent, alors que le premier, qui est ce que B a compris, provoque la réplique ironique B1. La pertinence de A1 vient clairement du fait que voir une antenne parabolique dans le voisinage est *a priori* très improbable (ce qui n'est pas le cas pour les antennes usuelles). Il semble que le formalisme logique pur soit incapable de représenter la signification de A1-A2. C'est pourquoi nous introduisons une nouvelle modalité, **IMPR**, qui dénote un fait hautement improbable. Nous obtenons une écriture du contexte qui peut être :

[ antenne\_parabolique(X) & voisinage(X) ]  $\Rightarrow$  **IMPR**

Ici (  $p \Rightarrow$  **IMPR** ) signifie simplement que  $p$  est *a priori* hautement improbable (voir annexe). Dans [ex\_Abraham, p.19], cette modalité aurait été aussi nécessaire pour exprimer le contexte logique : le fait de procréer en étant très âgé est *a priori* hautement improbable. Ici, la variable  $X$  désigne un objet générique, puisque l'estimation de la probabilité *a priori* vaut pour tout objet, et pas seulement pour l'objet spécifique aperçu par A.

Noter que les "ratés" que nous pouvons observer dans ces extraits ([ex\_Abraham, p.19] B2, [ex\_Mercedes] B1, [ex\_antenne] B1), ne proviennent pas de difficultés rencontrées par le second locuteur aux niveaux de la sémantique ou du contexte pragmatique : dans ces extraits, l'intervention précédente avait une signification non ambiguë, elle était reliée au contexte et donc convenait du point de vue de la pragmatique contextuelle. Le problème était par essence *logique*.

La plupart du temps, le contexte logique reste implicite, en particulier lorsque personne n'éprouve le besoin de demander une explicitation. Mais heureusement on peut imaginer d'effectuer l'expérience du "Et alors ?" comme une expérience de pensée (thought experiment). De cette façon un observateur au courant de la connaissance partagée peut souvent comprendre en quoi consiste le contexte logique. Dans l'exemple qui suit, nous pouvons imaginer ce que A aurait répondu si B avait dit "Pourquoi dis-tu ça ?" :

[ex\_carte\_postale]

contexte : A a acheté plusieurs cartes postales qu'il destine à des amis différents, et il a commencé à les écrire.

A1- Oh zut, je me suis trompée de carte ! J'ai écrit cette carte à C.

B1- Pff, qu'est-ce que ça peut faire ?

A a choisi ses cartes en fonction des amis auxquels il les destinait. Il réalise soudain qu'il a effectué une permutation. Ceci est indésirable, puisque chaque carte avait été achetée pour un ami spécifique. De nouveau, nous avons besoin d'une nouvelle modalité pour représenter ce contexte :

[destinée\_à(Carte, Ami1) & écrite\_à(Carte, Ami2) & (Ami1  $\neq$  Ami2)]  $\Rightarrow$  **IND**

Cette fois  $p \Rightarrow$  **IND** signifie "p est hautement indésirable". Cette modalité aurait été nécessaire aussi pour représenter le contexte logique de [ex\_train2, p.20] de la façon dont il était perçu par B après B2 : prendre un train bondé est hautement indésirable. Noter encore ici l'emploi des variables *Carte*, *Ami1* et *Ami2* pour rendre le fait que le contexte logique est plus général que la situation présente.

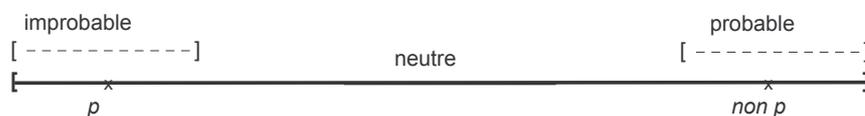
#### 6.4. La logique conversationnelle

On pourrait craindre qu'une nouvelle modalité soit nécessaire à chaque nouvel extrait. Mais l'examen des données montre que tel n'est pas le cas. En d'autres termes, trois modalités **F**, **IMPR**, **IND**, se révèlent suffisantes dans notre modèle pour représenter la signification logique des énoncés. **F** tient lieu de proposition toujours fautive et est utilisée pour réécrire la logique du premier ordre sous forme de clauses négatives, qui présentent l'avantage d'être symétriques symétriques. Ainsi ( $p1 \Rightarrow p2$ ) sera réécrit [ $p1 \& \text{non } p2$ ]  $\Rightarrow$  **F**. [ $a \& b$ ]  $\Rightarrow$  **F** signifie donc que a et b sont incompatibles. Par commodité nous utiliserons aussi la modalité **DES** (fait hautement désirable) comme synonyme de **non IND**, ainsi que **PROB** au lieu de **non IMPR**. Le tableau ci-dessous résume la sémantique de cette logique conversationnelle :

$p \Rightarrow \mathbf{F}$	p est fautive
$p \Rightarrow \mathbf{IMPR}$	p est hautement improbable
$\mathbf{PROB} \Rightarrow p$	p est hautement probable
$p \Rightarrow \mathbf{IND}$	p est hautement indésirable : l'occurrence de p suffit à rendre le locuteur insatisfait
$\mathbf{DES} \Rightarrow p$	p est hautement désirable : le locuteur ne peut pas être satisfait sans p
$p \Rightarrow \mathbf{DES}$	p est désirable : l'occurrence de p est suffisante pour rendre le locuteur satisfait
$\mathbf{IND} \Rightarrow p$	p est indésirable : le locuteur ne peut pas être insatisfait sans p

Noter qu'avec cette sémantique, les écritures  $p \Rightarrow \mathbf{PROB}$  ou  $p \Rightarrow V$  (avec  $V = \text{non } \mathbf{F}$ ) sont inutiles, car elles n'imposent aucune contrainte sur  $p$ .

Dans cette représentation, si  $p$  est improbable, alors [ $\text{non } p$ ] est probable (mais rien ne peut être dit si  $p$  n'est ni improbable, ni probable). Il nous faut insister sur ces liens entre logique et probabilité. La figure 1 montre un axe sur lequel on place chaque événement (représenté ici par une proposition  $p$ ) en fonction de sa probabilité *a priori*. Si  $p$  se situe dans la zone improbable, alors [ $\text{non } p$ ] est placé de manière symétrique dans la zone probable. Par exemple si gagner au loto est hautement improbable, ne pas gagner est hautement probable<sup>3</sup>.



----- figure 1 : axe de probabilité -----

Ce formalisme nous permet de faire la différence entre [ $p \Rightarrow \mathbf{IND}$ ] et [ $\mathbf{IND} \Rightarrow p$ ]. Nous aurons l'occasion ci-dessous de voir des exemples qui illustrent ces différents cas.

Bien sûr, [ $p \Rightarrow \mathbf{IND}$ ] équivaut syntaxiquement à [ $\mathbf{DES} \Rightarrow \text{non } p$ ] (voir annexe). La signification de ceci est que la satisfaction exige que les événements hautement indésirables soient supprimés. Inversement, [ $\mathbf{IND}$

<sup>3</sup> Cet axe ne doit pas être confondu avec un espace qui montrerait  $\{x | \text{imp}(x)\}$  et  $\{x | \neg \text{imp}(x)\}$  comme des ensembles complémentaires, où  $\text{imp}(x)$  serait un prédicat indiquant l'improbabilité d'événements  $x$ .

$\Rightarrow p$  équivaut à [*non*  $p \Rightarrow DES$ ], et signifie typiquement que [*non*  $p$ ] est un événement désirable digne d'être remarqué, sans que  $p$  ait des conséquences spécialement néfastes. Nous éviterons de mêler les deux modalités *IMPR* and *IND* dans les mêmes formules, parce que cela se révèle non nécessaire.

Nous allons pouvoir vérifier que ce formalisme peut être utilisé avec succès pour représenter la signification logique des répliques et de leur contexte. Nous commencerons par montrer comment certaines des contraintes qui limitent la liberté des interlocuteurs lors de la première intervention peuvent être aisément modélisées à l'aide de la logique conversationnelle.

*Résumé* : il est impossible d'étudier un fragment de conversation si l'on ne possède pas la connaissance partagée (qui est le plus souvent implicite). Le contexte logique fait partie de cette connaissance partagée. Certains dysfonctionnements de la conversation peuvent se produire lorsque le contexte logique est perçu de manière incorrecte, alors même que la sémantique du message et la pragmatique de la situation sont correctement détectées. L'expérience du "Et alors ?" est une manière d'expliciter le contexte logique, qu'un formalisme simple utilisant les modalités *IMPR* et *IND* permet de représenter.

## 7. Introduire un nouveau sujet dans la conversation

Les interlocuteurs ne débent pas un nouveau sujet de conversation de manière quelconque. Du point de vue logique qui nous concerne ici, nous observerons qu'ils le font seulement de trois manières, ce que notre modèle peut représenter. La première manière consiste à mentionner un fait improbable.

### 7.1. Le mode de l'improbabilité

Nous allons modéliser dans un premier temps un type de conversation très fréquent : celles qui débent par la mention d'un fait inhabituel. Ceci inclut les faits actuels (si une tortue est en train de traverser le jardin, on ne peut résister à l'envie d'attirer l'attention sur elle), ou les histoires "incroyables" passées. Tous ces faits ont en commun la propriété d'être *a priori* improbables, et c'est précisément ce que prédit la théorie de Shannon si on l'appliquait aux conversations. Commençons par donner un exemple :

[ex\_voiture]

contexte : A parle à son fils C dont la soeur a commandé une nouvelle voiture au début de la semaine.

A1- Je t'ai dit que B a commandé une nouvelle voiture ?

Dans ce contexte où B achète une voiture par décennie, C peut assigner *a priori* une probabilité très faible au fait que sa soeur achète une voiture au cours de la semaine : nous pouvons estimer que cette probabilité est inférieure à  $10^{-2}$  puisqu'il y a une cinquantaine de semaines par an. Il est facile d'expliciter ce contexte de la manière suivante :

$commande(X, voiture, cette\_semaine) \Rightarrow IMPR$

Comment relier une telle situation au modèle de Shannon ? Les conversations spontanées sont souvent présentées comme un échange d'information. Mais qu'est-ce exactement que l'information ? Lorsqu'une personne que vous connaissez est sur le point de vous adresser la parole dans une situation propice à la conversation (pendant un repas par exemple), vous êtes en droit de supposer qu'elle va vous fournir une information : par exemple que X a acheté une nouvelle voiture. Ce concept d'information est largement utilisé pour décrire les processus du langage. Cependant nous préférons l'employer ici dans le sens étroit que lui a donné Claude Elwood Shannon [1948]. La quantité d'information contenue dans un symbole parvenant de la source du message est mesurée par  $\log(1/p)$ . Dans cette formule,  $p$  mesure la probabilité *a priori* du symbole. Il semble en effet normal qu'un événement très improbable (réception d'un symbole rare) procure une grande information lorsqu'il se produit.

Le fait d'utiliser cette définition de l'information va avoir des conséquences surprenantes dans notre contexte. On retrouve l'idée qu'un événement improbable est plus informatif. Mais l'analogie va plus loin. Dans le paradigme de Shannon (communications numériques), la *source* du message (pour nous : le premier locuteur) émet des *symboles* (pour nous ici, des mentions de faits ou d'événements) vers le *destinataire* (ici l'auditeur) qui va les **RE**connaître. Ceci veut dire que le récepteur doit connaître à l'avance tous les symboles qui peuvent lui parvenir ainsi que la probabilité *a priori* de chacun d'eux ! Comment ces caractéristiques plutôt restrictives peuvent être transposées dans un contexte d'énoncé informatif ?

Peut-on imaginer que les auditeurs "connaissent" à l'avance tous les événements que l'on pourra mentionner devant eux ? D'une certaine façon, oui. Ceci signifie que tous les événements susceptibles d'être mentionnés en début de conversation dans le mode improbable doivent non seulement être reconnus par les auditeurs comme possibles, mais ceux-ci doivent de plus être capables d'estimer la probabilité *a priori* de tels événements. Le premier locuteur espère de plus que cette estimation sera très faible.

Ceci correspond tout à fait à la définition de Shannon. Toutefois, ces estimations de probabilité ne peuvent être que qualitatives, et ne sauraient recevoir de valeur quantitative précise [Savage 1972:31]. Nous avons rencontré cette situation dans l'extrait précédent, mais aussi dans [ex\_antenne, p.21] A1-A2 (voir une antenne parabolique dans le voisinage est *a priori* improbable) et dans la question [ex\_Abraham, p.19] A1 (Abraham était improbableement âgé quand il devint père).

La transposition du modèle de Shannon est particulièrement éclairante et s'accorde parfaitement avec nos données. Plus précisément, il nous permettra d'exprimer des contraintes qui limitent ce qui peut être répliqué après une introduction sur le mode improbable.

Dans ces exemples dans lesquels la conversation débute par la mention d'un événement improbable, l'information transmise du premier locuteur vers ses auditeurs résulte de la combinaison de deux éléments : le message proprement dit (qui mentionne un événement explicitement) et le contexte logique (implicite le plus souvent) dont le rôle est de permettre une estimation de l'improbabilité du message. Nous allons constater de nouveau l'intérêt de ce contexte logique en observant d'autres modes d'introduction.

## 7.2. *Le mode paradoxal*

Toutes les conversations ne fonctionnent pas sur le mode improbable. Autrement dit, on ne saurait considérer que toute réplique initiale apporte une information, dans le sens de Shannon. Nous allons décrire maintenant un type de conversation que l'on rencontre quotidiennement : les conversations commençant par un étonnement, comme celles qui commencent par un "C'est bizarre..." ou un "C'est marrant..."

[ex\_jumelage]

contexte : A vient de traverser en voiture un village du Sud-Ouest de la France. A l'entrée, un panneau mentionnait un jumelage avec une ville d'Alsace. A s'attendait à ce que deux villes jumelées appartenissent à deux pays différents.

A1- Tout à l'heure, j'ai vu un village jumelé avec une ville d'Alsace !

Les deux endroits mentionnés en A1 sont situés dans deux régions de France. Ceci n'est pas un fait *a priori* improbable, car le contexte logique n'est pas "il m'arrive rarement de constater de tels jumelages". Si nous demandions à A "Pourquoi dis-tu ça ?", il répondrait quelque chose comme : "Je pensais que des villes jumelées devaient appartenir à des pays différents". Nous pouvons représenter le contexte par :

[jumelées( Ville\_A, Ville\_B ) & non étrangères( Ville\_A, Ville\_B )] ⇒ F

Si, en utilisant cette connaissance, nous voulions calculer une probabilité *a priori* pour que deux villes jumelées appartiennent au même pays, nous trouverions que cette probabilité est rigoureusement nulle ! Peut-on raisonnablement considérer que A1 apporte une information infinie ? La réponse est non, et selon

notre modèle nous devons admettre que l'objectif de A1 n'était pas d'apporter une information en mentionnant un fait d'une rareté infinie.

Le premier locuteur fait part d'un fait qui lui semble paradoxal, d'un fait qui lui semble en contradiction avec sa connaissance, mais pas d'une rareté<sup>(4)</sup>. Nous avons la même situation avec [ex\_Goffman, p.15] A1, [ex\_Mercedes, p.20] A1 et [ex\_animiste, p.18] A1. Dans chacun de ces cas, le contexte logique permet aux interlocuteurs de déduire par une preuve logique la négation du fait mentionné : B n'aurait pas dû *a priori* connaître les livres d'E. Goffman, parce qu'il n'est pas sociologue; les Hongrois n'auraient pas dû posséder de grosses voitures parce qu'ils ne sont pas riches; les gens ayant fait des études supérieures n'auraient pas dû être animistes. Voilà pourquoi nous parlons de contradictions, ou de paradoxes.

Nous verrons ci-dessous d'autres raisons de faire une différence qualitative entre improbabilité et paradoxe, et de ne pas considérer que celui-ci est une limite de celle-là. L'une de ces différences résulte de la forme des répliques admissibles dans ces deux modes. Mais avant cela, il nous faut d'abord observer qu'il existe un mode d'introduction qui diffère radicalement du paradoxe et de l'improbabilité.

### 7.3. *Le mode (in)désirable*

Nous allons décrire ici l'introduction, très fréquente, d'un sujet comportant un fait (ou une éventualité) présenté comme indésirable.

[ex\_faim]

contexte : des étudiants conversent

A1- *Bon sang j'ai rien bouffé ce soir.*

B1- [mangeant un hamburger] *Tu veux un peu de mon hamburger ?*

A2- *Non, non, merci, pas de hamburger.*

C1- *Vous voulez des gâteaux ? J'ai des cookies.*

A3- *Et tu attends 9H pour les proposer ? !*

Nous aurions du mal à trouver ici quoi que ce soit de paradoxal ou d'improbable. A1 est clairement la manifestation d'un fait indésirable, ce que la représentation suivante du contexte laisse apparaître :

*dîner\_sauté*  $\Rightarrow$  *faim*

*faim*  $\Rightarrow$  **IND**

Nous avons rencontré un cas analogue avec [ex\_carte-postale, p.21] A1. Voici maintenant un exemple d'introduction sur le mode désirable. A parle d'une station de ski (l'une des plus vastes dans le monde) :

[ex\_ski]

contexte : Le groupe de skieurs s'arrête pour déjeuner. Ils sont dans une station dont on dit qu'elle est l'une des plus grandes du monde.

A1- *Ce qu'il y a de bien ici, c'est que c'est très étendu. C'est pas étouffant, la vallée.*

B1- *oui*

A2- *Partout où je suis allée,...*

Nous pouvons représenter le contexte par :

*étendu*  $\Rightarrow$  **non étouffant**

**non étouffant**  $\Rightarrow$  **DES**

<sup>4</sup> Nous verrons que des faits présentés comme paradoxaux par les locuteurs peuvent même être fréquents.

Ici nous avons choisi (non étouffant  $\Rightarrow$  **DES**) au lieu de (**DES**  $\Rightarrow$  non étouffant). En effet, A2 (dont l'enregistrement est incomplet) semble indiquer que A tolère habituellement des stations de ski "étouffantes", en sorte que (non étouffant) apparaît être une qualité suffisante pour changer une situation normale en situation désirable.

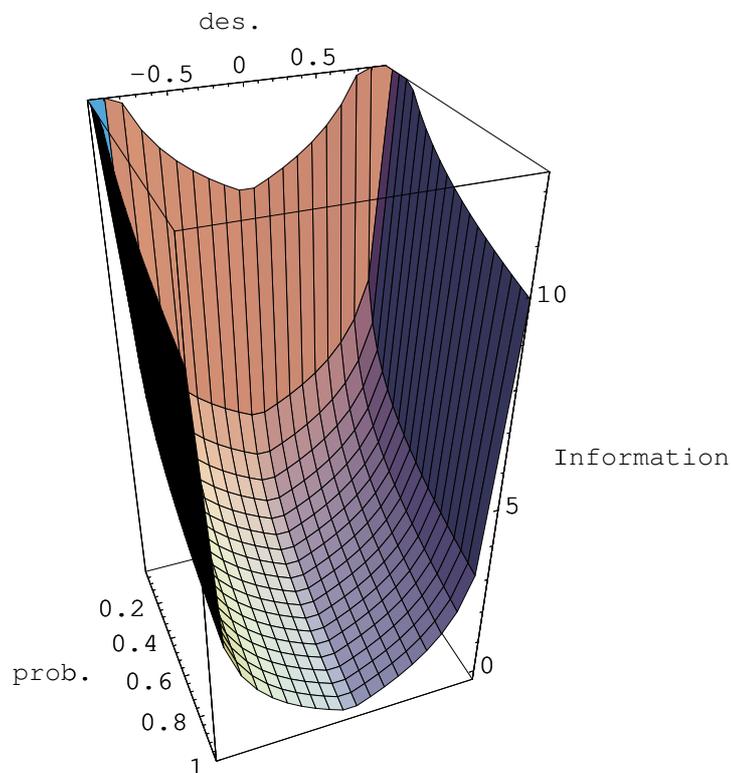
Ces introductions sur le mode (in)désirable ressemblent quelque peu au cas de l'improbabilité, dans la mesure où l'on peut supposer que les interlocuteurs sont capables d'assigner *a priori* une valeur positive ou négative de désirabilité à chaque événement imaginable. On peut même voir dans ce paramètre une nouvelle dimension de l'information. Par exemple, "X a été élu" n'apporte qu'un bit d'information au sens de Shannon si les deux candidats avaient des chances identiques *a priori*, mais cet événement peut avoir pour vous une grande valeur de désirabilité ou d'indésirabilité selon que votre favori est élu ou non, si vous êtes très concerné(e) par l'élection.

On peut ainsi imaginer étendre la définition de Shannon à une information à deux paramètres qui à chaque événement susceptible d'être mentionné associe un couple de nombres  $(p,d)$  où  $p$  mesure la probabilité *a priori* de l'événement, et  $d$  sa désirabilité *a priori* (entre -1 et 1). On peut définir une fonction d'*Information Conversationnelle* de cette façon :

Information conversationnelle :

$$I_c(p,d) = \log_2 \text{ Erreur !}$$

où  $p$  et  $d$  sont caractéristiques d'un événement donné. Cette formule est une extension simple de la formule de Shannon pour un événement isolé. L'Information Conversationnelle apparaît comme une fonction croissante de la désirabilité ( $d > 0$ ) ou de l'indésirabilité ( $-d > 0$ ), en atteignant des valeurs infinies aux valeurs extrêmes de  $d$  ( $|d| = 1$ ), ce qui lui donne une forme de vallée en V.



----- figure 2 : information conversationnelle -----

Noter que cette information permet de représenter les modes improbable et (in)désirable. En revanche, il n'est pas possible de définir l'information apportée par un fait logiquement impossible (un tel fait n'est pas un fait de probabilité nulle, cf. p. 42).

#### 7.4. Classification des conversations

Nous avons illustré trois modes différents d'introduction d'un nouveau sujet dans une conversation. Comme nous allons pouvoir le vérifier, l'enchaînement logique des répliques ultérieures dépend de manière cruciale du mode d'introduction, et il est donc essentiel d'en connaître la gamme complète. La principale prédiction du modèle exposé ici, pour les types de conversation considérés, est résumée dans le premier principe de pertinence logique, qui indique que les trois modes évoqués ci-dessus : improbabilité, paradoxe, (in)désirabilité, forment une classification exhaustive.

#### Premier principe de pertinence logique :

*Il n'y a pas de conversation sans problématique  
(improbabilité, paradoxe, (in)désirabilité).*

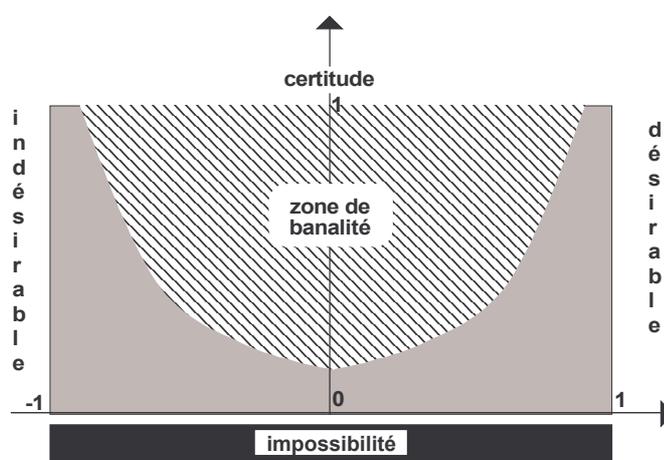
Une telle limitation a de quoi surprendre. Que penser des conversations qui débutent par une question ? Et ne peut-on pas imaginer d'autres sortes d'introduction ? Ne pourrait-on pas, par exemple, voir une conversation commencer par la simple assertion d'un fait vrai, ou par l'expression d'une opinion, sans la moindre mention d'événements improbables, paradoxaux ou (in)désirables ? Ceci est peut-être possible pour d'autres types de conversations (dans des situations expérimentales ou de résolution de problème). Toutefois, le modèle proposé ici avec ses trois modes tels qu'ils ont été décrits rend très bien compte des données que j'ai recueillies, et son objectif est de s'appliquer avec précision à la plupart des situations spontanées (telles que nous les avons définies p. 13).

Cette classification des propriétés logiques des introductions dans les conversations spontanées en trois classes : improbabilité, paradoxe, (in)désirabilité, est un modèle prédictif, et donc réfutable. Ce modèle prédit que des contraintes étroites affectent le comportement du premier locuteur : dans le type de conversations spontanées que nous décrivions au début de cette présentation, tout nouveau sujet doit être placé dans un contexte problématique<sup>5</sup>. Mais le comportement de l'interlocuteur est lui aussi contraint. Si le mode d'introduction ne lui apparaît pas comme appartenant à l'un des trois modes, alors le modèle prédit que l'on observera une *cassure* comme "Et alors ?" ou quelque chose d'analogue. Nous avons rencontré cette situation dans [ex\_Mercedes, p.20] B1, et dans la première partie de la réaction intentionnellement provoquée [ex\_train2, p.20] B1. Dans la situation naturelle de [ex\_Mercedes], A1 avait été formulée sur le mode paradoxal, mais B ne l'avait pas perçue car un élément du contexte logique lui faisait défaut.

On pourrait argumenter sur le fait que cette classification des introductions en trois types n'est pas constituée de classes disjointes. Considérons l'énoncé imaginaire : "J'ai croisé une voiture qui roulait à contresens sur l'autoroute". Il peut être perçue comme un paradoxe si l'interlocuteur a des raisons de penser que cet événement est impossible dans des conditions normales ("Et le conducteur ne s'apercevait de rien ?"), ou bien comme un fait improbable ("ça m'est arrivé une fois aussi"), ou comme un événement hautement indésirable ("comment as-tu réagi ?"). Cette critique est parfaitement justifiée au point où nous en sommes. Mais nous pouvons donner deux réponses. Premièrement, dans beaucoup de situations réelles, le mode apparaît sans ambiguïté aux interlocuteurs (voir [ex\_Côme, p.41]). Deuxièmement, même quand l'introduction peut être perçue simultanément comme appartenant à plusieurs modes, les réactions, elles, ne pourront agir que sur un seul mode à la fois, comme nous le verrons dans une section ultérieure.

<sup>5</sup> Ce terme ne prend pas en compte le mode désirable, mais il est commode pour résumer la forme que doit revêtir un nouveau sujet de conversation.

Il semble commode de placer chaque nouveau sujet dans un espace à deux dimensions. L'analyse des répliques ultérieures va montrer en effet que la probabilité et la désirabilité peuvent varier indépendamment, et peuvent donc être considérées comme deux degrés de liberté. Un nouveau sujet, pour être acceptable, doit être présenté comme ayant une valeur extrême pour l'un au moins de ces deux paramètres. La première intervention apparaîtra comme normale si et seulement si elle peut être située par les auditeurs en dehors de la "zone de banalité" de la figure 3. La frontière de cette zone a une forme hyperbolique qui correspond à une information conversationnelle constante (cf. p. 26). Le premier énoncé se situera dans la zone d'impossibilité si l'événement est paradoxal, juste au-dessus s'il est improbable, loin à droite s'il est désirable ou à gauche s'il est indésirable. Cette représentation bidimensionnelle permet à un nouveau sujet d'être par exemple indésirable et improbable simultanément, en situant son point représentatif dans le coin inférieur gauche du diagramme. Cette représentation va se révéler très utile lorsque nous étudierons l'effet des répliques sur le contexte logique.



----- figure 3 : espace de pertinence -----

### 7.5. Les exceptions apparentes

Avant d'affirmer qu'il y a des exceptions évidentes qui échappent aux prédictions du modèle, le lecteur est invité à les imaginer en contexte. En effet, il semble difficile d'imaginer quelqu'un introduire un nouveau sujet en affirmant un fait qui n'apparaîtrait pas clairement comme paradoxal, improbable, ou (in)désirable par rapport au contexte. Une telle assertion serait perçue comme banale et provoquerait des réactions de réparation.

Mais on pense immédiatement à des introductions qui ne consistent pas en une assertion, mais en une question. Notons tout d'abord que les nouveaux sujets sont assez souvent introduits d'une façon assertive, contrairement à ce que certains modèles linguistiques supposent<sup>(6)</sup>. Mais les questions sont toutefois fréquentes et doivent être intégrées à notre modèle. En effet, en posant une question, le premier locuteur ne dit pas nécessairement quelque chose d'(in)désirable, d'improbable ou d'impossible. Mais il y a des contraintes aussi sur les questions, comme le montre l'extrait [ex\_repas, p.14] : toute question n'est pas admissible dans une situation donnée, même une question aussi anodine que "Et toi, ça va ? Tu as bien mangé à midi ?".

En fait, les contraintes repérées par notre modèle vont porter non sur la question elle-même, mais sur les types de réponse anticipés par le questionneur. Voyons cela sur des exemples.

<sup>6</sup> Par exemple, dans le modèle IRF qui est souvent mentionné ([Coulthard 1977:135]; [Stubbs 1983:136]; [McTear 1985:35]), le rôle de l'initialisation (*initiation*) est presque toujours joué par une question.

[ex\_Channel] [Crystal & Davy 1975:52]

contexte : deux couples (B et C, A et D) viennent d'échanger leur expériences de vacances. La discussion arrive sur le meilleur moyen de traverser la Manche.

C3- *I see. How did you get - I mean how did you find that side of it, because...*

A4- *Marvellous*

C4- *you know some people say that... that driving a car across a ferry is the devil of a job*

A5- *well this was...*

D3- *across a...*

C5- *I mean taking a car across to the continent on a ferry is... is hell*

A6- *no it isn't at all*

D4- *why*

C6- *I don't know but...*

[. . .]

C9- *well I'll tell you the sort of thing I've heard, I mean ev... every summer, you see stories of tremendous queues at the...*

D8- *but they're people who haven't booked*

Ici, le contexte logique n'est pas pleinement explicité par C avant C9 (après les essais infructueux C4 et C5) qui montre que C teste un fait indésirable. En posant la question C3, C s'attendait à des commentaires relatifs aux queues, parce que les queues sont hautement indésirables ("it's hell").

Nous avons pu observer des questions dont la fonction était de tester l'occurrence d'un paradoxe avec [ex\_Goffman, p.15] A1, C1, A2, A4, et avec [ex\_Mercedes, p.20] A1. Les questions peuvent aussi contribuer à confirmer un événement improbable, comme dans [ex\_Abraham, p.19] A1 (si Abraham était réellement vieux lorsqu'il a procréé, alors l'événement est improbable) ou dans [ex\_voiture, p.23] A1 (si j'ai oublié de te dire que ta soeur a acheté une voiture, alors cela reste un événement improbable pour toi).

Bien sûr, nous pouvons poser des questions pour lesquelles nous n'avons pas d'idée de la réponse : "Quelle est la capitale du Burundi ?". Mais il faut alors imaginer une telle question en situation, comme un moyen d'introduire un nouveau sujet. Que répondrait-on après une réaction du type "Pourquoi veux-tu savoir ça ?" ? D'après le modèle, le premier locuteur révélerait alors une situation problématique<sup>(7)</sup>. Celle-ci peut aller du manque indésirable d'instanciation (par ex. si le locuteur joue aux mots-croisés) à un problème plus structuré (par ex. une incohérence apparente dans un passage de roman qui se passe en Afrique). Mais même le manque d'instanciation se doit d'être spécifiquement problématique. En d'autres termes, le modèle exclut les questions non motivées.

Notre modèle prédit que chaque fois qu'un locuteur introduit un nouveau sujet en posant une question au cours d'une conversation spontanée, il a une idée précise d'au moins une réponse possible ou de certaines de ses caractéristiques<sup>(8)</sup>, et que **cette réponse attendue est logiquement problématique** (soit paradoxale, soit improbable, soit (in)désirable). De plus, les interlocuteurs doivent la plupart du temps être conscients de cette anticipation (cf. [Fox 1987:375]). Quand il ne la perçoivent pas, comme dans [ex\_repas, p.14] A1, ils émettent une protestation (souvenez-vous que les interlocuteurs sont supposés bien se connaître). Dans [ex\_repas], A pensait que B avait eu un repas amélioré sur son lieu de travail, comme c'est la coutume juste avant Noël. A testait donc l'occurrence d'un fait désirable.

De ce fait, les questions ne sauraient constituer en tant que telles une exception à la règle d'introduction, et le modèle parvient à les prendre en compte quand le contexte logique est connu. Mais nous pourrions

<sup>7</sup> ou désirable.

<sup>8</sup> Dans l'exemple des mots-croisés, ces caractéristiques peuvent être certaines lettres connues, mais c'est un exemple extrême d'introduction de nouveau sujet.

trouver d'autres cas dans lesquels la première intervention semble en désaccord avec le modèle, comme cette réplique entendue après qu'une interlocutrice eut quitté la pièce en entendant son enfant pleurer :

[ex\_calmer]

contexte : un enfant pleure. Sa mère se lève et monte à l'étage pour le consoler. Quelqu'un commente :

A1- *Je ne crois pas que le fait d'y aller, ça aide à les calmer.*

Cette constatation n'a rien à voir avec un événement improbable ou impossible. "Y aller" n'est pas présenté comme ayant des conséquences désastreuses. Est-ce là un cas non prévu par le modèle ? On sent bien tout de même qu'il y a quelque chose d'indésirable dans cette situation (comme le reste de la conversation le montre) : la possibilité que l'enfant continue à pleurer. Chacun autour de la table a entendu un enfant pleurer, et aurait pu déclarer : "Ecoutez, il pleure de nouveau !", indiquant par là un événement indésirable. Cette intervention implicite précède logiquement A1, et notre modèle s'applique à elle, et non à A1. De telles situations où les premiers énoncés ne sont pas prononcés semblent être relativement rares (il n'y a que peu d'exemples dans le corpus, et tous coïncident avec des événements externes évidents). Les locuteurs, sauf lorsqu'ils cherchent à faire de l'humour (mais cela ne nous concerne pas ici), semblent faire des efforts particuliers pour être explicites, même au prix d'être redondants.

## 7.6. Une définition du sujet

Ce modèle de la première intervention peut être formulé de la façon suivante : un sujet sera correctement introduit si et seulement si le contexte logique (qui fait partie de la connaissance pragmatiquement pertinente associée au sujet) comprend un lien problématique que nous pouvons exprimer logiquement<sup>9</sup> par :

$$[ p_1 \ \& \ p_2 \ \& \ \dots \ \& \ p_n ] \Rightarrow \text{MOD}$$

où **MOD** représente l'une des modalités **F**, **IMPR**, **DES** et **IND**, et où toutes les  $p_i$  sont tenues pour vraies. Nous désignerons cette relation par le terme de *clause saturée*, ou *clause de référence*.

Ce modèle de la première intervention nous permet de donner une définition précise de ce qu'est le sujet courant d'une conversation. Le sujet est généralement considéré comme une réponse possible à la question "de quoi étiez vous en train de parler?". Mais une telle définition est trop vague, et peut aller d'un simple mot jusqu'au fait de rejouer la conversation.

Dans notre modèle, la clause saturée est essentielle pour caractériser une conversation donnée, et nous proposons de la prendre comme définition du sujet. Dès que cette clause saturée change au cours de la conversation, on peut dire que le sujet a changé.

Il nous faut maintenant étudier ce que peut être l'effet logique des répliques ultérieures sur le sujet, afin de mieux cerner ce que "pertinence logique" veut dire.

*Résumé* : selon notre modèle, un nouveau sujet doit apparaître de manière évidente soit comme paradoxal, soit comme hautement improbable, soit comme hautement désirable ou indésirable. Il doit apporter une grande quantité d'*information conversationnelle*. Sinon, le modèle prédit que l'introduction sera un échec et provoquera une cassure.

## 8. La deuxième intervention

Nous avons montré que la première intervention était fortement contrainte. Mais lorsque ces contraintes ont été perçues comme étant respectées, y a-t-il encore des limites à ce que les interlocuteurs peuvent répondre ? C'est le sujet de cette section.

<sup>9</sup> Nous n'écrivons ici que des propositions, par souci de simplicité.  $p_i$  peut comporter une négation. Voir annexe.

### 8.1. Réactions à un fait paradoxal

Considérons l'extrait suivant dans lequel A s'étonne d'un fait paradoxal

[ex\_poulet]

contexte : Beaucoup de gens préfèrent acheter des poulets "garantis fermiers" qui sont pourtant beaucoup plus chers. A s'étonne de ce comportement, puisque selon lui "garanti fermier" ne sont que des mots, et non une véritable garantie.

A1- Alors ce qui impressionne les populations, c'est le mot garantie . Un gars qui fait des poulets d'élevage, il met garanti fermier, alors tout le monde se précipite.

B1- Mais il y a des labels

A2- Oui, mais garantis par qui ?

B2- Mais ils le disent, par qui, parfois... par la chambre syndicale des machins...

Pour A, il est illogique d'acheter en se fiant à une mention qui n'est pas garantie. Pour lui, le contexte contient la connaissance que nous représentons par la clause suivante :

[ achat\_dû\_à(Mention) & non garanti(Mention) ]  $\Rightarrow$  F

Le contexte, au niveau de A1, comprend cette clause saturée ainsi que les éléments suivants, que nous notons aussi sous forme de clauses négatives<sup>(10)</sup> :

garanti(mention\_poulet\_fermier)  $\Rightarrow$  F

non achat\_dû\_à(mention\_poulet\_fermier)  $\Rightarrow$  F

Le paradoxe peut être représenté par la première clause dans laquelle tous les termes deviennent vrais quand les deux autres clauses sont prises en compte, [Mention] étant instanciée en [mention\_poulet\_fermier]. Quelle est l'action logique effectuée par B1 ? Avec B1, B essaie d'établir [garanti(mention\_poulet\_fermier)] en adjoignant une connaissance au contexte logique, ce que nous rendons par l'ajout de la clause suivante :

[ label(L) & accompagne(L,Mention) & garanti(L) & non garanti(Mention) ]  $\Rightarrow$  F

avec la supposition que quelque label garanti L existe. En effet, cette clause se réécrit :

[ label(L) & accompagne(L,Mention) & garanti(L) ]  $\Rightarrow$  garanti(Mention)

L'effet de B1 est donc d'invalider le paradoxe de A en niant directement l'un de ses termes : [non garanti(mention\_poulet\_fermier)]. Nous appellerons ce type de réaction *invalidation directe*.

Dans la mesure où la clause saturée comportait deux termes, notre traduction logique prédit une autre possibilité d'invalidation directe. B aurait pu dire quelque chose comme : "ils n'achètent pas à cause de la garantie; ces poulets-là ont tout de même meilleur aspect", niant par là-même le terme [achat\_dû\_à(mention\_poulet\_fermier)]. L'invalidation directe, qui consiste à nier l'un des termes de la clause saturée, ou au moins à le mettre en doute, offre un nombre plutôt limité de possibilités et peut donc être aisément anticipée par le modèle.

Noter que le contexte logique est modifié au cours de la conversation. Ici, B active une connaissance qui doit donc être intégrée au contexte logique. Nous allons observer maintenant des cas dans lesquels la clause saturée elle-même est modifiée. En effet, beaucoup de réactions à des introductions sur le mode paradoxal ne

<sup>10</sup> Toute base de connaissances exprimée en logique des propositions peut être présentée sous la forme d'un ensemble (conjonctif) de clauses négatives du type [ p<sub>1</sub> & p<sub>2</sub> & . . . & p<sub>n</sub> ]  $\Rightarrow$  F. Il s'agit d'une simple réécriture de la forme normale conjonctive-disjonctive. Nous étendons cette représentation à la logique conversationnelle grâce à l'emploi des modalités IMPR, IND et DES. Voir annexe.

se traduisent pas par des invalidations directes. Nous en avons vu un exemple avec [ex\_Goffman, p.15]. Au départ, le paradoxe résidait dans la clause saturée :

$$[ \textit{connaît}(\mathbf{X}, \textit{livres\_de\_Goffman}) \ \& \ \textit{non sociologue}(\mathbf{X}) ] \Rightarrow \mathbf{F}$$

L'intervention B5 ("I heard about it from a friend who was a sociologist, and he said read this book, ..."), n'a aucun effet sur les deux termes de cette clause saturée, et donc B5 n'est pas ce que nous avons appelé une invalidation directe. Toutefois on sent bien que le paradoxe est tout de même invalidé après B5. Ceci ne se comprend que si l'on reconnaît que la clause précédente est maintenant incorrecte et doit être remplacée par :

$$[ \textit{connaît}(\mathbf{X}, \textit{livres\_de\_Goffman}) \ \& \ \textit{non sociologue}(\mathbf{X}) \\ \& \ \textit{non recommande}(\mathbf{Y}, \mathbf{X}, \textit{livres\_de\_Goffman}) ] \Rightarrow \mathbf{F}$$

Autrement dit, en niant le dernier terme ci-dessus, B5 n'invalide pas la clause saturée initiale, mais une version modifiée, complétée.

Voici pourquoi nous parlons ici d'*invalidation indirecte*. Contrairement à l'invalidation directe, l'invalidation indirecte consiste à nier un terme "oublié" dans la clause saturée. En affirmant [*recommande*(ami, B, livres\_de\_Goffman)], B5 nie un terme qui apparaît comme une prémisses oubliée dans le contexte initial. La possibilité d'invalidation indirecte pour un contexte donné est fondamentale, et la distinction entre les invalidations directe et indirecte est syntaxiquement évidente dès que le contexte est exprimé sous forme logique.

La possibilité d'une invalidation indirecte ne devrait pas donner l'impression que cette façon de répliquer est peu contrainte. En effet toutes les modifications de la clause saturée ne sont pas admissibles ! Le paradoxe modifié doit être accepté par le premier locuteur. Imaginons une seconde la réaction de A si B avait rétorqué "c'est parce que j'ai faim" ! Plus précisément, si la clause saturée s'écrit (en logique propositionnelle) :

$$[ p_1 \ \& \ p_2 \ \& \ \dots \ \& \ p_n ] \Rightarrow \mathbf{F}$$

une invalidation directe est une négation de l'un des  $p_i$ , alors qu'une invalidation indirecte fait intervenir la négation d'une prémisses additionnelle  $p_{n+1}$  :

$$[ p_1 \ \& \ p_2 \ \& \ \dots \ \& \ p_n \ \& \ p_{n+1} ] \Rightarrow \mathbf{F}$$

Une invalidation indirecte n'est admissible que si le locuteur surpris peut l'accepter comme niant une prémisses oubliée  $p_{n+1}$ . En d'autres termes, ce locuteur doit accepter que ce que nous représentons à l'aide de la clause ci-dessus représente vraiment l'incompatibilité qu'il a soulevée.

Il n'est pas étonnant que des prémisses comme  $p_{n+1}$  puissent être ainsi "oubliées" par le premier locuteur. Après tout, toute incompatibilité relevée dans la vie courante présuppose que le monde existe encore, que les personnes se trouvent à un seul endroit à un moment donné, et ainsi de suite. Mais exiger qu'un fait donné  $p_{n+1}$  puisse être reconnu comme partie intégrante de l'incompatibilité initiale reste une contrainte très forte sur ce qui peut ou ne peut pas être considéré comme une invalidation admissible. Dans l'extrait précédent, A aurait accepté (sans nécessairement y souscrire) que B s'attaque à des hypothèses comme :

- on ne lit pas un livre par hasard
- on ne lit que les livres de son domaine
- la lecture d'un livre ne résulte pas d'un pari
- etc.

parce que ces termes peuvent être reconnus par elle comme faisant partie de la clause saturée. Mais elle n'aurait sans doute pas admis des arguments qui auraient nié :

- Goffman est blond
- les livres de Goffman ont un nombre pair de pages
- le climat est doux en Oregon

parce qu'elle aurait refusé d'intégrer ces termes dans l'incompatibilité, et que dans ce cas leur valeur de vérité est sans effet sur le paradoxe.

Le caractère admissible d'une invalidation dépend totalement du contexte logique et de la façon dont ce contexte pourrait être augmenté pour intégrer des prémisses oubliées. La procédure qui nous permet de dire si une telle prémisse est admissible est exactement la même que celle par laquelle nous avons obtenu le contexte logique. Ce rôle est normalement rempli par la Pragmatique. Mais en l'absence d'une méthode fiable, ce travail peut être réalisé par tout observateur connaissant parfaitement la situation. Nous verrons que dans certaines circonstances, un système artificiel peut reconnaître des invalidations indirectes lors de conversations avec des utilisateurs humains.

La gamme complète des répliques prédites par le modèle après une introduction sur le mode paradoxal est ainsi plutôt limitée. Toute réplique doit être reconnue comme une invalidation admissible de la clause saturée (à moins que cette invalidation ne soit retardée par une réaction de co-étonnement, comme nous le verrons p. 43). Les réactions de A et de C, dans [ex\_Goffman, p.15], illustrent parfaitement ce point : B1, B2, B3 et B4 ne pouvaient pas satisfaire A et C, puisqu'elles n'avaient aucun effet sur la clause saturée. Le fait que B lût le livre n'invalidait aucun terme de la clause saturée. Au contraire ! Ils n'obtinrent une invalidation (indirecte) qu'en B5.

Toute invalidation d'un contexte paradoxal peut toujours être décrite comme une *explication*. A la suite d'une intervention sur le mode paradoxal (ce que nous décrivons dans [Dessalles 1992e] comme une surprise logique (model-based surprise)), une explication est due. Inversement, toute réplique spontanée qui ressemble à l'explication d'une surprise émise ou anticipée peut être considérée comme l'indice d'un paradoxe sous-jacent.

Dans cette section, nous avons montré comment le modèle contraint la seconde réplique dans le cas d'une introduction sur le mode paradoxal. Examinons maintenant quelles répliques peuvent être acceptables après une introduction sur le mode improbable.

## 8.2. Réactions à un fait improbable

### 8.2.1. Invalidation d'un fait improbable

L'extrait suivant est sur le mode improbable :

[ex\_dard]

contexte : A appelle B pour qu'il voie un insecte muni d'un étrange appendice

A1- Tiens, regarde le dard qu'elle a

B1- ça alors !

A2- C'est marrant

B2- C'est pas un dard, c'est une trompe

A apporte une information, car les dards d'insectes sont *a priori* supposés être beaucoup plus courts. Celui-là était improbableement long. Nous pouvons écrire ceci de la manière suivante :

[ *dard*(S) & *longueur*( S, L) & L > 1cm ]  $\Rightarrow$  IMPR

Ce contexte devient caduc à la suite de l'invalidation B2. Nous voyons qu'une invalidation, ici une invalidation directe, peut aussi se produire dans une conversation sur le mode improbable. On peut donc se

demander s'il est légitime de maintenir une distinction nette entre le mode improbable et le mode paradoxal, ou si l'on doit considérer que le second est un cas limite du premier. Il se trouve toutefois qu'il est absolument nécessaire de différencier ces deux modes, car comme nous allons le voir il existe une façon de répliquer dans le mode improbable qui n'est pas admissible après une introduction sur le mode paradoxal.

### 8.2.2. Banalisation d'un fait improbable

[ex\_soif]

contexte : A et B sont en train de parler de D, leur petit fils. D (enfant de un an) semble se souvenir d'eux après une séparation de plusieurs mois. B prétend ne pas être surprise : l'enfant riait en entendant leur voix au téléphone. Quelques secondes plus tard, C remarque que D ingurgite une grande quantité d'eau.

A1- Apparemment, il nous avait pas oubliés.

B1- Non. Il nous a pas oubliés quand ... il riait aux éclats quand il entendait notre voix.

[pause]

C1- Hé ben, il avait soif !

B2- Oui, il avait soif. Je m'en suis doutée, qu'il avait soif !

Le premier sujet abordé dans cet extrait porte sur un enfant (D) que A et B retrouvent après une longue séparation. D est si jeune qu'il était *a priori* improbable qu'il se souvînt d'eux. A1 apporte une information au sens de Shannon. B1 n'invalide en aucune façon le raisonnement menant à la conclusion **IMPR**. Si nous pouvions demander à B pourquoi elle a émis la réplique B1, elle "répondrait" certainement :

$riait\_en\_entendant(D, A\_et\_B) \Rightarrow \mathbf{non\ oubli}(D, A\_et\_B)$

Le contexte logique était initialement :

$[jeune(D) \& séparation\_longue(D,A\_et\_B) \& \mathbf{non\ oubli}(D,A\_et\_B)] \Rightarrow \mathbf{IMPR}$

Donc B1 ne réalise pas une invalidation. Au contraire, elle semble confirmer A1 ! Cet extrait est assez remarquable, car B fait preuve d'exactly le même comportement une seconde plus tard, à propos du deuxième sujet. C est impressionnée par la quantité d'eau que D est en train d'avalier, et apporte ainsi une information en attirant l'attention sur un fait improbable. B2 ressemble une fois encore à une confirmation.

De nombreux auteurs considèrent que la conversation est une sorte de coopération au cours de laquelle l'apport d'information et la confirmation jouent un rôle important. Pourtant, au niveau logique auquel nous examinons cet extrait, on peut voir que les répliques de B, dans les deux cas, ne sont pas de simples acquits qui signalent la bonne réception des énoncés de A et de C. Il est très important de comprendre que B1 et B2 visent à *diminuer l'originalité* de A1 et C1 respectivement. Vues ainsi, B1 et B2 perdent totalement leur aspect coopératif !

Ce que A et C affirment, ce n'est pas tant l'événement qu'ils ont remarqué. C'est son improbabilité *a priori*. B, par deux fois, modifie cette probabilité. Après B1, par exemple,  $\text{Prob}(\mathbf{non\ oubli})$  doit être remplacé par  $\text{Prob}(\mathbf{non\ oubli} \mid riait\_au\_téléphone)$ , qui est beaucoup plus élevée (elle vaut en fait 1 si on considère que  $riait \Rightarrow \mathbf{non\ oubli}$ ). En d'autres termes, avec la connaissance que D riait au téléphone en entendant les voix de A et de B, il est beaucoup moins improbable qu'il se souvienne d'eux. Avec B2 de nouveau, B indique simplement que la probabilité que D soit assoiffé n'était pas si basse puisqu'elle suspectait qu'il l'était.

Nous appellerons ce genre de réaction une réaction de *banalisation*. En émettant une invalidation, le deuxième locuteur détruit le raisonnement menant au constat d'improbabilité, et celle-ci disparaît. Au contraire, une réaction de banalisation diminue simplement la valeur de cette improbabilité, sans l'éliminer.

Nous définirons l'effet de la banalisation comme la différence entre l'information apportée par l'événement  $Ev$  mentionné avant et après la réaction de banalisation :

$$\text{Effet de la banalisation : } \log_2 \frac{1}{\Pr_1(Ev)} - \log_2 \frac{1}{\Pr_2(Ev)} = \log_2 \frac{\Pr_2(Ev)}{\Pr_1(Ev)}$$

Dans le cas où, comme dans l'exemple B1, la banalisation est obtenue par mention d'un événement  $Ev_1$ , on a :

$$\text{Banalisation par complément d'information : } \log_2 \frac{\Pr(Ev|Ev_1)}{\Pr(Ev)}$$

### 8.2.3. Banalisation d'un événement rare

Ce modèle permet de prédire plusieurs moyens pour un locuteur Y d'augmenter la probabilité *a priori*  $P(Ev_0)$  d'un événement donné  $Ev_0$  présenté par X comme improbable. Y peut simplement indiquer que sa propre estimation est supérieure à celle de X, comme ce fut le cas dans B2 :

$$\Pr_Y(Ev_0) \gg \Pr_X(Ev_0)$$

Y peut aussi révéler qu'il connaît des faits additionnels  $F_1, \dots, F_n$  tels que :

$$\Pr(Ev_0 | F_1, \dots, F_n) \gg \Pr(Ev_0)$$

C'est ce que B fit en B1. Maintenant, Y peut aussi adopter une autre stratégie pour diminuer l'information apportée par X, lorsque  $Ev_0$  est un événement *rare* :  $Ev_0$  est susceptible de se produire plusieurs fois, quoiqu'extrêmement rarement (penser à l'achat de la voiture dans [ex\_voiture, p.23]). Si nous écrivons  $Ev_0 = Ev(\text{Sit}_0)$  pour représenter le fait que  $Ev_0$  est la réalisation d'un événement générique  $Ev()$  dans la situation présente (ou rapportée)  $\text{Sit}_0$ , on peut modéliser l'estimation de la probabilité *a priori* d'un tel événement par la mesure statistique :

$$\Pr(Ev(\text{Sit}_0)) = \sum_{\text{Sit} \in \Omega} \frac{\delta(Ev(\text{Sit}))}{\text{card}(\Omega)}$$

$Ev(\text{Sit})$  est vrai si l'événement  $Ev$  s'est produit dans la situation  $\text{Sit}$ , et  $\delta(Ev(\text{Sit})) = 1$  ssi  $Ev(\text{Sit})$  est vrai.  $\Omega$  est l' "univers" des situations prises en compte. Il contient toutes les situations qui sont analogues à celle l'événement rapporté, c'est-à-dire les situations qui sont obtenues en considérant que certains paramètres dans  $\text{Sit}_0$  peuvent prendre d'autres valeurs. L'univers  $\Omega$  est construit par les interlocuteurs à partir de l'énoncé du fait improbable. Ce phénomène important de construction d'un univers pertinent pour n'importe quel événement devrait faire l'objet d'études détaillées. Nous nous contenterons de quelques indications qui devraient suffire à notre propos.

L'univers  $\Omega$  est construit à partir de  $\text{Sit}_0$  et de  $Ev_0$ . La situation  $\text{Sit}_0$  est généralisée par abandon de caractéristiques non pertinentes. Par exemple, l'antenne parabolique de [ex\_antenne, p.21] a été vue dans une rue située à 300m du domicile des interlocuteurs. L'univers va comprendre toutes les antennes, paraboliques ou non, qui sont situées dans un rayon de 300m. La position de la maison qui portait l'antenne, le nom de la rue, etc. étaient pragmatiquement non pertinents :

**Sit<sub>0</sub>** : antenne TV blanche vue sur le toit d'une grande maison au n°124 de la rue des Roissys par beau temps ...

**Sit** : antenne TV vue dans un rayon de 300m autour du domicile

Une caractéristique très importante de  $\Omega$  est qu'il est *égocentrique*. Les particularités de  $\text{Sit}_0$  sont généralisées selon une hiérarchie conceptuelle centrée sur le locuteur (puis sur l'auditeur). Dans l'exemple précédent, la rue dans laquelle l'antenne parabolique a été aperçue n'est pas généralisée en une rue quelconque, mais en toute rue située dans le voisinage du domicile des interlocuteurs. On peut dire que  $\text{Sit}_0$  et **Sit** sont "égocentriquement analogues" et parler de *sphère égocentrique* à propos de  $\Omega$ , puisque les caractéristiques non pertinentes de la situation sont éliminées sous la contrainte suivante : la "distance conceptuelle" au locuteur ne doit pas augmenter. Cette distance peut être concrète, comme pour l'antenne, mais elle est le plus souvent abstraite. Elle est évaluée dans l'espace, ou dans le temps, ou selon le degré de

parenté, selon la familiarité, etc.. Toute localisation dans l'espace ou le temps, ou toute caractéristique impliquant un point de vue définit une telle distance. Par exemple le changement de *vue* en *entendu parler de* dans la généralisation précédente  $Sit_0 \rightarrow Sit$  augmenterait la distance conceptuelle.

Noter que toutes choses égales par ailleurs, plus le "rayon" de cette sphère égocentrée est faible, plus  $Ev_0$  sera improbable. L'art d'introduire un sujet sur le mode improbable implique donc de savoir sélectionner des faits "égocentriquement proches". Nous en verrons un exemple p. 37 avec l'effet de récence.

Deux interlocuteurs X et Y n'utiliseront pas le même univers pour le même événement :  $\Omega(X)$  et  $\Omega(Y)$  n'auront pas le même "centre". Une stratégie très largement employée par le deuxième locuteur Y pour banaliser un événement rare  $Ev_0$  est justement de montrer que l'estimation statistique de  $P(Ev_0)$  sur  $\Omega(X)$  est peut-être incorrecte. Examinons-là avec l'extrait suivant qui a eu lieu à la suite d'une discussion sur la relative douceur du temps en ce premier janvier :

[ex\_Nouvel-An]

contexte : cette conversation intervient après quelques remarques sur le temps qui était assez doux pour permettre aux enfants de jouer dehors. La température semblait exceptionnellement douce pour cette période de Nouvel An.

A1- De toutes manières, la température qu'on a aujourd'hui, c'est le MAXIMUM qu'on puisse espérer jusqu'à, jusqu'à la mi-février !

B1- Ca c'est vrai

C1- Je te dis, moi, en 77 y avait eu un temps de Föhn, l'hiver [...] Il faisait 20° au 1er janvier.

B2- Je me rappelle une année, quand nous étions petites, je m'en souviens très bien. Le 20 décembre, à peu près. [...] On avait, D et moi, des robes...les robes d'été ! Des robes d'été ! Et ça avait été une année vraiment très exceptionnelle.

Avec A1, A fait observer qu'une température aussi élevée que, disons, 16° est improbableement douce pour une période de Nouvel An :

$$\begin{aligned} \Pr(\text{température}(\text{1er\_janvier\_1987}) \geq 16^\circ) &= \Pr(\text{température}(D) \geq 16^\circ \mid \text{Nouvel\_An}(D)) \\ &= \sum_{D \in \Omega} \frac{\delta(\text{température}(D) \geq 16^\circ)}{\text{card}(\Omega)} \end{aligned}$$

où  $\Omega$  est l'ensemble des situations "égocentriquement analogues" à la situation courante. Si on remplace  $\Omega$  par  $\Omega(A)$ ,  $\Omega(B)$  ou  $\Omega(C)$  en fonction de la personne qui parle,  $\Omega(X)$  est l'ensemble des périodes de Nouvel An que X a vécues ou dont il peut se souvenir dans un passé récent (*i.e.* ici quelques années). A1 est perçue comme affirmant qu'aucune instance de  $(\text{température}(D) \geq 16^\circ)$  ne peut être satisfaite dans  $\Omega(A)$ . Donc la probabilité *a priori* de cet événement doit être très faible :

$$A1 : \quad \Pr(\text{température}(\text{1er\_janvier\_1987}) \geq 16^\circ) \ll 1 / \text{card}(\Omega(A))$$

Mais C est capable de mentionner au moins un exemple analogue de son propre échantillon qui entre dans la somme ci-dessus, indiquant par là que l'évaluation par A de la probabilité de la situation présente est sous-estimée.

$$C1 : \quad \text{température}(\text{1er\_janvier\_1977}) \geq 20^\circ$$

$$\begin{aligned} \Pr(\text{température}(\text{1er\_janvier\_1987}) \geq 16^\circ) &> \Pr(\text{température}(\text{1er\_janvier\_1987}) \geq 20^\circ) \\ &\cong 1 / \text{card}(\Omega(C)) \end{aligned}$$

L'effet de banalisation (voir p. 37) peut être estimé ici par le gain d'au moins un ordre de grandeur sur les probabilités, c'est-à-dire (pour fixer les idées) un peu plus de 3 bits ( $\log_2 10$ ). On voit aussi comment le fait

que C "en rajoute" en mentionnant une température de 20° augmente l'effet de banalisation (nous pouvons faire l'hypothèse que  $\text{card}(\Omega(X))$  est indépendant de  $X$ ).

#### 8.2.4. Rôle de l'analogie dans la banalisation d'un événement rare

Lorsque l'interlocuteur veut réaliser une banalisation d'un événement rare  $Ev_0$  en mentionnant un événement  $Ev_1$ , il est important que  $Ev_1$  appartienne à l'univers  $\Omega(X)$  sur lequel la probabilité  $P(Ev_0)$  a été estimée. Mais ce n'est pas toujours possible.

Par exemple la réplique B2 réalise elle aussi une banalisation. Mais elle démarre avec un handicap puisqu'elle remplace [*Nouvel\_An(D)*] par [*proche(D, Nouvel-An)*] : le nouvel événement [*température(D) ≥ 16° | proche(D, Nouvel-An)*] est plus facilement réalisé, d'autant plus qu'il inclut des dates *a priori* moins froides que le Nouvel An. Il est donc plus facile de mentionner une autre instance de cet événement affaibli.

Ainsi, lorsque l'événement  $Ev_1$  mentionné en guise de banalisation n'est pas parfaitement analogue à l'événement improbable  $Ev_0$ , l'univers  $\Omega(X)$  doit être augmenté pour englober à la fois  $Ev_0$  et  $Ev_1$ . Mais l'estimation de  $P_{\Omega'}(Ev_0)$  sur cet univers augmenté  $\Omega'(X)$  sera plus élevée que  $P_{\Omega}(Ev_0)$ . L'effet de banalisation (voir p. 34) est mesuré par la différence entre l'information accordée à  $Ev_0$  avant et après l'intervention, mais ici, les informations doivent être estimées par rapport au même univers  $\Omega'(X)$  :

$$\text{Banalisation par un événement analogue : } \log_2 \frac{\Pr_{\Omega'}^2(Ev_0)}{\Pr_{\Omega'}^1(Ev_0)}$$

Le changement de  $\Omega(X)$  en  $\Omega'(X)$  peut être néfaste à l'efficacité de la banalisation puisque  $\Pr_{\Omega'}^1(Ev_0) \geq \Pr_{\Omega}^1(Ev_0)$ . Nous comprenons maintenant pourquoi plus  $Ev_1$  est analogue à  $Ev_0$ , plus la banalisation est efficace. L'événement mentionné en B2, qui s'est produit aux environs du Nouvel An, n'est donc pas si bon que l'événement de C1 dans son action de banalisation. Voici pourquoi B insiste tant, d'un autre côté, sur le caractère improbable de la température atteinte en cette année exceptionnelle, renforçant en cela la contrainte ( $T \geq 16^\circ$ ).

Cet extrait est un court exemple de *tournoi narratif* (story round, [Tannen 1984]). Des tournois narratifs débutent fréquemment après une introduction sur le mode improbable. Après une première histoire racontant un fait particulièrement improbable, le locuteur suivant raconte une autre histoire à propos d'un fait analogue. Plus le fait est analogue, et plus il est efficace pour diminuer l'information apportée par la première histoire, puisque les deux événements appartiennent à la même catégorie (c.à.d. au même univers).

Cependant, l'histoire suivante d'un tournoi mène le plus souvent à un changement de sujet, et le nouveau sujet est alors considéré comme un fait improbable en soi. Mais les mécanismes d'enchaînement des conversations sont au delà de ce que nous voulons traiter ici.

#### 8.2.5. Effet de la récence dans la banalisation d'un événement rare

Très souvent, les interlocuteurs font part d'un fait récent lorsqu'ils introduisent un sujet sur le mode improbable, ou même attirent l'attention sur un fait présent ou un événement actuel. Cet effet de "récence" est un cas particulier de la nécessité d'avoir une sphère égocentrique de faible rayon (cf. p. 36). Un fait récent a plus de chances d'être *a priori* improbable. Avoir vécu un événement improbable dans les dix ans qui précèdent n'est pas en soi si improbable. Cela le devient si l'on regarde les dix jours ou les dix heures qui précèdent (en fonction de la personne à qui vous parlez). Ainsi, commencer une conversation en trouvant une histoire "incroyable" qui s'est produite depuis la dernière rencontre apporte plus d'information que s'il s'agit d'une vieille histoire. Vous pouvez dire "Devine ce qui m'est arrivé hier !", pour souligner le caractère inattendu d'un événement récent. Mais ce genre de question n'est pas admissible pour introduire une aventure trop ancienne, car beaucoup de réponses deviennent alors possibles.

Mais la "récence" joue aussi un rôle lors de la banalisation. Si l'on mentionne un fait  $Ev_1$  analogue à un événement  $Ev_0$  dans le but de banaliser celui-ci, alors un  $Ev_1$  récent sera plus efficace. Il s'agit d'une illustration du rôle joué par l'analogie dans la banalisation. Mais il est possible de modéliser plus précisément ces deux rôles de la récence, lors de l'introduction et lors de la banalisation.

L'événement  $Ev_0$  mentionné initialement est de toute façon un événement rare. Considérons que la probabilité *a priori* pour qu'un tel événement se produise  $n$  fois par unité de temps soit donné par une loi de Poisson :  $e^{-\lambda} \lambda^n/n!$ .  $\lambda$  est le nombre moyen d'occurrences de  $Ev_0$  par unité de temps. Sa valeur est inconnue, mais doit être considérée comme fixée au moment où  $Ev_0$  est émis (en fait le premier locuteur considère que  $\lambda$  est très petit).

Si  $Ev_0$  se produit au temps  $-t_0$  (l'instant présent étant pris comme origine), alors nous devons estimer la probabilité *a priori* que  $Ev_0$  se soit produit au moins une fois pendant la durée  $t_0$ . Cette probabilité est donnée par :

$$\Pr(Ev_0) = 1 - e^{-\lambda t_0} \cong \lambda t_0$$

Nous voyons donc qu'un événement récent ( $t_0$  faible) est plus improbable qu'un événement ancien, et apporte donc davantage d'information. Examinons maintenant la raison pour laquelle un autre événement  $Ev_1$ , qui vise à banaliser  $Ev_0$ , se doit aussi d'être récent.

L'objectif de  $Ev_1$  est de donner une estimation différente pour  $\lambda$ . En mentionnant  $Ev_1$ , le deuxième locuteur indique qu'un événement analogue à  $Ev_0$  (c.à.d. appartenant au même univers) s'est produit au moins une fois au cours de la durée écoulée  $t_1$ . L'effet est de suggérer que  $\lambda$  n'est pas si faible, et doit être remplacé par  $1/t_1$ . La nouvelle probabilité *a priori* pour que  $Ev_0$  se soit produit est donnée maintenant par :

$$\Pr(Ev_0) = 1 - e^{-t_0/t_1}$$

et la banalisation par l'événement récent est mesurée par :

$$\log_2 \frac{1 - e^{-t_0/t_1}}{1 - e^{-\lambda t_0}} \cong \frac{1}{\lambda t_1}$$

Ceci explique pourquoi l'effet de banalisation est meilleur si  $t_1$  est plus faible, parce que  $\Pr(Ev_0)$  est augmenté davantage. A cet égard, la réplique C1 n'est pas très efficace : si C avait dit "il y a dix ans nous avons eu la même température", la banalisation aurait été très faible. Mais C renforça la contrainte (température  $> 16^\circ$ ) et obtint ainsi un meilleur effet. B2, qui évoque un événement s'étant produit non pas dix ans, mais cinquante ans auparavant, est encore pire du point de vue de la récence. Mais encore une fois le locuteur augmente la contrainte de température en comparant la période du Nouvel An avec l'été.

### 8.3. Réactions à un fait (in)désirable

L'invalidation et la banalisation sont toutes deux possibles après une introduction sur le mode (in)désirable. Donnons en premier un exemple d'invalidation directe :

[ex\_placard]

contexte : A et B ont eu des problèmes d'humidité dans leur maison. Celle-ci n'a pas été chauffée pendant le week-end, et les vêtements sont encore froids dans le placard.

A1- C'est humide même là [dans le placard]

B1- C'est pas humide, c'est froid

B explique que toucher des vêtements froids donne l'impression indésirable qu'ils sont humides, mais qu'en réalité ils ne le sont pas.

La possibilité d'une invalidation, indirecte cette fois, dans le mode indésirable est illustrée par [ex\_channel, p. 29] D8. Mais le rôle d'invalidation indirecte peut être joué par des *suggestions*<sup>(11)</sup> comme [ex\_faim, p.25] B1 ou C1. Il n'y a pas de différence logique, mais les suggestions portent sur des actions,

<sup>11</sup> Noter que les suggestions peuvent apparaître comme coopérative au niveau sociologique. C'était le cas pour les cookies ou le hamburger dans [ex\_faim, p.25]. Mais du point de vue logique, elles jouent le rôle d'invalidations.

alors que d'autres types d'invalidations utilisent des faits. Cette distinction entre actes et faits n'est pas particulièrement intéressante ici.

Une réaction de *banalisation* peut aussi être produite dans le mode (in)désirable. Si nous étendons le concept de banalisation à toute diminution quantitative de l'Information Conversationnelle (comme nous l'avons définie p. 26), alors nous devons nous attendre à observer un effet de banalisation dans le mode (in)désirable chaque fois que  $|d|$  est diminuée. Nous avons pu observer une réaction de banalisation avec [ex\_carte-postale, p.21] B1. L'effet de banalisation peut être extrême, comme ici :

[ex\_corrosion]

contexte : B et son épouse vont acheter une voiture d'occasion. A les prévient des dangers liés au risque de rouille.

A1- *Votre problème, c'est d'avoir une voiture qui résiste à la corrosion.*

B1- *Non, notre problème, c'est pas la corrosion*

Si nous exprimons le contexte logique tout simplement par :

**non résistant\_corrosion**  $\Rightarrow$  *rouillé*

*rouillé*  $\Rightarrow$  **IND**

alors nous voyons que B1 ne nie aucune prémisse utilisée pour conclure à l'indésirabilité, et ne saurait donc être considérée comme une invalidation. B1 nie la modalité elle-même. Après B1, l'éventualité de la rouille qui était présentée comme indésirable par A1 devient neutre. En fait B n'a que faire du fait que sa voiture rouille. B1 agit donc comme une banalisation.

Maintenant, la situation ressemble à celle des introductions sur le mode improbable. Toutefois, certaines répliques admissibles dans le mode (in)désirable ne sauraient être considérées comme des invalidations ou des banalisations.

[ex\_voile]

contexte : discussion à propos d'un marin français qui concourait pour la Coupe de l'Amérique, en étant supporté par le public français.

A1- *Alors Marc Pajot s'est fait écraser encore.*

B1- *Deux minutes, c'est pas écrasé !*

A2- *Deux minutes quarante secondes, si.*

C1- *Oui, mais enfin, tu sais, hein, il a quand même tenu le coup jusqu'au dernier moment, hein, lui.*

A3- *Quel coup ?*

C2- *Hé ben enfin, il a pas sombré, il est pas tombé, il est arrivé !*

A4- *Tu veux dire qu'il est arrivé, il est arrivé en demi-finale*

C3- *en demi-finale, écoute !*

La première réplique de C ne remet pas en cause le fait indésirable que M. Pajot ait été battu lors de sa dernière régata dans la Coupe de l'Amérique (contrairement à la banalisation effectuée par B1). C parle d'un autre aspect de la situation, à savoir le fait désirable que M. Pajot ait atteint la demi-finale :

contexte de A1 : [*français(X) & important( Course ) & battu( X, Course )*]  $\Rightarrow$  **IND**

contexte de C1 : [*français(X) & important( Course ) & atteint\_demi-finale(X, Course)*]  $\Rightarrow$  **DES**

La même situation est présentée par C comme suffisant à provoquer simultanément l'insatisfaction et la satisfaction. Nous appellerons ce genre de réplique une *réaction antagoniste*. Le modèle prédit en fait quatre possibilités pour les réactions antagonistes, comme indiqué dans le tableau suivant :

première intervention	réplique antagoniste
$p \Rightarrow \text{IND} \text{ ou } \text{DES} \Rightarrow \text{non } p$	$p \Rightarrow \text{DES} \text{ ou } \text{IND} \Rightarrow \text{non } p$ $\text{non } p \Rightarrow \text{IND} \text{ ou } \text{DES} \Rightarrow p$
$p \Rightarrow \text{DES} \text{ ou } \text{IND} \Rightarrow \text{non } p$	$p \Rightarrow \text{IND} \text{ ou } \text{DES} \Rightarrow \text{non } p$ $\text{non } p \Rightarrow \text{DES} \text{ ou } \text{IND} \Rightarrow p$

L'extrait précédent appartient clairement au premier de ces quatre cas. Nous aurons l'occasion d'analyser des extraits illustrant les autres cas. La réaction antagoniste consiste ainsi à équilibrer les avantages d'une situation ou d'une solution par ses inconvénients, ou l'inverse, ce qui revient, en d'autres termes, à opposer le "pour" et le "contre".

Tous les résultats obtenus à ce point sont résumés dans la figure 4.

	paradoxe	improbabilité	(in) désirabilité
invalidation	+	+	+
banalisation	-	+	+
r. antagoniste	-	-	+

----- figure 4 : réactions admissibles en fonction de la modalité -----

Les cas absents de la figure sont "interdits", autrement dit on peut prédire qu'ils entraîneront des réactions d'incompréhension ou de "réparation", voire des réactions agressives.

Selon le modèle, tel que nous l'avons exposé jusqu'ici, il n'y a pas de possibilité de donner une réplique *pertinente* autrement qu'en invalidant ce qui vient d'être dit, ou en le banalisant, ou encore en donnant un argument antagoniste (deux autres possibilités de moindre importance vont être mentionnées dans une prochaine section p. 43). Ces trois possibilités ont en commun une caractéristique qui est peut-être à la base de l'art de converser.

#### 8.4. L'art de converser

Nous pouvons utiliser la représentation bidimensionnelle de la figure 1 pour illustrer l'effet de la seconde réplique sur le contexte (figure 5).

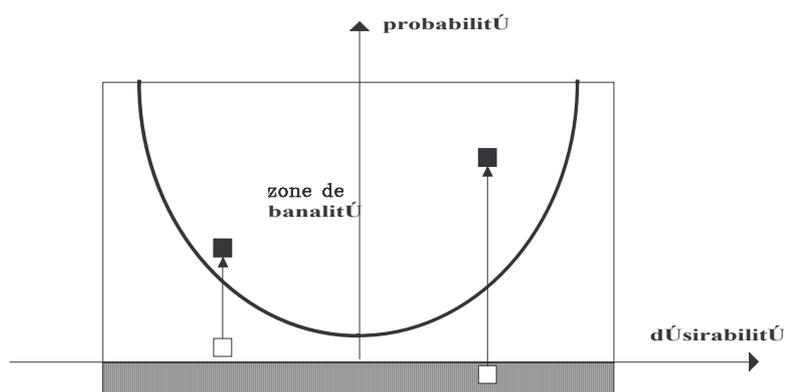


fig.5-a : effet d'une réaction de banalisation et d'une invalidation dans le mode improbable et le mode paradoxal respectivement

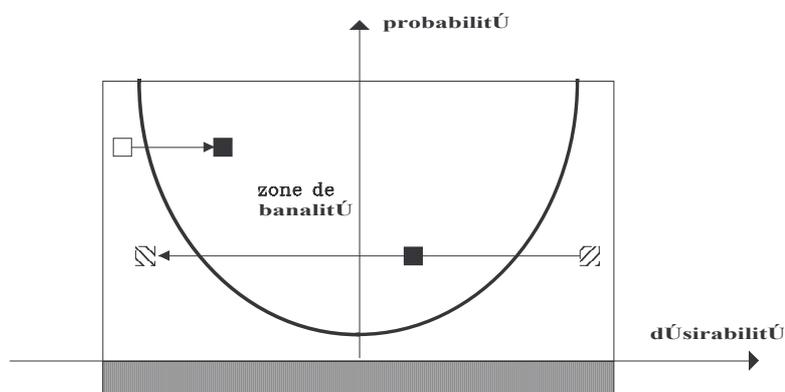


fig.5-b : effet d'une réaction de banalisation et d'une réaction antagoniste dans le mode indésirable

----- figure 5 : effet logique des répliques sur le contexte (voir fig. 3 p.28)-----

Chaque réplique déplace le point représentatif de la situation observée ou mentionnée d'un bord du cadre vers la zone de banalité. Après une introduction dans le mode paradoxal, un fait cesse d'être situé dans la zone d'impossibilité, mais il retrouve des coordonnées spécifiées par le reste de la connaissance partagée. La représentation de la réaction de banalisation se comprend d'elle-même (que ce soit dans le mode improbable ou dans le mode (in)désirable) et correspond à une diminution de l'intensité de la modalité. Une réaction antagoniste a pour résultat un compromis qui vient d'un équilibre entre deux valeurs de désirabilité extrêmes.

La première impression que l'on peut retirer de cette représentation est que, bien que la première intervention puisse être multi-modale (par ex. simultanément improbable et indésirable) si elle place le sujet dans un coin inférieur du diagramme, ce n'est pas le cas des répliques ultérieures. Il est en effet difficile de concevoir que la même réplique puisse par exemple augmenter la probabilité et la désirabilité en même temps. Voyons cela sur l'extrait suivant :

[ex\_Côme] [Morel 1983:44]

contexte : Le lac de Côme est un lac très allongé. A raconte son voyage en Italie.

A1- enfin ce Lac de Côme est insup - est vraiment insupportable -

B1- tous les bords de lac sont insupportables -

A2- il fait deux cents kilomètres de long - c'est incroyable ce truc -

En tant qu'observateurs extérieurs, nous pourrions percevoir A1 comme rapportant un fait indésirable. Mais la réaction B1 est une banalisation typique du mode improbable, et c'est bien ce mode que A2 souligne : ce lac est probablement ("incroyablement") long, et parcourir sa rive est probablement ennuyeux.

[*lac( Lac\_de\_Côte, Longueur, Largeur) & Longueur >> Largeur*] ⇒ IMPR

Il y a ainsi une distinction très nette entre les modalités improbable et indésirable lorsqu'il s'agit de répliquer. Mais qu'en est-il de la distinction paradoxe / improbabilité ? Voici un exemple tout à fait intéressant de ce point de vue :

[ex\_ping-pong]

contexte : A est intrigué par la technique étrange employée par un étudiant chinois qui joue au tennis de table.

A1- *T'as vu comment il tient sa raquette ?*

B1- *C'est un Chinois, les Chinois ils tiennent tous leur raquette comme ça*

A2- *C'est pas naturel ! Comment ils arrivent à jouer ?*

B2- *Oh ! C'est pas un sport ça !*

A3- *Enfin, tu transpires !*

B3- *Un sport où tu cours pas, c'est pas un sport.*

A4- *Alors le tennis, c'est la même chose...*

B4- *Ouais. Le basket tu cours oui.*

C1- *Le volley aussi.*

B5- *Tu parles !*

C2- *Enfin, à haut niveau, smasher et tout...*

A5- *Mais comment ils font pour smasher, les Chinois ?*

B1 est clairement une réaction de banalisation, ce qui montre que B comprend A1 comme rapportant un événement improbable. Mais il n'y a aucun doute pour nous que A, dans A1, était intrigué par un paradoxe, puisqu'il montre son étonnement en A2 ("Comment ils arrivent à jouer ?") et de nouveau en A5. C'est pourquoi il est nécessaire de faire une distinction qualitative entre paradoxe et improbabilité, puisque dans le cas d'un paradoxe, comme c'est le cas ici, une réaction de banalisation n'est pas recevable. A ne peut pas être satisfait par B1 parce qu'il a en tête un modèle logique qui l'amène à conclure que personne ne peut jouer correctement en tenant la raquette de cette façon. Et cette conclusion reste valide quand bien même un milliard de chinois joueraient de cette façon. L'étonnement de A n'a rien à voir avec une improbabilité, et ne pouvait être éliminé que par une invalidation qui n'est d'ailleurs jamais arrivée dans cet extrait<sup>(12)</sup>.

Ceci nous amène à quelques remarques sur l'art de converser. La plupart des linguistes considèrent la conversation comme un exemple typique de coopération entre les locuteurs, au niveau sociologique. Ceux-ci sont observés en train de construire une connaissance en commun en ajoutant des éléments de vérité à l'occasion de chaque réplique. Cette opinion influente est souvent attribuée à Grice :

"After Grice [1975], it is universally accepted that cooperation is the crucial feature which any theory of dialogue has to account for." [Airenti & al. 1989:148].

<sup>12</sup> Cet extrait est assez remarquable pour plusieurs raisons. L'argument B2 apparaît comme un méta-argument: si le ping-pong n'est pas un sport sérieux, alors la question de savoir s'il est possible de jouer en tenant la raquette de cette façon n'est plus intéressante. Les méta-arguments se rencontrent aussi par exemple quand un interlocuteur est accusé d'être de mauvaise foi. La discussion B2 --> C2 est donc une méta-discussion simplement parce que l'intérêt du sujet initial est en jeu (ce qui semble vite oublié). A tous autres égards, c'est une conversation normale, avec les invalidations A3, A4, B5, C2. Noter la parfaite explicitation d'une règle du contexte en B3, et la tentative de démonstration par l'absurde A4. Noter aussi l'effet du mot "smasher" qui réveille le paradoxe resté en suspens dans l'esprit de A en A5.

Toutefois, les exemples ci-dessus et leur interprétation dans le cadre de notre modèle indiquent que cette façon de considérer le comportement conversationnel des *interlocuteurs* est peu pertinente quand il s'agit de décrire la structure logique des *conversations* (voir note 1 p. 12). Au niveau logique, une bonne réplique n'est pas une simple élaboration constructive sur ce qui vient d'être dit. Au contraire, comme nous l'avons vu, c'est une sorte de destruction logique, fortement contrainte.

On pourrait argumenter sur ce point en montrant que ce que nous appelons une destruction (invalidation, banalisation, etc.) au niveau logique apparaît comme un acte coopératif à un niveau supérieur. Cela est possible. Par exemple une invalidation serait ainsi perçue comme un moyen d'aider une personne intriguée par un problème de compréhension. Mais on pourrait voir inversement une agression sociale dans l'acte qui consiste à souligner un manque de connaissance ou une faute de raisonnement dans le discours de l'autre [Ducrot 1972]. De toutes façons, nous ne voyons aucune raison de considérer systématiquement toute réplique comme sociologiquement coopérative. Discuter de questions sociologiques est hors de notre propos. Notre modèle ne porte pas sur les interlocuteurs, mais sur la structure logique de leurs productions.

En ce qui concerne ce niveau logique, nous pouvons donc difficilement décrire toutes les répliques comme coopératives<sup>(13)</sup>. Nous allons toutefois observer qu'elles ne sont pas toutes destructives.

### 8.5. *Taxonomie des répliques*

Les invalidations, les réactions de banalisation et les réactions antagonistes sont-elles les trois seules façons possibles de répliquer à l'introduction d'un nouveau sujet ? Observons l'extrait suivant :

[ex\_sapin]

contexte : trois semaines après Noël, le sapin est toujours vert.

A1- *Mais c'est toujours le sapin ? C'est formidable, ça !*

B1- *Mais oui ! Il n'est pas encore sec !*

A2- *Et sans racines ?*

B2- *Sans racines. Il perd même pas ses... On a déjà coupé toutes les branches du bas.*

A3- *C'est curieux qu'il ne perde pas ses...*

B3- *Ben oui, je ne comprends pas.*

A4- *Il est en plastique*

Les répliques de B, B1 et B2, pourraient être perçues comme de simples réponses aux questions de A (ne comportant aucun caractère "destructif"). Toutefois, comme B3 le révèle parfaitement, B perçoit le paradoxe exprimé en A1. Mais elle ne trouve aucun argument pour l'invalider (noter l'invalidation humoristique A4). Les répliques de B, tout particulièrement B3, sont des réactions de *co-étonnement*. Ce genre de réplique peut être rencontré à la suite d'introductions sur le mode paradoxal, mais aussi sur le mode improbable (par ex. [ex\_dard, p.33] B1). Après un début sur le mode (in)désirable, nous pouvons rencontrer des co-réjouissances ou des co-lamentations.

[ex\_vent]

contexte : le vent fait cruellement défaut pour la planche à voile

A1- *Y a pas beaucoup de vent*

B1- *Oui, c'est pas terrible !*

Ici le contexte est celui de la planche à voile, et le manque de vent est indésirable. Ces "co-réactions" ne modifient pas le contexte logique. On aurait peine à dire qu'elles contribuent à l'argumentation. Toutefois, elles offrent souvent l'occasion au deuxième locuteur d'en "rajouter" :

<sup>13</sup> Dans le cas de plus de deux interlocuteurs, on pourrait certes dire de certaines répliques qui sont dirigées contre une même intervention qu'elles coopèrent (par ex. [ex\_ping-pong, p.42] B4 et C1).

[ex\_armes]

contexte : Cette discussion arrive après un long échange sur l'efficacité comparée de différentes armes.

A1- *Enfin, voilà bien du pognon, hein ?*

B1- *Tu parles des armements ?*

A2- *mmm*

B2- *C'est faramineux*

A3- *Et ça diminue pas*

B3- *C'est le gros problème des russes. C'est pour ça que [...]*

Ici A1 et A3 sont sur le mode indésirable. B2 repousse le fait mentionné plus loin vers la gauche du diagramme de la figure 3, mais aussi vers le bas. Il s'agit peut-être de la part de B d'une tentative pour changer la modalité vers l'improbabilité.

Jusqu'à maintenant, nous avons pu voir que le modèle reconnaissait plusieurs possibilités pour les répliques : les répliques "destructives" (invalidations, banalisations, réactions antagonistes), et les "co-réactions". Mais qu'en est-il des *accords* ? De nombreux auteurs qui étudient les conversations au niveau sociologique décrivent incorrectement comme de simples accords des réactions qui ont en fait une fonction spécifique au niveau logique : des invalidations (comme [ex\_ping-pong, p.42] B4 où B valide une conclusion considérée comme absurde par A), des banalisations (comme dans [ex\_soif, p.34] B2), des co-réactions ([ex\_armes] B2), ou encore des clarifications (comme [ex\_armes] A2, voir prochain paragraphe). Les accords réels sont certes parfaitement compatibles avec le modèle. Toutefois, nous n'avons pas spécialement attiré l'attention sur eux parce qu'ils sont plutôt rares dans le corpus (peut-être parce qu'il ont tendance à clore le sujet, cf. plus loin [ex\_portes, p.46] B4).

Il faut mentionner une dernière possibilité de réplique. Certaines répliques ont pour objectif d'explicitier davantage le contexte logique. Nous avons pu en observer quelques exemples avec [ex\_train2, p.20] B1, [ex\_Goffman, p.15] A6, [ex\_armes] B1, A2 ou [ex\_Channel, p.29] C4, C5, D4, C9. Voici un exemple moins évident :

[ex\_gifle]

contexte : A est enseignant. Il passe pour être une personne fort paisible.

A1- *Jeudi dernier, c'est ma onzième année d'enseignement, j'ai mis la première gifle de ma..., depuis onze ans.*

B1- *Mais il était spécialement dur, comme gamin ?*

Le contexte logique est explicité dans A1. L'événement est présenté comme rare puisque A va jusqu'à donner une estimation de sa fréquence : il prétend que l'événement ne s'est produit qu'une seule fois en onze ans. Nous pouvons exprimer cela ainsi :

[ *ne\_gifle\_jamais(A) & gifle(A, garçon)*  ]  $\Rightarrow$  **IMPR**

([*ne\_gifle\_jamais(A)*] signifie que A n'a jamais giflé d'enfant en onze ans). Nous aurions pu demander à B pourquoi elle a émis B1. C'est ce que nous avons fait, quoique longtemps après que cette conversation s'est déroulée : "ça veut dire que c'était un gamin spécial, parce que sinon il n'aurait pas attendu onze ans pour mettre sa première gifle". Transformons légèrement cette formulation pour faciliter la traduction logique :

*ça veut dire que c'était un gamin spécial,  
sinon [il ne l'aurait pas giflé ou bien] il n'aurait pas attendu onze ans pour mettre sa première gifle*

Nous proposons de considérer ces deux phrases comme des contrapposées, et de représenter la première par :

[ *ne\_gifle\_jamais(A) & gifle(A, garçon)* ]  $\Rightarrow$  *spécialement\_dur(garçon)*

et la deuxième par :

**non** *spécialement\_dur(garçon)*  $\Rightarrow$  [ **non** *gifle(A, garçon)* **ou non** *ne\_gifle\_jamais(A)* ]

ce qui est totalement équivalent. Ainsi B demande une validation du contexte plus complet :

[ *ne\_gifle\_jamais(A) & gifle(A, garçon)* ]  $\Rightarrow$  *spécialement\_dur(garçon)*

*spécialement\_dur(garçon)*  $\Rightarrow$  **IMPR**

transférant ainsi le "scoop" sur la conduite supposée improbablement insupportable du garçon. B1 apparaît donc comme une réplique de *clarification de contexte*. Résumons les différentes possibilités de former une réplique admissible :

répliques "destructives"	co-réactions	accord	clarification de contexte
<i>invalidation</i> <i>banalisation</i> <i>réaction antagoniste</i>	<i>co-étonnement</i> <i>co-lamentation</i> <i>co-réjouissance</i>		

On pourrait penser que l'introduction de "co-réactions" rend le modèle non réfutable. Chaque fois que la réplique du deuxième locuteur va dans la même direction que la première intervention, nous pourrions dire qu'il s'agit d'une co-réaction; autrement elle serait décrite comme une réplique "destructive", et ainsi aucun exemple ne pourrait jamais contredire le modèle. Mais il ne saurait en être ainsi, car le modèle prédit bien plus :

- la structure d'une réplique "destructive" est parfaitement déterminée : une telle réplique doit effectuer une action logique définie sur le contexte (invalidation, banalisation, réaction antagoniste).
- les co-réactions sont aussi bien définies. Le second locuteur les emploie pour montrer qu'il est *au moins* aussi troublé (ou réjoui) par l'événement mentionné, soulignant en cela la modalité.
- les réactions de clarification sont aussi logiquement déterminées. Elles explicitent une autre version de la clause saturée.
- les réactions "vides", c.à.d. les réactions sans contenu spécifique, ne sont pas prises en compte dans le modèle (excepté les accords). Par exemple, des réactions comme "je ne pense pas que tu as raison" ne peuvent pas remplacer une invalidation, elles ne jouent aucun rôle logique, et sont perdues lors de la traduction logique sans dommage pour la représentation de l'argumentation logique.
- toutes ces réactions logiques admissibles laissent la place pour tout un éventail de réactions "logiquement neutres", c.à.d. des réactions "non vides" n'ayant pas d'effet sur le contexte logique. Mais **le modèle les exclut**, ou plutôt prédit que de telles réactions logiquement neutres seront perçues comme des marques de non-compréhension et encourageront le premier locuteur à expliciter davantage le contexte logique (par ex. [ex\_channel, p.29] A4 du point de vue de C).

En d'autres termes, selon notre modèle, une réplique doit avoir un effet en relation avec le contexte logique : soit un effet positif en soulignant la modalité (co-réactions) ou en explicitant le contexte logique (clarifications), soit un effet "destructif". Ceci nous conduit à énoncer le deuxième principe de pertinence logique :

### Deuxième principe de pertinence logique :

*Toute réplique vise à renforcer ou inversement à diminuer ou détruire la problématique.*

L'élaboration ou l'échange d'information, qui sont souvent considérés comme une caractéristique essentielle des conversations au niveau sociologique, doivent émerger de ces quelques possibilités d'actions logiques, principalement des actions "destructives".

Le plus surprenant ici est que le premier locuteur s'attend à des répliques destructives. Elles constituent en quelque sorte le "moteur" des conversations, comme le montre le fait que beaucoup de conversations s'arrêtent après une ou deux co-réactions ou acceptations. Cet effet dynamique des répliques destructives vient de la possibilité qu'elles soient à leur tour attaquées, comme nous allons le voir dans la prochaine section.

*Résumé* : Notre modèle est aussi restrictif pour les répliques qu'il l'est pour la première intervention. Une réplique logiquement pertinente doit avoir une action logique sur le contexte logique : invalidation, banalisation, action antagoniste. Ces réactions ont un effet "destructif" sur le contexte logique de la première intervention. Les autres possibilités sont : co-réactions, accord, clarification. Ceci exclut les répliques logiquement neutres et impose de fortes contraintes sur la deuxième intervention. La banalisation a fait l'objet d'une description plus détaillée : nous avons montré comment un locuteur pouvait diminuer l'information apportée par une première intervention dans le mode improbable.

## 9. Le déroulement des conversations

### 9.1. L'enchaînement des répliques

Pour montrer comment les répliques sont logiquement liées les unes aux autres, examinons un extrait plus long :

[ex\_portes]

contexte : A raconte qu'il est en train de repeindre les portes chez lui, et qu'il a décidé de commencer par décaper la vieille peinture, ce qui se révèle être un travail difficile.

A1- *Ben moi, j'en bave actuellement parce qu'il faut que je refasse mes portes, la peinture. Alors j'ai décapé à la chaleur. Ça part bien. Mais pas partout. C'est un travail dingue, hein ?*

[. . .]

B2- *Quelle chaleur ? La lampe à souder ?*

A3- *Ouais, avec un truc spécial.*

B3- *Faut une brosse, dure, une brosse métallique.*

A4- *Oui, mais j'attaque le bois.*

B4- *T'attaques le bois.*

[pause 5 secondes]

A5- *Enfin je sais pas. C'est un boulot dingue, hein ? C'est plus de boulot que de racheter une porte, hein ?*

B5- *Oh, c'est pour ça qu'il vaut mieux laiss... il vaut mieux simplement poncer, repeindre par dessus*

A6- *Ben oui, mais si on est les quinzèmes à se dire ça*

B6- *Ah oui.*

A7- *Y a déjà trois couches de peinture, hein, dessus.*

B7- *Remarque, si elle tient bien, la peinture, là où elle est écaillée, on peut enduire. De l'enduit à l'eau, ou*

A8- *Oui, mais l'état de surface est pas joli, quoi, ça fait laque, tu sais, ça fait vieille porte.*

A1 est une introduction sur le mode indésirable. B2 est une tentative de suggestion (c.à.d. une forme d'invalidation indirecte, cf. p. 38) : B avait semble-t-il en tête des systèmes à air chaud qu'il considère comme

moins efficaces, auquel cas la lampe à souder constitue une bonne suggestion. La suggestion B3 agit bien comme une invalidation indirecte :

*repeindre\_fastidieux* ⇒ **IND**

[*décaper & non brosse\_métallique & non repeindre\_fastidieux*] ⇒ **F**

B3 nie l'hypothèse supposée oubliée [**non brosse\_métallique**]. Noter que la brosse métallique est présentée comme *nécessaire* en B3 ("... faut une brosse"). Ceci peut être expliqué par une réécriture de la dernière clause :

[*décaper & non repeindre\_fastidieux*] ⇒ *brosse\_métallique*

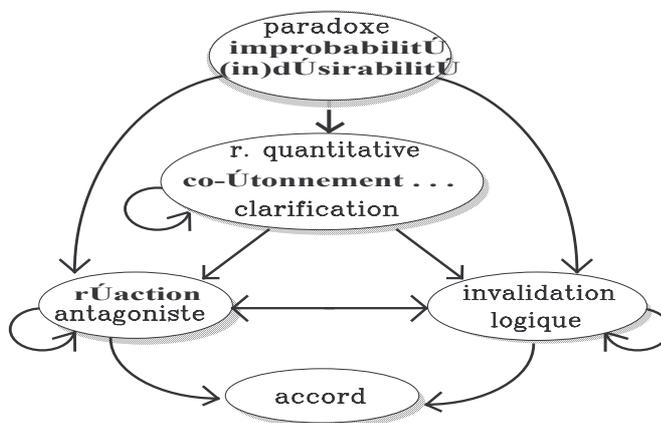
A4 apparaît alors comme une réaction antagoniste typique contre B3. L'usage d'une brosse métallique détruit le raisonnement conduisant à [*repeindre\_fastidieux*], et hérite en cela d'une valeur désirable. A4 ne met pas en question cette désirabilité de la brosse métallique qui est susceptible de faciliter le travail de décapage, mais elle montre que son usage a des inconvénients par ailleurs, à savoir que cela abîme le bois.

*bois\_abîmé* ⇒ **IND**

[ *brosse\_métallique & bois\_tendre & non bois\_abîmé* ] ⇒ **F**

Nous observons ici comment une réaction antagoniste peut être opposée à une invalidation. Nous avons la même séquence avec la suggestion B5 et la réaction antagoniste A6-A7, et une fois encore avec la suggestion B7 et la réaction antagoniste A8. Noter les "oui mais" qui débute souvent les réactions antagonistes.

[ex\_poulet, p.31] B1, A2, B2 nous ont offert un exemple d'une succession d'invalidations, qui est caractéristique d'une conversation sur le mode paradoxal : chaque réplique invalide la précédente. Le diagramme de la figure 6 résume les différentes possibilités d'enchaînement de répliques.



----- figure 6 : diagramme de transition des répliques -----

Nous donnons ci-dessous une description plus précise de l'enchaînement des répliques sous la forme d'une grammaire non contextuelle (figure 7).

CONV	-->	SUJET	CONV						
CONV	-->	.		LUT	-->	éton	LUT		
				LUT	-->	inv	inv	LUT	
SUJET	-->	PAR   IMP   IND   DES							
				IMP	-->	LUT	IMP		
PAR	-->	LUT	PAR	IMP	-->	dérar	IMP		
PAR	-->	inv	accord	IMP	-->	inv	accord		

IMP	-->	.				<u>légende :</u>
DES	-->	réj	DES			CONV conversation
DES	-->	dédram	DES			SUJET sujet de la conversation
DES	-->	anta	IND	DES		LUT lutte d'invalidations
DES	-->	LUT	DES			PAR paradoxe
DES	-->	inv	accord			IMP improbabilité
DES	-->	.				IND indésirabilité
						DES désirabilité
IND	-->	lam	IND			
IND	-->	dédram	IND			accord accord
IND	-->	anta	DES	IND		inv invalidation
IND	-->	LUT	IND			éton (co)-étonnement
IND	-->	inv	accord			dérar banalisation de l'improbabilité
IND	-->	.				réj (co)-réjouissance
						lam (co)-lamentation
						dédram banalisation de l'(in)désirabilité
						anta réaction antagoniste

----- figure 7 : grammaire conversationnelle -----

Dans ce schéma, les "co-réactions" et les clarifications sont présentées comme des réactions de transition. Les invalidations et les réactions antagonistes peuvent alterner ou boucler entre elles. Finalement, un accord peut venir clore le sujet, mais cela est loin d'être systématique.

Les boucles antagonistes se produisent souvent lorsque plusieurs solutions sont en compétition :

[ex\_remorque]

contexte : B et C veulent acheter une voiture break. A fait une autre suggestion, car les breaks ont été dépeints précédemment comme chers et difficiles à garer.

[...]

A1- Maintenant... il y a une solution pour économiser le break, c'est la remorque [...]

B1- Ouais, mais il faut voir aussi

A2- et on met dessus, quand on enlève la caisse, on y met [des tas de choses]

B2- Oui, mais le break, c'est pas mal aussi !

C1- Faut un garage [pour la remorque]

A3- Oui, mais si on l'utilise deux fois par an, balader un break !

C2- La remorque, moi, je peux pas la stocker.

B3- C'est pas plus gros qu'une voiture normale, un break.

Ici, la remorque est présentée en A1 comme une suggestion contre le coût élevé des breaks dont il a été question avant. A2 anticipe B1 comme une réplique antagoniste : selon B, mais pas pour A, un break offre davantage d'espace. Ensuite B2, C1, A3 et peut-être C2 peuvent être comprises comme des réactions antagonistes. En ce qui concerne C2, la question est de savoir si le problème de stockage est définitif, auquel cas C2 serait une invalidation. L'invalidation directe B3 termine l'extrait.

Pour comprendre comment la pertinence logique est à l'oeuvre ici, nous devons considérer que remorque et break s'excluent ici et constituent localement un ensemble exhaustif de solutions :

[ remorque & break ]  $\Rightarrow$  F

[ non remorque & non break ]  $\Rightarrow$  F

A2 explique qu'une remorque est désirable parce qu'elle offre beaucoup d'espace. B réplique en B2 que la négation de la remorque (*i.e.* le break) est aussi désirable. Avec C1, la remorque apparaît cette fois comme indésirable, mais sa négation est aussi indésirable en A3. Ainsi B et A utilisent ici deux des quatre moyens prévus par le modèle pour former des répliques antagonistes (voir p. 40).

Les flèches de la figure 6 ne doivent pas être prises trop littéralement comme indiquant des tours de paroles du jeu conversationnel. Nous allons illustrer ce point maintenant.

## 9.2. *La dynamique des répliques*

Une première situation à prendre en compte vient des cas où les locuteurs se répliquent à eux-mêmes, comme dans l'extrait suivant :

[ex\_break]

contexte : B et C veulent acheter une voiture break. A fait observer que ce genre de véhicule est cher.

A1- *Mais, heu, quand je disais une voiture spacieuse,... Ah oui, break. Evidemment, il y a de la place dans un break. Seulement les breaks sont rares et chers, je crois, ils sont recherchés.*

B1- *Non, non, ou ils sont*

C1- *Non, j'en ai vu une à vendre, mais il [B] a pas voulu acheter.*

A2- *et ensuite, ils ont des kilométrages impressionnants. Ils ont 200.000 bornes quand tu les...*

A1 débute avec la répétition d'une réplique donnée dans une conversation précédente au cours de laquelle il apprit que B et C allaient acheter une voiture non pas seulement spacieuse, mais break. Mais A se souvient soudain qu'il a déjà obtenu une réponse, et trouve un nouvel argument antagoniste : les breaks sont chers. Il produit encore un autre argument antagoniste en A2, comme si B1 et C1 n'avaient pas été émises. Nous voyons que les intervenants sont par moment capables de converser seuls, et que les transitions du diagramme de la figure 6 doivent être comprises comme indiquant des répliques logiques, indépendamment de celui ou de celle qui les a émises (voir note 1 p. 12).

Mais ce diagramme donne un sentiment de fuite en avant, chaque réplique étant détruite par la suivante. Nous avons pu observer ce type de progression avec [ex\_remorque] A1-A2 --> B2, C1 --> A3 --> C2. Mais les répliques ne s'enchaînent pas toujours de cette façon. [exportes, p.46] A5 ou [exping-pong, p.42] A5, par exemple, montrent que les locuteurs remontent (backtrackent) souvent pour répliquer à une intervention antérieure. Cette remontée peut être très courte ou bien assez distante, comme dans [ex\_break] A1. Illustrons le cas de remontées de courte portée :

[ex\_roues] [Maynard 1990:206]

contexte : "then the parties systematically introduced and pursued talk regarding cars, experiences with working on them, and also what kinds of cars the participants would ideally like, what they could afford, and other matters".

A1 - *Have you ever heard anything about wire wheels ?*

B1 - *They can be a real pain. They, you know, they go out of [ ? ? ] and*

A2 - *Yeah - The... if you get a flat you have to take it to a special place to get the flat repaired*

B2 - *why is that ?*

A3 - *because they're really easy to break. I mean to bend and damage*

B3 - *Oh really ?*

A4 - *And most people won't touch them unless they have the special, you know, equipment or they... they have the know how*

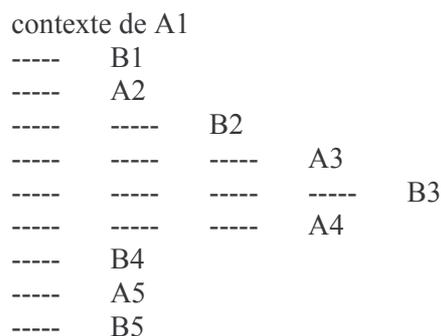
B4 - *They're like about two hundred bucks a piece or something too*

A5 - Yeah they and they're real expensive to have aligned

B5 - Yeah, you get them just chromed and that's the only way to have them just about too, you know

A6 - Yeah

Dans cet extrait, les deux participants rivalisent pour trouver la meilleure réaction antagoniste. Le contexte de A1 inclut le fait que les roues à rayons sont très appréciées des acheteurs de voitures de sport. Ainsi B1, A2, B4, A5, B5, qui mentionnent des inconvénients présentés par ce genre de roues, apparaissent comme des réactions antagonistes. Elles sont au même niveau, car elles sont toutes dirigées contre le contexte de A1. Nous pouvons représenter la hiérarchie dynamique de cet extrait (à la manière d'A. Trognon [1990]) de la façon suivante :



Ce schéma montre clairement le phénomène de remontée. On peut aussi noter le fait que A4 réalise aussi une remontée, puisque comme A3 il s'agit d'une explication (*i.e.* une invalidation) du contexte de B2<sup>1</sup>). De telles remontées de courte portée sont en conflit avec les prédictions de la grammaire conversationnelle de la figure 7. Mais les grammaires formelles sont connues pour être incapables de représenter de tels phénomènes de retour-arrière [Sabah 1988]. Nous pouvons voir dans les conversations la trace (avec seulement les symboles terminaux) de l'application des règles de cette grammaire.

Certaines remontées sont parfois rendues nécessaires parce que des conversations entières jouent le rôle d'une réplique unique. Nous avons eu des illustrations de ceci dans les tournois narratifs, où chaque nouvelle histoire joue le rôle d'une banalisation par rapport à la précédente. Nous avons signalé aussi le cas de la méta-conversation [ex\_ping-pong, p.42] B2-->C2 dans laquelle la pertinence de A2 est en jeu (voir note 12 p. 42), et qui est suivie par la remontée A5.

De nombreuses remontées peuvent être aussi observées lors de conversations assez longues, si bien que certaines répliques peuvent attaquer des interventions antérieures fort lointaines. Le corpus contient une conversation de 355 répliques sur la conception d'un repas gastronomique qui s'étend sur plus d'une demi-heure. La conversation commença par une question : "Qu'est qu'on va faire à manger quand Bill va venir ?", qui révèle le fait indésirable de n'avoir encore aucune idée du menu qui sera offert à la personne invitée. Le fait que cette personne soit étrangère fait de cet état de choses un problème non trivial. Ce fait va être invalidé périodiquement au cours de la conversation :

[invité(Bill) & non connu(menu)] ⇒ IND

[connu(menu) & non connu(plat\_de\_résistance)] ⇒ F

[connu(menu) & non connu(hors\_d'oeuvre)] ⇒ F

[connu(menu) & non connu(fromage)] ⇒ F

...

<sup>1</sup> Nous interprétons B2 de la manière suivante: B a des raisons de penser que le type de la roue est sans influence sur les problèmes concernant les pneus. A2 contredit cela. B s'étonne de ce paradoxe, et exprime sa surprise en posant la question B2.

Comme indiqué ci-dessous, suggérer ici un plat de résistance est une façon d'invalider l'un des raisonnements possibles qui mènent à **IND** :

**non connu**(plat\_de\_résistance)  $\Rightarrow$  **non connu**(menu)

[*invité*(Bill) & **non connu**(menu)]  $\Rightarrow$  **IND**

Mais [**non connu**(plat\_de\_résistance)] hérite de l'attribut indésirable, et toute conversation s'ensuit, autour du problème de trouver un plat de résistance convenable. Les interlocuteurs peuvent ensuite remonter pour invalider un autre raisonnement :

**non connu**(hors\_d'oeuvre)  $\Rightarrow$  **non connu**(menu)

[*invité*(Bill) & **non connu**(menu)]  $\Rightarrow$  **IND**

et ainsi de suite jusqu'à ce que personne ne puisse plus trouver de moyen de prouver [**non connu**(menu)].

Le chaînage des répliques tel qu'il est prédit par le modèle est représenté statiquement dans la figure 6. Nous venons de donner quelques indications sur sa dynamique. Ce modèle montre que des contraintes strictes limitent l'éventail des actions logiques admissibles qu'une réplique peut produire lorsqu'elle est émise. Mais nous n'avons donné jusqu'ici que peu d'indications sur la façon dont les répliques sont effectivement conçues et sur la manière d'utiliser la connaissance des contraintes conversationnelles pour décider de la pertinence logique d'une réplique donnée. Le chapitre I.3 a pour objectif de donner des moyens de progresser dans cette direction.

*Résumé* : Nous avons montré dans cette section que les conversations consistent entièrement en une succession d'invalidations, de réactions antagonistes, etc. (et de rien d'autre !) et que de nombreuses combinaisons sont possibles. Mais la contrainte de pertinence logique s'applique tout au long de la conversation : chaque réplique doit avoir un effet logique sur le contexte. Toutefois, ce contexte n'est pas nécessairement celui de l'intervention qui précède : les répliques peuvent remonter (backtracker) et agir sur des contextes antérieurs.

## 10. Implications cognitives

L'analyse logique de la conversation contribue à l'analyse du langage naturel au niveau qui est le sien, comme il est indiqué en figure 8.

...
sociologie
...
stratégies argumentatives
pragmatique logique
sémantique
syntaxe
morphologie
phonologie

----- figure 8 : niveaux d'analyse du langage -----

Il y aurait plusieurs raisons de distinguer plusieurs niveaux dans le langage, mais la plus évidente vient du fait que des productions linguistiques peuvent être acceptables à un niveau et incorrectes à un autre niveau. Par exemple "le jardin de la porte mange le ciel" est syntaxiquement correcte, mais semble sans signification. La phrase "il y a des gens dans le monde qui ont plus de douze ans" est sémantiquement parfaitement acceptable, mais il y a des contextes dans lesquels elle ne peut pas être pertinente. Pour illustrer la nécessité de considérer la pragmatique comme logique, il suffit de noter que certaines répliques peuvent

être associées au contexte, mais n'exercent pour autant aucun effet logique et n'ont donc pas de pertinence logique (par ex. [ex\_channel, p.29] A4). Nous traiterons cette question de la pragmatique dans le chapitre I.2.

Certaines des lois identifiées à un niveau linguistique sont parfois considérées comme générales, voire universelles. Dans quelle mesure les contraintes logiques décrites ici pourraient-elles être universelles ? Sont-elles limitées à la culture occidentale ? Les deux réponses possibles à cette dernière question seraient du plus haut intérêt. De nombreuses études montrent que la logique joue un rôle important dans le discours, dans des cultures non apparentées (cf. [Hutchins 1980:67], [Taylor 1989:277]), mais des études spécifiques sur la conversation spontanée sont encore nécessaires. Et parmi les personnes qui partagent la même culture, peut-on trouver des groupes ou des individus qui n'accorderaient pas d'importance à la pertinence logique ? Une réponse facile serait de dire qu'un tel fait, s'il était vrai, serait connu.

Il est en fait connu, mais pour une sorte très particulière de personnes dont on dit qu'elles présentent des "troubles mentaux". Voici une conversation avec une personne malade mentale, enregistrée par A. Trognon [1988] :

[ex\_chanter]

contexte : cette conversation a été enregistrée par l'enquêteur alors qu'il parlait avec une malade mentale.

A1- *Vous ne chantez plus, là ?*

B1- *Je chante mais c'est faux.*

A2- *C'est faux ?*

B2- *C'est faux.*

A3- *Oh faut peut-être continuer un peu, non ?*

B3- *Oh mais j'ai été à l'école, moi !*

A4- *Ouais ?*

B4- *A l'école, on m'a appris à connaître la vie hein... connaître les gens quoi... tout ça les décès les mariages*

A5- *Oui*

B5- *soit c'est des décès, soit c'est des mariages... mais c'est pas bien les décès par exemple... vaut mieux pas décéder.*

Il semble malaisé de communiquer avec des malades mentaux. Ici, les contraintes logiques ne sont pas les seules règles transgressées : B3 est un changement de sujet abrupt qui ne tient pas compte des règles de transition. Mais c'est aussi une mauvaise introduction de sujet. Nous ne pouvons pas trouver de contexte vraisemblable qui rende B3+B4 paradoxal, improbable ou (in)désirable. Finalement, B parvient à émettre une proposition indésirable à la fin de B5, mais aucun contexte intéressant ne la rend pertinente.

L'analyse logique pourrait permettre de donner une définition scientifique et un diagnostic de ce que l'on nomme troubles de la conversations. Certains de ces troubles ont été d'ailleurs corrélés avec des dysfonctionnements logiques [Watzlawick & al. 1977], [Bateson & al. 1956]. Mais les malades mentaux ne sont pas les seuls à émettre des répliques qui sont logiquement inappropriées. Les jeunes enfants sont souvent déconcertants avec leur raisonnements erronés, leurs justifications étranges, etc. Pourtant, les extraits de McTear [1985] montrent que des enfants de moins de six ans sont parfois capables de discuter entre eux selon un schéma logique précis. Il semblent sensibles aux trois modalités (paradoxe, improbabilité, (in)désirabilité), mais leurs contextes logiques semblent très simples.

Il devient clair maintenant que la conversation n'est pas une interaction sociale dépourvue de structure logique, ou dotée d'une structure logique plus ou moins aléatoire qui dépendrait de circonstances externes. Les conversations ont une structure logique intrinsèque qui, à n'en pas douter, contribue à rendre possible ce moyen de communication sociale.

Mais il peut être aussi intéressant de noter que les contraintes logiques que les protagonistes s'imposent mutuellement sont cognitivement justifiées. Par exemple, en discutant de paradoxes, ils ont l'occasion de

"réparer" les incohérences qui corrompent leur connaissance. Cela doit être bénéfique, si l'on se souvient à quel point le premier locuteur de [ex\_Goffman, p.15], de [ex\_ping-pong, p.42] ou de [ex\_sapin, p.43] insistait pour obtenir une invalidation adéquate. En discutant à propos d'événements improbables, les interlocuteurs vont revoir les coefficients de probabilité qu'ils assignent à chaque événement. Cela se révèle essentiel si l'on pense que chaque décision de la vie courante nécessite plusieurs estimations probabilistes. Enfin, en conversant sur le mode (in)désirable, ils profitent de l'expérience des autres, ils imaginent et testent ensemble des plans d'action qui leur éviteront des expériences désagréables ou les amèneront plus près de situations souhaitables.

Nous n'avons pas parlé ici d'autres types de contraintes qui limitent la liberté de l'interlocuteur, parce qu'elles ne sont pas strictement "logiques", et sortent donc du cadre de ce travail. Mais de telles contraintes, comme par exemple les règles de transition entre sujets, ont elles aussi leurs implications cognitives.

Ayant pris conscience de l'importance des contraintes logiques dans la conversation, nous pouvons imaginer plusieurs applications dans des domaines comme l'intelligence artificielle, l'Enseignement Assisté par Ordinateur (chap. III.2), la reconnaissance de la parole, le transfert d'expertise (chap. III.4), etc.

## 11. Conclusion

Décrire les conversations comme un échange d'information semble être un constat bien faible et bien vague. Les conversations sont le processus privilégié par lequel les individus se transmettent mutuellement une connaissance logique pertinente. La conversation, et ses règles, obligent celui qui veut s'engager dans une interaction sociale de type conversationnel à traduire au moyen du langage une partie de sa connaissance, et à en faire part d'une manière bien codifiée. Lorsque nous voulons exprimer une idée, nous utilisons notre compétence linguistique pour coder cette idée en mots. Nous prenons grand soin d'utiliser des mots existants et une syntaxe acceptable. Nous énonçons des phrases qui ont un sens de manière à être compris. Mais cet encodage de nos pensées n'est pas terminé si nous ne les plaçons pas dans un contexte où elles jouent un rôle logique.

Ignorer la dimension logique des conversations, c'est se condamner à ne pas comprendre leur fonction essentielle. Nos cerveaux ont besoin d'échanger non pas seulement des faits, mais aussi et surtout des relations conceptuelles de type logique. Ils peuvent élaborer ces dernières par l'expérience ou l'imitation, mais le moyen de loin le plus rapide et le plus efficace est de profiter directement des règles élaborées par les autres. Comment s'étonner que ce transfert d'une connaissance, qui peut s'avérer dans certains cas complexe, nécessite que des règles précises soient respectées ?

**Remerciements :** Je suis reconnaissant à Lamia Choukair, Alain Grumbach, Georges Sallé et à un lecteur anonyme qui ont bien voulu lire des versions antérieures de ce papier et me faire part de leurs critiques.



## Annexe : quelques règles de réécriture logiques

Le lecteur n'a besoin que d'une connaissance très sommaire de la logique formelle pour saisir le sens des formules qui expriment le contenu logique des conversations. Pour sa commodité, je rappelle ici quelques transformations qui s'avèrent utiles pour transformer syntaxiquement ces formules.

Les symboles de base :

$p \Rightarrow q$	implication logique	(Si p, alors q)
<b>non</b> p	négation	(le contraire de p est vrai)
p & q	conjonction	(p et q sont vrais simultanément)
p <b>ou</b> q	disjonction	(de p et de q, l'un des deux au moins est vrai)

Les symboles p, q, etc. tiennent lieu de propositions logiques. Ils peuvent être remplacés par n'importe quelle formule syntaxiquement correcte liant des propositions logiques, ce qui permet d'enchaîner les manipulations syntaxiques.

Quelques expressions équivalentes :

<b>non</b> (p & q)	équivalent à	<b>non</b> p <b>ou</b> <b>non</b> q
<b>non</b> (p <b>ou</b> q)	équivalent à	<b>non</b> p & <b>non</b> q
$p \Rightarrow q$	équivalent à	<b>non</b> q $\Rightarrow$ <b>non</b> p
$p \Rightarrow q$	équivalent à	<b>non</b> p <b>ou</b> q
<b>non</b> (p $\Rightarrow$ q)	équivalent à	p & <b>non</b> q
$p \Rightarrow$ (q <b>ou</b> r)	équivalent à	( p & <b>non</b> q) $\Rightarrow$ r
$p \Rightarrow$ (q & r)	équivalent à	( p $\Rightarrow$ q) & (p $\Rightarrow$ r)
(p <b>ou</b> q) $\Rightarrow$ r	équivalent à	( p $\Rightarrow$ r) & (q $\Rightarrow$ r)

Modalités utilisées pour traduire les conversations :

$p \Rightarrow$ <b>F</b>	fausseté	(p est fausse)
<b>V</b> $\Rightarrow$ p	vérité	(p est vraie)
$p \Rightarrow$ <b>IMPR</b>	improbabilité	(p est hautement improbable)
<b>PROB</b> $\Rightarrow$ p	probabilité	(p est hautement probable)
$p \Rightarrow$ <b>IND</b>	indésirabilité	(p suffit à me rendre insatisfait(e))
<b>IND</b> $\Rightarrow$ p	indésirabilité	(si je suis insatisfait(e), alors forcément p)
$p \Rightarrow$ <b>DES</b>	désirabilité	(p suffit à me rendre satisfait(e))
<b>DES</b> $\Rightarrow$ p	désirabilité	(si je suis satisfait(e), alors forcément p)

Ces modalités sont conçues pour se comporter comme des propositions dans les manipulations syntaxiques, avec les conventions suivantes :

<b>non</b> <b>F</b>	équivalent à	<b>V</b>
<b>non</b> <b>IMPR</b>	équivalent à	<b>PROB</b>
<b>non</b> <b>IND</b>	équivalent à	<b>DES</b>

Toute formule de la logique des propositions qui est syntaxiquement correcte peut être réécrite comme la conjonction de *clauses paradoxales* (clauses négatives) :

$$[ p_1 \& p_2 \& \dots \& p_n ] \Rightarrow \mathbf{F}$$

La signification d'une telle clause est que  $p_1, p_2, \dots, p_n$  sont incompatibles : elles ne peuvent pas être vraies toutes simultanément. L'introduction des modalités de probabilité ou de désirabilité conduira à des clauses du type suivant :

$$[ p_1 \& p_2 \& \dots \& p_n ] \Rightarrow \mathbf{IMPR}$$

$$[ p_1 \& p_2 \& \dots \& p_n ] \Rightarrow \mathbf{IND}$$

$$[ p_1 \& p_2 \& \dots \& p_n ] \Rightarrow \mathbf{DES}$$

Le calcul des prédicats comporte l'usage de variables, sur lesquelles peuvent porter deux quantificateurs :

$\forall \mathbf{x} \in X; p(\mathbf{x})$	quantificateur universel	(la propriété $p()$ s'applique à tous les éléments de l'ensemble $X$ )
$\exists \mathbf{x}_0 \in X; p(\mathbf{x}_0)$	quantificateur existentiel	(la propriété $p()$ s'applique au moins à l'un des éléments de l'ensemble $X$ )

Les manipulations syntaxiques sont les mêmes que pour les propositions, avec la règle supplémentaire suivante :

$$\mathbf{non} ( \forall \mathbf{x} \in X; p(\mathbf{x}) ) \quad \text{équivaut à} \quad \exists \mathbf{x} \in X; \mathbf{non} p(\mathbf{x})$$

Toutes les équivalences données ici peuvent bien entendu s'appliquer dans les deux sens et se combiner entre elles.

# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

On reconnaît généralement que le phénomène du langage peut être modélisé selon différents *niveaux*. Notre modélisation de la communication conversationnelle aborde le niveau que l'on qualifie de *pragmatique*. Pourtant elle ne fait pas directement appel aux modèles de la communication développés par les "pragmaticiens", et réciproquement les théories pragmatiques fondatrices ne prennent pas en compte, contrairement à nous, l'aspect bidirectionnel de l'échange conversationnel.

Nous avons jugé dans un premier temps que les théories pragmatiques, notamment celle de Grice [1975], pour intéressantes qu'elles soient, ne nous permettaient pas de progresser dans notre problème d'analyse des extraits conversationnels réels. Notre propre approche de la pertinence conversationnelle [Dessalles 1985] est ainsi étrangère aux cadres théoriques classiques édictés dans le domaine de la Pragmatique. Cependant, certains aspects de théories comme celle de Sperber et Wilson [1986] rejoignent parfois des préoccupations pour lesquelles les contraintes conversationnelles que nous avons mises en évidence semblent fournir une réponse.

Nous proposons ici de comparer l'essentiel de la théorie de Grice, ainsi que celle de Sperber et Wilson, telles que nous sommes capables de les décrire, avec les prédictions de notre propre modélisation. Nous essaierons d'indiquer que ces théories pragmatiques, lorsqu'on les applique à la conversation, sous-estiment l'efficacité de la communication en contexte et sous-déterminent le comportement conversationnel.

## 2. Le problème de la communication en situation

### 2.1. *Le modèle digital*

Dans un acte de communication, ce qui est réellement échangé n'est pas de même nature que ce qui est communiqué. Voyons cela sur des exemples<sup>(1)</sup> :

- (1) je viendrai demain
- (2) Babette s'est achetée une glace
- (3) Pierre est vraiment un chic type
- (4) une tasse de café m'empêchera de dormir

---

<sup>1</sup> Contrairement à notre habitude, les exemples que nous prenons ici sont artificiels, comme semble l'exiger la tradition pragmatique à laquelle nous nous référons dans ce texte. Pour que la comparaison soit plus aisée, nous choisissons la plupart de ces exemples artificiels parmi ceux fournis par Sperber et Wilson dans leur livre de 1986 (traduit en français).

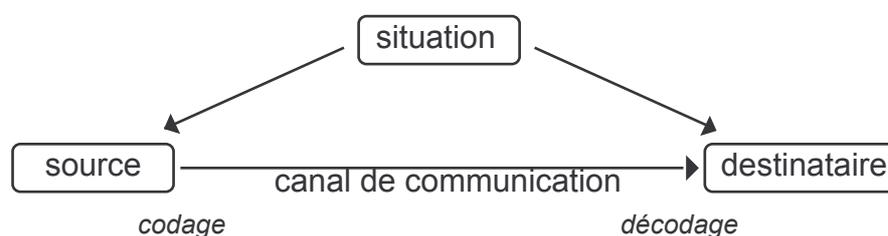
(5) Robert a acheté le Figaro

Ce qui est communiqué, dans chaque cas, n'est pas, littéralement, contenu dans ce qui est dit. L'interprétation de (1) suppose par exemple que l'on sache qui est "je" et à quelle date la phrase est prononcée. Dans (2), le mot *glace* est ambigu. (3) peut être sincère ou ironique. Si (4) répond à une offre, s'agit-il d'un refus ou d'une acceptation ? Robert a-t-il, dans (5), acheté l'entreprise qui fabrique le journal "Le Figaro" ? De plus, ce qui est communiqué va peut-être bien au-delà de ce que nous venons de suggérer. Par exemple, (2) signifie peut-être que Babette va grossir, ou que le locuteur souhaite aussi avoir une glace, ou que le fait d'acheter une glace par grand froid est inhabituel, etc.

Or c'est bien là que réside le problème principal que la Pragmatique s'attache à résoudre. De même que les théories syntaxiques se doivent de reproduire la performance des individus qui sont capables de juger de la grammaticalité des phrases, les théories de la Pragmatique ont pour premier but de déterminer, à partir de ce qui est dit, ce qui est "réellement" communiqué. Or ce problème posé à toute théorie de la pragmatique est un problème résolu tout naturellement par les individus qui communiquent ! "Ce qui est communiqué" acquiert ainsi une existence empirique, et le défi lancé aux "pragmaticiens" est d'expliquer et de prévoir cet objet d'expérience. Mais par quel moyen va-t-on déterminer ce que les phrases ci-dessus communiquent ?

Il est clair pour tout le monde que la situation d'énonciation est absolument nécessaire pour que l'interprétation soit possible. Mais est-elle suffisante ?

Dans (1), la connaissance de la situation permet de savoir qui viendra et à quelle date. La connaissance de la situation permet donc bien d'aller au-delà de ce qui est dit. On pourrait penser qu'une connaissance suffisamment complète de la situation d'énonciation permettrait une détermination complète de ce qui est communiqué, comme figuré dans le schéma suivant :



Ce schéma est largement inspiré du paradigme introduit par Claude Shannon dans la théorie des communications. Il suggère que les contraintes du canal de communication imposent des processus de codage et de décodage, autrement dit que la communication soit de type "digital" et non pas "analogique". Dans une communication de type analogique (penser à la transmission d'un morceau de musique joué par un orchestre), toute l'information que l'émetteur souhaite transmettre est contenue dans le signal émis qui lui *ressemble* : par exemple le signal hertzien de la radio ressemble à la pression acoustique produite par les instruments de musique. Dans une communication de type digital, en revanche, cette ressemblance est (en grande partie) perdue. Les opérations de codage et de décodage changent complètement la nature de l'information. Par ex. la partition d'orchestre n'entretient quasiment aucune ressemblance avec les vibrations acoustiques de la salle de concert. Il faut connaître le code de la transcription musicale pour profiter de la partition, alors qu'aucun code n'était nécessaire pour profiter de la retransmission radiophonique.

La compréhension du langage que l'on a acquise au cours de ce siècle tend à indiquer que la plupart des caractéristiques de la communication verbale humaine sont de type digital. La langue permet un encodage de la signification dans des phonèmes, dans des mots, dans des phrases. Qu'en est-il du niveau pragmatique ?

Nous avons constaté que le signal émis était beaucoup plus restreint que l'effet produit. La communication en situation ne peut donc pas être de type analogique<sup>(2)</sup>. Existe-t-il pour autant un code qui permet d'accéder à toute la signification, comme le schéma précédent le suggère ? Les avis des pragmaticiens

<sup>2</sup> Il y a peut-être certains aspects analogiques dans la communication. Par exemple dans l'accentuation ("*Cet enfant est impossible*"). Mais cela joue un rôle marginal dans notre propos.

sur ce point fondamental sont partagés. Il existe des modèles de ce type ("modèles du code"), et notre propre modèle en est un à bien des égards. Mais la plupart des théories pragmatiques, tout en acceptant implicitement le mode digital de la communication, refusent l'existence d'un code. Ce sont notamment les théories dites "inférentielles" (qui sont parfois considérées, à tort, comme s'opposant au modèle de la communication digitale).

## 2.2. Le modèle inférentiel

Un modèle inférentiel<sup>(3)</sup> va à l'encontre d'un décodage qui se limiterait à une simple mise en correspondance message / signification. Un tel décodage ne peut pas rendre compte de l'interprétation de (6a) selon (7) :

(6a) : Paul n'est jamais en retard

(6b) : soit c'est Marie qui est en avance, soit c'est Paul qui est en retard

(7) : Marie est en avance

Un modèle inférentiel est un processus complexe faisant appel à des procédures comme le syllogisme. Si (6a) est le message, la signification (7) est produite grâce à la connaissance (6b) qui n'est pas un fait objectif de la situation (il s'agit plutôt d'une connaissance, d'une théorie sur le comportement de Paul et de Marie). On peut définir des inférences qui ne soient pas strictement de type logique, comme les inférences graduelles, qui semblent permettre de rendre compte de phénomènes argumentatifs liés à la langue [Racah 1990; 1991]. Mais l'important ici est de constater que dans un modèle inférentiel, la signification est *construite*, et que ce processus de construction ne se satisfait pas de la seule connaissance de la situation objective d'énonciation.

Pour préciser ce point, nous pouvons écrire de manière symbolique de quels paramètres dépend l'interprétation  $I$  élaborée par l'auditeur  $j$  d'un message  $m$  émis par le locuteur  $i$  :

$$I(m(s_i, c_i), s_j, c_j)$$

avec les notations suivantes :

$I()$  : interprétation construite par le destinataire

$m()$  : message émis par la source

$c_i$  : connaissance possédée par  $i$

$s_i$  : situation évoquée ou présente, perçue par  $i$

L'adoption d'un modèle inférentiel pose alors un problème redoutable : d'où vient cette connaissance nécessaire à l'interprétation ? Si chacun utilise sa propre connaissance, comment garantir que la signification construite par le destinataire sera identique (ou aura au moins un rapport étroit) à la signification voulue par la source du message ?

Les réponses apportées par les théories pragmatiques diffèrent. Nous évoquerons le point de vue de H.P. Grice, puis celui de D. Sperber et D. Wilson, pour ensuite faire nos propres suggestions.

## 3. Grice : la communication vue comme une coopération

Nous nous basons sur l'article [Grice 1975] pour donner notre propre interprétation de la théorie de H.P. Grice. Le modèle de Grice, comme tous ceux que nous évoquerons par la suite, est un modèle *inférentiel*<sup>(4)</sup>.

<sup>3</sup> Nous empruntons largement la terminologie et la problématique du modèle inférentiel, ainsi qu'une partie de notre critique du modèle de Grice, au livre de Sperber et de Wilson [1986].

Toutefois, la construction de la signification est guidée par un certain nombre de *normes* qui limitent les risques de divergence entre la signification voulue par la source et celle que le destinataire va produire.

Pour Grice, ces normes dérivent toutes d'un principe unique, le principe de coopération. Toute contribution à la conversation se doit d'être coopérative, c.à.d. d'aller dans la "direction mutuellement acceptée" de l'échange :

"Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts ; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction. This purpose or direction may be fixed from the start (*e.g.*, by an initial proposal of a question for discussion), or it may evolve during the exchange ; it may be fairly definite, or it may be so indefinite as to leave considerable latitude to the participants (as in casual conversation). But at each stage, SOME possible conversational moves would be excluded as conversationally unsuitable."  
[Grice 1975].

Le principe de coopération permet à l'auditeur, selon Grice, de faire les inférences qui vont lui permettre de passer du message à la signification. Si la source du message peut être considérée comme coopérative, alors de nombreuses suppositions peuvent être faites, et notamment :

- le message contient juste les éléments d'information qui permettent la construction de sa signification (maxime de quantité)
- le message ne contient pas d'éléments faux ou douteux (maxime de qualité)
- le message est pertinent (maxime de relation)

Ces maximes permettent à l'auditeur de sélectionner une signification parmi toutes celles que le message aurait pu représenter. Par exemple, si l'on reprend (5) :

- (5) Robert a acheté le Figaro
- (5a) Robert a acheté un exemplaire du Figaro
- (5b) Robert a acheté l'entreprise qui publie le Figaro

seule (5a) sera possible dans certaines situations. Dans les cas où (5b) est manifestement impossible, (5b) est exclue de l'interprétation en raison de l'hypothèse qui empêche que le locuteur ait dit quelque chose qu'il tient pour faux (maxime de qualité).

Le problème du modèle "Gricéen" vient principalement de son manque de précision et de prédiction. Grice reconnaît que les maximes peuvent être violées, par ex. lorsque deux maximes entrent en conflit, ou lorsqu'un participant veut faire une insinuation (*implicature*). Mais la difficulté majeure réside dans le fait que les maximes ont une définition beaucoup trop floue pour que l'on puisse décider de leur éventuelle violation. Reprenons l'exemple (4) :

- (4a) Pierre : Veux-tu du café ?
- (4b) Marie : Le café m'empêche de dormir.

Au lieu de répondre par oui ou par non, Marie énonce un message à propos des vertus anti-dormitives du café. Le principe de coopération permet à Pierre de penser que Marie a voulu être pertinente, que ce qu'elle a dit est relié à une réponse donnée à (4a). Il peut aller chercher la connaissance "Marie souhaite dormir" pour inférer un refus du café.

Or cette maxime de la pertinence et le type d'inférences qu'elle autorise sont complètement laissés dans l'ombre par Grice. Au point que la théorie ne prend aucun risque, ce qui revient à dire qu'elle paraît irréfutable. Il semble qu'elle ne fait qu'entériner *a posteriori* ce que notre compétence d'interlocuteur nous dicte au vu de chaque exemple particulier. Dans (4ab), nous voyons (4b) comme un refus, nous comprenons que cela vient du fait que Marie veut dormir, et le principe de pertinence est sauf. S'il s'était agi d'une

<sup>4</sup> Voir note (3).

acceptation, en passant par le fait que Marie veut conduire et donc ne veut pas dormir, le principe de pertinence aurait aussi fonctionné. Si Marie avait dit "j'ai vécu 3 semaines chez ma tante dans le Jura", nous n'aurions pas compris, et nous aurions décidé que le principe de pertinence était en défaut<sup>(5)</sup>. Cette maxime de pertinence ne semble donc rien prédire. Elle semble simplement paraphraser notre compétence, ce qui la rend irréfutable. Il en est de même des autres maximes.

Grâce au principe de coopération, nous avons appris que l'auditeur sait qu'il pourra reconstituer la signification du message, un peu comme un enfant sait que l'énoncé d'un problème de mathématique contient juste les éléments qui permettent de le résoudre. Mais alors que nous connaissons en gros les techniques que l'enfant est susceptible d'employer pour parvenir à la solution de son problème scolaire, le principe de Grice ne donne aucune indication sur la méthode employée par l'auditeur pour faire les bonnes inférences et parvenir à la signification<sup>(6)</sup>. Or la question posée à la pragmatique était justement celle-là.

## 4. Sperber et Wilson : le principe de pertinence optimale

Dans leur célèbre ouvrage [1986], D. Sperber et D. Wilson font une critique de la pragmatique de Grice, et tout en reconnaissant l'apport de celui-ci, ils tentent de substituer un principe plus simple, plus général et sans doute mieux défini, le principe de pertinence, au principe de coopération et aux maximes qui l'accompagnent, moins parcimonieux selon eux.

"A bien des égards, l'analyse de Grice reste aussi floue que la représentation intuitive qu'on a de la communication. Plusieurs des concepts essentiels qui apparaissent dans les maximes ne sont pas définis. C'est le cas par ex. du concept de pertinence." (p.61)

Ces auteurs s'attachent en effet à rechercher les conditions qui permettent de dire qu'un message est pertinent. La solution qu'ils donnent à ce problème résout selon eux toutes les questions pour lesquelles les maximes de Grice et le phénomène d'implicature (insinuation) ont été postulées. Notons par commodité SW la théorie de ces auteurs. Selon SW, la pertinence d'un message est définie par rapport à un *contexte* qui dépasse les éléments objectifs de la situation :

Un message  $M$  émis par  $S$  en direction de  $D$  sera *pertinent* pour  $D$   
si et seulement si  
 $M$  a un *effet contextuel* sur la connaissance de  $D$ .

Voyons ce que ces auteurs entendent par effet contextuel :

(8) Vous êtes en train de dormir profondément

Une telle phrase n'est pas pertinente dans la plupart des cas, car vous ne pouvez pas la croire. Elle ne modifie pas votre connaissance. Dans SW, un message  $M$  n'est PAS pertinent dans l'un des trois cas suivants :

- les connaissances que l'on peut construire à partir de  $M$  sont sans lien avec le contexte
- les connaissances que l'on peut construire à partir de  $M$  sont déjà présentes dans le contexte
- les connaissances que l'on peut construire à partir de  $M$  sont en contradiction avec le contexte et la *force* du message (liée à la confiance que l'on a en lui) n'est pas suffisante pour que le contexte soit révisé (cas de (8)).

<sup>5</sup> Bien entendu, tout énoncé est interprétable par rapport à quelque contexte. Simplement, ici, la situation d'énonciation que nous sommes susceptibles d'imaginer en lisant (4a) ne rend pas acceptable la réponse de Marie concernant sa tante du Jura.

<sup>6</sup> Le principe de l'insinuation (*implicature*) engage l'auditeur à rechercher des inférences moins immédiates. Là encore, le mécanisme n'est pas élucidé par Grice.

Cette définition de la pertinence pose un problème. Elle suppose définie la notion de contexte. Or le contexte ne préexiste que partiellement au message. Il est issu, mais seulement en partie, de la prise en compte de la situation objective, et des déclarations antérieures. Autrement dit, une partie du contexte est construite "en temps réel" au moment de l'audition du message ! A partir du message, l'auditeur réalise des inférences et construit ainsi de nouvelles connaissances. Le contexte, après l'audition d'un message, comporte la connaissance que ce message permet de construire, mais aussi les conséquences logiques de cette connaissance.

Où s'arrête le contexte ? En effet, comme le note Johnson-Laird :

"There are always [...] infinitely many valid conclusions that can be drawn from any given set of premises. It is for this reason that I argued that an inferential heuristic, containing the particular conclusions that are drawn, is so crucial a part of a psychological theory of reasoning." [Johnson-Laird 1983]

Dans SW, il est clairement affirmé que le contexte n'a pas de frontières définies. Les connaissances évoquées par les messages successifs sont évoquées avec plus ou moins de *force* et plus ou moins d'*effort cognitif*. Lorsque la force devient trop faible ou le coût cognitif devient trop élevé, les inférences s'arrêtent.

La force d'une construction décroît chaque fois qu'une inférence utilise une hypothèse peu sûre. L'effort cognitif est une conséquence de la structuration de la mémoire : l'évocation d'une proposition *b* par une proposition *a* est plus facile si *a* et *b* ont été associées dans l'expérience de l'individu.

Maintenant, on est en droit de se demander ce qui va pousser l'auditeur à réaliser certaines inférences plutôt que d'autres ? Dans SW, la réponse vient du principe de pertinence lui-même, principe non appris, inhérent à toute situation de communication : *l'auditeur supposera toujours que le communicateur a émis le message le plus pertinent pour communiquer un ensemble de connaissances.*

Si le message *M* a été émis, c'est que l'auditeur doit avoir les moyens de relier *M* au contexte (pertinence de *M*) avec un coût raisonnable, et qu'il peut faire toutes les inférences de force suffisante accessibles pour un coût raisonnable à partir de *M*. Il obtiendra ainsi avec une bonne probabilité les connaissances que le communicateur voulait lui transmettre.

## 5. Comparaison et critique des théories de Grice et de Sperber et Wilson

L'intérêt, souligné par Sperber et Wilson, de la théorie SW par rapport à celle de Grice est qu'elle est plus parcimonieuse. Elle ne postule pas une bienveillance particulière et complexe du communicateur à l'égard de l'auditeur. Le communicateur ne peut pas empêcher que son message soit vu comme devant être optimalement pertinent (présomption de pertinence optimale), et il se conduit en conséquence.

Sur ce point, SW est tout de même assez proche du principe de coopération de Grice. Simplement, dans SW, les maximes apparaissent comme superflues, car elles découlent du fait que quoiqu'il arrive, le message est considéré comme candidat à la pertinence.

SW va plus loin que Grice sur la question de savoir comment l'interprétation est construite. Toutes les inférences accessibles pour un coût cognitif raisonnable à partir du message *M* seront faites, notamment celles qui permettront de relier *M* au contexte, à condition qu'elles n'accumulent pas des hypothèses trop douteuses.

Nous voudrions critiquer les aspects suivants de la théorie SW :

- la définition de la pertinence d'un message et la notion de lien message-contexte
- le principe d'un contexte graduel construit de manière "centrifuge"
- l'invocation d'un coût cognitif pour limiter la croissance du contexte
- la force des connaissances construites et son utilisation pour délimiter le contexte
- le caractère partiel de l'approche SW

**lien message-contexte** : selon SW, la pertinence d'un message suppose l'existence d'un "lien" entre le message et le contexte. Or la nature de ce lien n'est pas explicitée. C'est d'autant plus regrettable que c'est dans ce lien, selon nous, que réside l'entièreté de la pertinence. Nous verrons que certains liens sont licites, et d'autres non.

De plus, selon SW, un message *M* n'est PAS pertinent lorsque *M* est déjà dans le contexte, ou est en contradiction avec le contexte. Nous pensons que cela est erroné, comme le montrent les deux exemples qui suivent :

- (9) Tiens, tu es là toi !
- (10) C'est drôle, les tickets sont plus chers en banlieue.

L'intervention (9) est pertinente (dans certains contextes), et pourtant l'information qu'elle apporte est connue de son destinataire. Dans notre modèle, (9) est pertinente dans la mesure où elle exprime un sentiment d'incohérence. De même, et toujours contrairement aux prédictions de SW, (10) est pertinente pour quelqu'un qui est persuadé que les tickets ont un prix indépendant de leur lieu de vente, bien qu'elle entre en contradiction manifeste avec cette connaissance. Dans notre modèle, (10) est pertinente justement pour cette raison.

**contexte graduel** : l'idée d'un contexte graduel aux contours mal définis est caractéristique d'un modèle inférentiel, par opposition à un modèle du code qui permettrait de constituer sans ambiguïté un nouveau contexte à partir de l'ancien contexte et du message (cf. section 2.1.). Nous sommes résolument partisans d'un modèle digital qui repose sur un code, et nous considérons que la conception de la communication manifestée dans SW est, à cet égard, largement sous-estimée. Par ailleurs, le mode de construction du contexte, tel qu'il est suggéré dans SW, c.à.d. en "chaînage avant" à partir de la connaissance directement évoquée par le message, ne correspond pas à ce que suggèrent nos expérimentations. Dans notre modèle, le contexte est constitué en chaînage arrière, à partir de la relation problématique (incohérence, improbabilité, etc.) vers les hypothèses qui l'établissent.

**le coût des inférences et la force des connaissances construites** : pour des raisons d'ordre assez général, le fait d'expliquer des phénomènes fondamentaux comme la communication humaine à l'aide de contraintes floues (*i.e.* graduelles) comme la charge cognitive ou le degré de certitude semble fort peu satisfaisant. Pour nous, les phénomènes de communication présentent des aspects qualitatifs essentiels. Face à une remarque perçue comme non pertinente, les interlocuteurs protestent ("Pourquoi dis-tu ça ?"). Il s'agit là d'un phénomène tout-ou-rien. Les contraintes floues ne peuvent expliquer de telles réactions de rejet. De plus, l'argument est difficile à tenir pour d'autres raisons : par exemple, notre capacité à communiquer varierait sensiblement avec notre concentration, ce qui est loin d'être établi.

**un modèle partiel de la pertinence** : enfin, nous critiquerons la théorie SW parce qu'elle n'est pas testée sur des cas réels. On pourrait la qualifier de "pragmatique de fauteuil", par opposition avec des études qui s'efforcent de rendre compte de situations réelles. Au moment de rappeler l'importance du contexte, Sperber et Wilson avouent :

"Nous sommes obligés d'utiliser des exemples artificiels [...]. Cela ne veut pas dire qu'ils sont traités et compris hors de *tout* contexte." (p. 275)

Une telle démarche semble admise dans certaines sciences du langage. Mais que penserait-on d'une science physique qui se contenterait d'interpréter les résultats imaginaires d'expériences imaginaires ?

Grice et SW emploient souvent le mot "conversation", qui se retrouve d'ailleurs dans le titre de l'article de Grice. Pourtant, à aucun moment, ces auteurs ne se posent la question fondamentale liée à toute situation de communication bidirectionnelle : comment le message  $M'$ , qui répond à  $M$ , est-il calculé ? Nous allons prolonger cet aspect de notre critique en montrant les insuffisances des deux modèles pragmatiques que nous avons évoqués, lorsqu'il s'agit de rendre compte d'interactions réelles.

## 6. Vers une pragmatique conversationnelle

Si l'on considère les prédictions du principe de coopération de Grice ou du principe de pertinence de SW, et qu'on cherche à les appliquer à une situation de *conversation*, on s'aperçoit qu'elles sous-déterminent largement les phénomènes de dialogue. Dans l'extrait suivant, B *sait* que quelque chose "ne va pas" dans l'intervention de A :

- (11) A1- T'as vu la Mercedes ?  
B1- Hé bé quoi ?

Pourtant une voiture de cette marque venait effectivement de doubler le véhicule de A et B, sur l'autoroute qui va de Paris vers l'Allemagne. Ce type de voiture se rencontre fréquemment sur cet itinéraire.

B pouvait inférer beaucoup de choses de façon sûre et sans effort : la Mercedes allait plus vite que leur véhicule; elle était nettement plus luxueuse (inférences immédiates); B devrait économiser pour pouvoir acheter ce genre de voiture; ils seraient sûrs d'arriver à temps; il serait possible de rouler plus vite avec la même sécurité (insinuations); etc. La réplique A1 ne viole aucune maxime de Grice : la mention de la Mercedes suffisait pour que B puisse faire des inférences, il était inutile de mentionner sa couleur ou sa vitesse ou sa longueur (maxime de quantité), l'événement mentionné était vrai (maxime de qualité), et il était relié au contexte de la circulation sur autoroute (maxime de relation, principe de pertinence). De même, par rapport à SW, la réplique A1 est pertinente. Elle a un effet sur le contexte, puisque B peut faire des inférences, et B peut considérer que A a donné un message optimal pour que ces inférences-là soient construites.

Pourtant A1 telle quelle est inadmissible, et B émet une réaction montrant qu'il n'a pas compris sa pertinence. Et effectivement la pertinence de A1 vient d'ailleurs, et B *savait* qu'il lui manquait quelque chose. Ce quelque chose est donné dans A2 :

- (12) A2- Tu savais que les hongrois, ils avaient des voitures comme ça ?

En effet, la Mercedes était immatriculée en Hongrie, ce que B n'avait pas remarqué. A2 fait ainsi mention d'un événement qui entre en contradiction avec la connaissance de B. A cette époque, le régime politique en Hongrie rendait impossible le fait qu'un ressortissant hongrois pût posséder officiellement une voiture aussi chère. B a toutes les raisons de penser que ce fait est impossible, et il n'est pas disposé à changer son point de vue. Pourtant, contrairement à ce que prédit SW, A2 lui apparaît tout à fait pertinente. Elle est même pertinente pour lui du fait même qu'elle entre en contradiction avec sa propre connaissance !

Dans notre modèle (voir chapitre I.1), nous avons défini les conditions de la pertinence logique des répliques. Pour la première intervention, il s'agit d'apparaître comme suggérant une situation *incohérente*, très *improbable*, ou très *désirable* / *indésirable* (chapitre I.1, Premier principe de pertinence logique). C'était le cas de (11) A1+ (12) A2, qui mettait en évidence une situation incohérente. Pour comprendre par ex. (5) en tant qu'intervention initiale pertinente, il nous faut absolument un contexte qui rende l'achat d'un exemplaire du Figaro *incohérent* (par ex. si l'acheteur est connu pour ses opinions d'extrême gauche et que l'on considère comme admis qu'il s'agit d'un journal de droite), ou un contexte qui rende l'achat de l'entreprise possible mais *improbable* (si Robert, qui possède d'autres entreprises de presse, vient d'acquérir "Le Figaro" aujourd'hui), ou un contexte qui rende l'achat *désirable* (nous allons enfin avoir les nouvelles), etc.

Noter que le contexte est alors défini sans ambiguïté. Il comprend tous les faits qui entrent dans la relation d'incompatibilité (ou d'improbabilité ou d'(in)désirabilité), cette relation elle-même, plus tous les faits et relations qui établissent ces faits à partir de connaissances tenues pour vraies. Ce contexte, contrairement à ce que laisse entendre SW, est recruté selon un "chaînage arrière", allant des faits à établir vers les prémisses qui entraînent logiquement ces faits.

Le critère de pertinence fourni par notre modèle s'applique aussi aux répliques ultérieures. Une telle réplique doit obligatoirement être une invalidation logique, une banalisation, une réaction antagoniste, ou une co-réaction (voir chap. I.1, Deuxième principe de pertinence logique). Ceci élucide, en la limitant, la nature du lien entre une réplique et le contexte. Contrairement à ce que prédit SW, il ne suffit pas d'être simplement relié au contexte pour être pertinent. Comment A aurait-elle réagi dans (11-12) si B avait répondu B2a ou B2b après A2 ?

- (13) B2a : les Mercedes sont solides  
B2b : les Hongrois fabriquent des voitures

Ces phrases sont parfaitement reliées au contexte, elles permettent de faire davantage d'inférences, et pourtant elles sont inadmissibles. La seule façon d'être pertinent après (12)A2, ainsi que le montre notre modèle, est soit de se co-étonner ("Tiens, oui, c'est bizarre !"), soit d'invalider l'incohérence ("Il existe des Hongrois privilégiés", "Il y avait un H mais ce n'est pas une plaque d'immatriculation hongroise", etc.).

## 7. Conclusion

Selon notre vision des modèles de Grice et de Sperber et Wilson, ces auteurs édifient des contraintes *autour* du phénomène de pertinence. Du moment qu'un message peut être pertinent ou au contraire manquer de l'être, le communicateur se doit d'aider le destinataire à percevoir la pertinence de son message, soit parce qu'il décide de coopérer (vision gricéenne), soit parce que de toute façon son message sera reçu comme s'il s'était conduit de manière optimalement coopérante (vision SW). Mais cela ne définit pas la pertinence.

La tentative, dans SW, de définir la pertinence par l'effet contextuel est malheureusement un peu courte. L'effet contextuel est imparfaitement défini : quelle est la nature du "lien" invoqué dans SW ? Cette insuffisance est, selon nous, regrettable, car c'est véritablement elle qui peut fonder le concept de pertinence.

Nous pensons que notre modèle est propre à amener ce qui fait tant défaut à ces théories pragmatiques. Si, comme nous le croyons, notre modèle est en adéquation avec ce que l'on peut expérimentalement observer, alors il définit de manière stricte les conditions de la pertinence. Il permet aussi de définir sans ambiguïté le contexte auquel un message doit sa pertinence. Enfin, notre modèle va plus loin : il prend en compte l'aspect symétrique de la communication humaine, en ce qu'il permet de prédire les répliques pertinentes qu'un message peut provoquer.

**remerciements** : je remercie Gilles Ricour pour son aide dans la mise en parallèle des travaux de Grice avec ceux de Sperber et Wilson. Je remercie aussi Laurent Karsenty et Christian Brassac pour leurs remarques... pertinentes.



# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

On considère généralement que les modèles de l'argumentation ne sont pas assez prédictifs pour conduire à des implémentations informatiques. Le traitement automatique du langage naturel inclut donc rarement les aspects supérieurs du langage. Nous tentons de montrer ici qu'il est possible de synthétiser une suite d'arguments à partir d'une connaissance logique statique. Notre programme PARADISE est capable de reconstruire l'argumentation échangée au cours d'une conversation réelle. Des résultats comme ceux que nous avons obtenus en développant PARADISE nous permettent d'espérer que la compétence conversationnelle des interlocuteurs peut être décrite à l'aide d'un ensemble limité de stratégies simples.

## 2. Analyse et synthèse des arguments conversationnels

Plusieurs aspects du langage naturel ont été modélisés de manière suffisamment précise pour permettre à l'ordinateur de reproduire certains aspects de la performance humaine. Mais cela a rarement été tenté pour les niveaux "supérieurs" du langage comme celui qui a trait à l'argumentation. Certes, reproduire la façon qu'ont les interlocuteurs d'argumenter au cours d'une conversation spontanée peut sembler soit impossible, soit trivial. Cela semble impossible lorsqu'on constate que les interlocuteurs font appel aux connaissances "de bon sens", qui font manifestement défaut aux machines. D'un autre côté, la reproduction de l'argumentation peut apparaître comme triviale si l'on part d'un modèle descriptif d'une conversation particulière qui inclut des connaissances *ad hoc* comme la force d'un argument donné dans un contexte donné.

Pourtant le problème vaut la peine d'être étudié. Une analyse correcte des arguments peut nous apporter une meilleure compréhension de l'usage spontané du langage naturel. De plus, si nous sommes en mesure d'engendrer des arguments qui seront perçus comme pertinents, nous pouvons nous attendre à une amélioration significative de l'interaction humain-machine dans les situations où une négociation des connaissances est en jeu : Enseignement Assisté par Ordinateur, explication dans les Systèmes à Base de Connaissances, acquisition de connaissances, etc.

Un certain nombre de travaux ont été consacrés à l'analyse du discours et des conversations (voir [Goodwin & Heritage 1990] pour une revue synthétique), mais la plupart des modèles ne sont pas suffisamment prédictifs pour permettre une implantation sur ordinateur. Certains auteurs se préoccupent des aspects linguistiques de l'argumentation (*e.g.* les connecteurs comme "mais" [Anscombe & Ducrot 1988]), d'autres s'intéressent à l'interaction verbale et regardent les participants comme des acteurs (coopération, négociation [Baker 1991], actes de langage [Searle 1969]), d'autres encore étudient les aspects structurels de l'argumentation ([Heritage 1990], [Tannen 1984], [Reichman 1985], [Trognon 1990]). Mais très peu d'études

concernent la question spécifique de prédire le contenu de ce qui va être dit à un moment donné d'une conversation.

Pour rendre possible ce type de prédiction, les auteurs qui cherchent à concevoir des modèles computationnels de l'argumentation doivent, dans un premier temps, examiner le contenu des arguments produits en situation naturelle et analyser leur effet logique sur les déclarations qui les précèdent [Guez 1990], [Flowers et al. 1982]. C'est ce que nous avons fait [Dessalles 1985]. Ceci nous a conduits à constater que plusieurs contraintes de type logique limitent la liberté des interlocuteurs (chap. I.1). Dans de nombreuses situations concrètes, ces contraintes laissent très peu de possibilités pour choisir un argument pertinent au sein de la connaissance disponible, et ces limitations rendent la prédiction possible, ainsi que nous allons le montrer en donnant l'exemple d'une reconstruction de conversation.

### 3. Le problème de la reconstruction d'une conversation

La reconstitution des arguments qui ont été réellement émis par des interlocuteurs humains au cours d'une interaction spontanée est sans aucun doute le meilleur moyen de mettre à l'épreuve la précision d'un modèle qui décrit et prédit le contenu des répliques. Nous allons illustrer ceci en donnant un extrait et en montrant en détail comment il peut être reconstruit.

[ex\_portes]

contexte: A raconte qu'il est en train de repeindre les portes chez lui, et qu'il a décidé de commencer par décaper la vieille peinture, ce qui se révèle être un travail difficile.

A1- *Ben moi, j'en bave actuellement parce qu'il faut que je refasse mes portes, la peinture. Alors j'ai décapé à la chaleur. Ça part bien. Mais pas partout. C'est un travail dingue, hein ?*

[...]

B2- *Quelle chaleur ? La lampe à souder ?*

A3- *Ouais, avec un truc spécial.*

B3- *Faut une brosse, dure, une brosse métallique.*

A4- *Oui, mais j'attaque le bois.*

B4- *T'attaques le bois.*

[pause 5 secondes]

A5- *Enfin je sais pas. C'est un boulot dingue, hein? C'est plus de boulot que de racheter une porte, hein ?*

B5- *Oh, c'est pour ça qu'il vaut mieux laiss... il vaut mieux simplement poncer, repeindre par dessus*

A6- *Ben oui, mais si on est les quinzièmes à se dire ça*

B6- *Ah oui.*

A7- *Y a déjà trois couches de peinture, hein, dessus.*

B7- *Remarque, si elle tient bien, la peinture, là où elle est écaillée, on peut enduire. De l'enduit à l'eau, ou*

A8- *Oui, mais l'état de surface est pas joli, quoi, ça fait laque, tu sais, ça fait vieille porte.*

Il est strictement impossible d'effectuer l'analyse et la reconstruction d'un tel extrait sans une connaissance précise du contexte. Même des observateurs extérieurs humains peuvent avoir les plus grandes difficultés à saisir la pertinence de certaines répliques s'ils ne disposent pas des éléments contextuels qui, parfois, sont très spécifiques. La question est donc : en quoi consiste la connaissance contextuelle et d'où vient-elle ?

Dans le cadre que nous nous fixons ici, la connaissance contextuelle va consister en un ensemble minimum de relations logiques entre des faits, qui permettent de représenter logiquement la signification de chaque réplique. Selon notre modèle, les participants d'une conversation utilisent ce type de connaissance pour analyser et concevoir les arguments. Nous allons illustrer ceci en montrant en détail comment notre programme PARADISE réalise ce processus.

Mais d'où vient cette connaissance logique sur laquelle le programme va travailler ?

Dans certaines situations favorables, il est possible d'utiliser une base de connaissances existante et le problème se ramène alors à une présélection d'un sous-ensemble de règles pragmatiquement pertinentes (en utilisant des principes du genre de ceux que proposent Sperber & Wilson [1986]). Ce peut être le cas lorsqu'il s'agit de traiter des dialogues didactiques ou techniques. Mais dans le cas des dialogues de la vie courante comme celui que nous venons de reproduire, nous ne pouvons pas anticiper quelle sera la connaissance utile. Nous devons la fournir "à la main". Mais qu'est-ce qui garantira son caractère objectif ? Essayons de clarifier ce point en donnant la connaissance contextuelle relative à l'extrait précédent, dans un premier temps sous une forme textuelle.

*contexte de A1: Repeindre est un travail fastidieux. Cela est indésirable. Si l'on décape les portes, les portes ont une jolie surface, ce qui est désirable.*

*contexte de B2: Si l'on décape les portes sans utiliser un chalumeau (par ex. avec un dispositif à air chaud), alors le décapage est fastidieux [chalumeau nécessaire].*

*contexte de B3: Si l'on ne veut pas que le décapage soit un travail fastidieux, alors on doit utiliser une brosse métallique [brosse métallique nécessaire].*  
*Si l'on utilise un chalumeau et une brosse métallique, alors le décapage n'est pas fastidieux [chalumeau + brosse métallique suffisants].*

*contexte de A4: le bois est tendre, donc une brosse métallique abîme le bois, ce qui est indésirable.*

*contexte de B5: Pour obtenir une surface correcte, il faut décapier ou poncer les portes [décapage ou ponçage nécessaires]*  
*si les portes sont poncées, alors la surface est correcte, ce qui est désirable [ponçage suffisant]*

*contexte de A7: Si l'on ne décape pas les portes, et s'il y a plusieurs couches, alors la surface ne sera pas correcte [décapage nécessaire].*

*contexte de B7: Si l'on ne décape pas les portes, s'il y a plusieurs couches et si l'on n'applique pas de l'enduit, alors la surface ne sera pas correcte [décapage ou enduit nécessaires].*  
*Si l'on ponce les portes, s'il y a plusieurs couches, si la vieille peinture tient bien, et si l'on applique de l'enduit, alors la surface sera correcte [enduit suffisant].*

*contexte de A8: Si l'on ne décape pas les portes, s'il y a plusieurs couches, alors la surface ne sera pas jolie [décapage nécessaire].*

Normalement, le contexte évolue au fur et à mesure du déroulement de la conversation, et donc chaque réplique doit être analysée par référence à un contexte légèrement différent. La liste ci-dessus doit donc être lue comme une liste cumulative, avec des règles ou des prémisses qui s'ajoutent à chaque étape. Nous allons toutefois constater que l'évolution de ce contexte peut être *calculée* à partir d'un ensemble non ordonné de connaissances dont disposent les interlocuteurs sur le sujet de la conversation.

L'essentiel de cette connaissance est fourni par la situation d'énonciation, qui est connue avec précision (il serait presque impossible de travailler sur des extraits pour lesquels les faits contextuels seraient inconnus). Les détails de cette connaissance (forme logique, mention de telle ou telle prémisses) sont le résultat d'une "traduction". Celle-ci a la même valeur qu'une traduction entre langues naturelles : elle peut être rendue aussi précise que nécessaire, et sa validité résulte d'un accord sur sa capacité à représenter le sens. Ici, par exemple, la distinction entre "joli" et "correct" a été introduite pour exprimer la signification de

A8. D'autres paraphrases logiques auraient été possibles. Elles auraient conduit, selon nous, à des reconstructions similaires.

#### 4. De la connaissance statique à l'argumentation

Nous pouvons exprimer la connaissance contextuelle sous forme logique, en employant des clauses négatives :

$$[ p_1 \& p_2 \& \dots \& p_n ] \Rightarrow \mathbf{F}$$

$\mathbf{F}$  tient lieu de proposition fausse. Chaque  $p_i$  est une proposition ou un prédicat, éventuellement nié. Une telle clause signifie que  $p_1, p_2 \dots p_n$  ne peuvent pas être tous vrais simultanément. En d'autres termes ils sont incompatibles. Les éventualités indésirables sont, quant à elles, exprimées à l'aide de la modalité  $\mathbf{IND}$  :

$$[ p_1 \& p_2 \& \dots \& p_n ] \Rightarrow \mathbf{IND}$$

Cette clause signifie que  $p_1, \dots p_n$  ensemble sont indésirables. Nous donnons maintenant une représentation de la connaissance contextuelle de l'extrait précédent en utilisant ces clauses négatives.

##### connaissance commune

- \*1. *repeindre-fastidieux*  $\Rightarrow \mathbf{IND}$
- \*2. *bois-abîmé*  $\Rightarrow \mathbf{IND}$
- \*3. **non** *surface-correcte*  $\Rightarrow \mathbf{IND}$
- 4. [*décapage & non surface-jolie*]  $\Rightarrow \mathbf{F}$
- 5. [*surface-correcte & non décapage & non ponçage*]  $\Rightarrow \mathbf{F}$
- 6. [*surface-correcte & non décapage & plusieurs-couches & non enduit*]  $\Rightarrow \mathbf{F}$
- 7. [*peinture-tient-bien & ponçage & enduit & non surface-correcte*]  $\Rightarrow \mathbf{F}$
- 8. [*ponçage & non plusieurs-couches & non surface-correcte*]  $\Rightarrow \mathbf{F}$
- 9. [*décapage & non chalumeau & non repeindre-fastidieux*]  $\Rightarrow \mathbf{F}$
- 10. [*décapage & non brosse-métallique & non repeindre-fastidieux*]  $\Rightarrow \mathbf{F}$
- 11. [*décapage & chalumeau & brosse-métallique & repeindre-fastidieux*]  $\Rightarrow \mathbf{F}$
- 12. [*brosse-métallique & bois-tendre & non bois-abîmé*]  $\Rightarrow \mathbf{F}$
- \*13. [*surface-jolie & non surface-correcte*]  $\Rightarrow \mathbf{F}$
- 14. [**non** *décapage & plusieurs-couches & surface-jolie*]  $\Rightarrow \mathbf{F}$

##### connaissance spécifique de A

- 15. **non** *bois-tendre*  $\Rightarrow \mathbf{F}$
- 16. **non** *plusieurs-couches*  $\Rightarrow \mathbf{F}$
- 17. **non** *peinture-tient-bien*  $\Rightarrow \mathbf{F}$
- 18. **non** *surface-jolie*  $\Rightarrow \mathbf{IND}$

Le point fondamental est que cette connaissance est totalement déclarative : il n'y a pas d'ordre entre les clauses, et il n'y a pas d'ordre non plus entre les propositions des clauses (plus précisément, le programme doit donner le même résultat quel que soit l'ordre). Tout le travail de PARADISE est donc de choisir le bon argument au bon moment. Nous pouvons faire trois remarques pour estimer la difficulté de la tâche.

**Calculer les effets des arguments :** tout d'abord, le type de connaissance utilisé par PARADISE ne donne aucune indication directe sur les raisons de préférer tel ou tel argument. Dans [Guez 1990] par exemple, nous trouvons des règles comme celle-ci :

stronger-opp(lot-of-work, nice-weather, opposite(go-for-a-walk))

ce qui signifie que *lot-of-work* est un argument plus fort que *nice-weather* pour prendre une décision à propos de *go-for-a-walk*.

Nous mettons PARADISE dans une situation bien plus difficile : il doit concevoir une argumentation, alors que le rôle potentiel de certaines propositions comme arguments, ainsi que la force relative qui peut être la leur, ne sont pas donnés en tant que tels, mais doivent être *calculés*.

**Une connaissance préexistante :** une deuxième remarque concerne l'évolution de la connaissance au cours de la conversation. Nous considérons (et d'ailleurs PARADISE contribue à le prouver) que la connaissance n'est pas apprise à la suite de chaque réplique entendue. La majeure partie de la connaissance pragmatiquement pertinente est partagée par les interlocuteurs, ou peut être activée (mais seulement activée) dans leur esprit. Dans notre description de l'extrait précédent, seulement quatre faits ont été instanciés dans l'esprit de B à la fin de l'interaction (clauses 15 à 18). Toutes les autres connaissances logiques peuvent être considérées comme connues des deux participants, *i.e.* ils peuvent activer et utiliser de telles connaissances si nécessaire. C'est pourquoi la connaissance est donnée à PARADISE comme un ensemble statique de règles ordonnées arbitrairement. Nous ne cherchons pas à biaiser le choix des arguments avec une progression qui serait cachée dans l'ordre des règles.

**Trouver les bons arguments, dans le bon ordre :** notre troisième remarque concerne le fossé qui existe entre une connaissance statique non ordonnée et l'aspect dynamique de la conversation. Pour donner une idée de la difficulté de choisir des arguments pertinents, nous comparons trois dispositifs capables de donner les cinq arguments qui ont été produits par B dans l'extrait précédent : *chalumeau*, *brosse-métallique*, *ponçage*, *enduit* et *[absence d'argument]* (= arrêt de la conversation).

Le tableau suivant indique les probabilités de trouver ces arguments dans l'ordre.

<i>dispositif</i>	<i>probabilité</i>
choix aléatoire : tirage sans remise d'une proposition parmi les 12 propositions du contexte, avec une négation possible	$0.3 \times 10^{-6}$
choix d'un argument dans une clause contenant une proposition mentionnée par l'interlocuteur dans sa réplique, avec une négation possible	$0.5 \times 10^{-4}$
PARADISE	0.5

Comme nous allons le voir, PARADISE évoque le chalumeau avant de parler de brosse-métallique, mais cela aurait pu être l'inverse. Tous les autres arguments arrivent nécessairement à leur tour, ce qui explique la probabilité 0.5 donnée pour PARADISE. Nous allons maintenant décrire le comportement de PARADISE .

## 5. Fonctionnement de la reconstruction

Voici ce que PARADISE produit dans le rôle de B (nous commenterons la présence de B3' et de A4' plus loin) :

<i>conversation d'origine</i>	<i>reconstitution par PARADISE</i>
<i>A1- Ben moi, j'en bave actuellement parce qu'il faut que je refasse mes portes, la peinture. Alors j'ai décapé à la chaleur. Ca part bien. Mais pas partout. C'est un travail dingue, hein?</i>	A1 <ul style="list-style-type: none"> <li>• on décape les portes</li> <li>• repeindre constitue un travail dingue</li> </ul>
<i>B2- Quelle chaleur? La lampe à souder?</i>	B2 <ul style="list-style-type: none"> <li>• On peut éviter que repeindre constitue un travail dingue; il n'y a qu'à considérer que on utilise un chalumeau</li> </ul>
<i>A3- Ouais, avec un truc spécial.</i>	A3 <ul style="list-style-type: none"> <li>• on utilise un chalumeau</li> </ul>
[...]	
<i>B3- Faut une brosse, dure, une brosse métallique.</i>	B3 <ul style="list-style-type: none"> <li>• On peut éviter que repeindre constitue un travail dingue; il n'y a qu'à considérer que on utilise une brosse métallique</li> </ul>
<i>A4- Oui, mais j'attaque le bois.</i>	A4 <ul style="list-style-type: none"> <li>• on attaque le bois</li> </ul>
<i>B4- T'attaques le bois.</i>	B3' <ul style="list-style-type: none"> <li>• Mais il est possible que le bois des portes n'est pas tendre</li> </ul>
[pause 5 secondes]	A4' <ul style="list-style-type: none"> <li>• le bois des portes est tendre</li> </ul>
<i>A5- Enfin je sais pas. C'est un boulot dingue, hein? C'est plus de boulot que de racheter une porte, hein?</i>	B4 <ul style="list-style-type: none"> <li>• Certes... Si le bois des portes est tendre alors on n'utilise pas une brosse métallique sinon on attaque le bois</li> </ul>
<i>B5- Oh, c'est pour ça qu'il vaut mieux laiss... il vaut mieux simplement poncer, repeindre par dessus</i>	B5 <ul style="list-style-type: none"> <li>• On peut éviter que on décape les portes et que repeindre constitue un travail dingue; il n'y a qu'à considérer que on ponce les portes</li> </ul>
<i>A6- Ben oui, mais si on est les quinzièmes à se dire ça</i>	
<i>B6- Ah oui.</i>	
<i>A7- Y a déjà trois couches de peinture, hein, dessus.</i>	A7 <ul style="list-style-type: none"> <li>• il y a plusieurs couches superposées</li> </ul>
<i>B7- Remarque, si elle tient bien, la peinture, là où elle est écaillée, on peut enduire. De l'enduit à l'eau, ou</i>	B7 <ul style="list-style-type: none"> <li>• Mais il est possible que on passe de l'enduit</li> </ul>
<i>A8- Oui, mais l'état de surface est pas joli, quoi, ça fait laque, tu sais, ça fait vieille porte.</i>	A8 <ul style="list-style-type: none"> <li>• l'état de surface n'est pas joli</li> <li>• C'est vrai, tu as raison... on décape les portes sinon l'état de surface n'est pas joli puisque il y a plusieurs couches superposées</li> </ul>

Nous fournissons en annexe la reconstruction lorsque PARADISE joue le rôle de A, ainsi que la reconstruction d'un autre dialogue.

PARADISE se compose de deux modules principaux : un moteur de clauses négatives et un module conversationnel :

- Le moteur réalise des déductions selon une méthode de réfutation : pour prouver  $p$ , il montre que **non**  $p$  est faux. Pour cela, il recherche une clause négative contenant **non**  $p$  et tente récursivement de prouver les autres termes de cette clause, s'ils existent. Noter que ce moteur traite les clauses indépendamment de leur modalité (voir annexe du chapitre III.2).
- C'est le module conversationnel qui va retenir plus particulièrement notre attention ici. Son fonctionnement est organisé suivant un principe fondamental et assez bien reconnu : les conversations obéissent à une structure de pile. Dans notre extrait, la pile peut être décrite de la manière suivante.

<i>pile conversationnelle</i>	<i>pile des clauses de référence (du point de vue de B)</i>
A1: travail fastidieux	[ 1 ]
B2: chalumeau	[ 1-9 ]
A3: (acceptation)	[ 1-9 ]
B3: brosse-métallique	[ 1-10 ]
A4: bois-abîmé	[ 1-10-12 ]
B3': <b>non</b> bois-tendre	[ 1-10-12 ]
A4': bois-tendre	[ 1-10-12-15 ]
B4:	[ 1-10-12-15 ]
(acceptation)	
A5: travail fastidieux	[ 1 ]
B5: poncer	[ 1-10-5 ]
A6: plusieurs-couches	[ 1-10-5-6 ]
B6: (acceptation)	[ 1-10-5-6 ]
A7: plusieurs-couches	[ 1-10-5-6 ]
B7: enduit	[ 1-10-5-6 ]
A8: <b>non</b> surface-jolie	[ 1-14 ]

Les nombres de la colonne de droite font référence aux clauses listées plus haut. La clause du sommet de la pile est la dernière clause qui est utilisée du point de vue du programme : cette clause peut soit être saturée (tous ses termes sont vrais) ou bien être utilisée par un interlocuteur pour prouver un terme. Par exemple, la clause :

$$[ p_1 \ \& \ p_2 \ \& \ \dots \ \& \ p_n ] \Rightarrow \textit{mod}$$

devient la clause de référence courante...

1) si tous ses termes se retrouvent vrais (ce qui traduit un étonnement si  $\textit{mod} = F$ , ou une indésirabilité si  $\textit{mod} = IND$ ),

2) si la clause est utilisée par l'un des interlocuteurs pour établir un terme  $q$ . Par ex. si  $q = \textit{non } p_i$ , la clause précédente se réécrit :

$$[ p_1 \ \& \ \dots \ \& \ p_{i-1} \ \& \ p_{i+1} \ \& \ \dots \ \& \ p_n ] \Rightarrow q$$

Dans ce type de réécriture, la modalité n'intervient pas (plus précisément,  $\textit{mod}$  ou  $q$  est considéré comme égal à  $q$ , ce qui correspond bien à la sémantique attendue).

Voyons comment PARADISE utilise sa pile pour calculer les arguments. Au début de l'interaction, PARADISE, qui joue ici le rôle de B, analyse A1 comme rendant la clause 1 saturée (introduction sur le mode indésirable). La clause 1 devient ainsi la première clause de référence. PARADISE estime ensuite que la clause 9 a été utilisée pour prouver *travail-fastidieux*. La clause 9 est "empilée" comme clause de référence. PARADISE suggère l'emploi d'un chalumeau, car c'est pour lui un moyen d'invalidiser cette nouvelle clause.

La suggestion étant acceptée, PARADISE dépile la clause 9 et réexamine la situation. Mais la clause 1 reste saturée, car son unique terme peut résulter d'une autre clause, la clause 10, qui va être empilée à son tour. PARADISE fait donc une deuxième tentative : il tente d'invalider la clause 10 en suggérant l'emploi d'une brosse-métallique. Puis vient A4 (bois abîmé) que le programme analyse comme une invalidation de sa dernière suggestion à travers la clause 12.

Après A4' (bois tendre), *non brosse-métallique* est établi de manière définitive, et ainsi *travail-fastidieux* semble désespérément vrai. A ce niveau, le programme a compris non seulement le problème (indésirable) qui motive A1, mais il dispose aussi de la preuve des éléments de ce problème.

PARADISE va tout de même tenter encore de "casser" cette situation problématique. Pour cela, il prend la décision de remettre en question un fait appris de l'interlocuteur. Il regarde comment *décapage* a pu être déduit. Il suppose que l'interlocuteur a utilisé pour cela la clause 5, qui se réécrit de la manière suivante :

[surface-correcte & non ponçage]  $\Rightarrow$  décapage

La clause 5 est à son tour empilée, et la suggestion B5 (*ponçage*) est ainsi calculée comme une invalidation de cette clause 5. La présence de la clause 5 en sommet de pile permet d'analyser la réplique A7 (*plusieurs-couches*) comme une défense du terme mis en doute (*décapage*), grâce à la clause 6. PARADISE empile 6 avant de l'invalider avec B7 (*enduit*). A8 est aussi analysée comme une défense de *décapage* par la clause 14.

On le voit, la structure de pile est essentielle. C'est elle qui permet à PARADISE d'analyser chaque réplique à travers son effet sur la clause de référence, c.à.d. la dernière clause qui a été utilisée par le programme ou l'interlocuteur lui-même. Nous allons indiquer maintenant ce qui permet à PARADISE de constituer cette pile, en montrant comment il détermine quelle clause sera la prochaine clause de référence.

## 6. Les stratégies de PARADISE

PARADISE ne reconnaît et n'utilise que peu de stratégies pour former des répliques pertinentes. C'est en cela que réside son intérêt en tant que modèle. Si  $CR_k$  désigne la dernière clause de référence (i.e  $CR_k$  est la dernière clause saturée ou utilisée pour prouver un fait), la technique de base pour invalider consiste à nier un terme de  $CR_k$ . Lorsque ce n'est pas possible, PARADISE fait intervenir une nouvelle clause  $C$  pour former une réplique  $R$  qui invalidera  $CR_k$ . Le tableau suivant décrit les trois stratégies d'invalidation :

	stratégie pour former l'invalidation $R$	repr. symbolique	exemple
①	[ <i>non R</i> ] est présente dans $CR_k$	$R \text{ } \text{✋} \text{ } CR_k$	B3', B7
②	$R$ est présente dans la clause $C$ . $R'$ peut être prouvée grâce à $C$ . [ <i>non R'</i> ] est présente dans $CR_k$ . $C$ est empilée ( $C = CR_{k+1}$ ).	$R \rightarrow C \rightarrow R' \text{ } \text{✋} \text{ } CR_k$	A4, A7
③	[ <i>non R</i> ] est présente dans la clause $C$ . $R'$ peut être prouvée grâce à $C$ , et $R'$ est présente dans $CR_k$ . $C$ est empilée ( $C = CR_{k+1}$ ).	$R \text{ } \text{✋} \text{ } C \rightarrow R' \rightarrow CR_k$	B2, B3, B5

Ce sont les moyens de base utilisés ou reconnus par PARADISE pour former et analyser les invalidations. Toutefois, si  $CR_k$  est utilisé pour prouver un terme  $q$  :

[  $p_1$  & ... &  $p_{i-1}$  &  $p_{i+1}$  & ... &  $p_n$  ]  $\Rightarrow$   $q$

(ici  $q = \text{non } p_i$ ), alors les trois stratégies peuvent être utilisées pour nier  $q$  (i.e prouver  $p_i$ ) au lieu de nier l'un des  $p_j$ . Les trois stratégies deviennent symboliquement :

❶'	$R \text{ ✎ } q$
❷'	$R \rightarrow C \rightarrow R' \text{ ✎ } q$
❸'	$R \text{ ✎ } C \rightarrow R' \rightarrow q$

Nous devons maintenant expliquer brièvement comment ces stratégies sont utilisées par PARADISE .

### 6.1. Analyse des arguments de l'interlocuteur

L'analyse d'une réplique  $R$  fournie par l'interlocuteur est systématique et est dirigée par les données : PARADISE recherche une clause contenant  $R$ , puis il cherche le lien entre cette clause et la clause de référence  $CR_k$  selon l'une des trois stratégies évoquées plus haut. C'est tout pour le principe de base.

Il y a une exception à ce processus, et elle est particulièrement intéressante. Elle concerne l'analyse des interventions initiales, l'analyse des répliques antagonistes et celle des raisonnements par l'absurde. Dans ces cas-là, lorsque  $R$  est émis, c'est [*non*  $R$ ] qu'il faut entendre. Pour détecter ces situations, PARADISE dispose d'une information (un drapeau) indiquant pour chaque clause s'il s'agit d'une connaissance "évidente" (de bon sens) ou non. Ce drapeau est indiqué par une étoile dans la connaissance donnée page 70. Supposons que la réplique  $R$  à analyser soit trouvée dans une clause de bon sens  $C$  :

$$\begin{array}{c}
 * \text{ clause } C : \quad [ p_1 \ \& \ p_2 \ \& \ \dots \ \& \ R ] \Rightarrow \textit{mod} \\
 \leftarrow \hspace{1.5cm} \rightarrow \\
 \textit{termes vrais}
 \end{array}$$

Si tous les termes de  $C$  autres que  $R$  sont vrais, alors  $C$  prouve *non*  $R$ .  $R$  se retrouve ainsi fausse, trivialement fausse, puisque  $C$  est une clause "évidente". PARADISE considère que l'interlocuteur n'a pas pu commettre une telle "erreur". Si  $R$  est la première intervention, alors il s'agit bien sûr d'une introduction sur le mode paradoxal ou indésirable, selon la valeur de *mod* (cf. chap. I.1). Sinon,  $R$  est analysée comme une réplique "par l'absurde" ou comme une réplique antagoniste, toujours selon la valeur de *mod*. Cela revient à considérer que *non*  $R$  a été émise.

Par exemple, *travail-fastidieux* dans A1 est trivialement fausse par la clause 1. Toutefois, A1 est acceptée comme une introduction sur le mode indésirable. En A4, *bois-abîmé* se retrouve trivialement faux par la clause de bon sens n° 2. A4 est pourtant analysée comme une réplique antagoniste, et agit logiquement comme *non bois-abîmé* (stratégie ❷' par la clause 12).

PARADISE est ainsi capable d'analyser les répliques de son interlocuteur. Il nous faut maintenant expliquer comment il synthétise ses propres arguments.

### 6.2. Synthèse des arguments par PARADISE

Lors de son démarrage, PARADISE recherche une clause saturée dans sa connaissance. S'il peut en trouver une, il émet l'un de ses termes et réalise ainsi une introduction sur le mode paradoxal ou indésirable, selon la valeur de la modalité de la clause (par ex. A1 lorsque PARADISE joue le rôle de A, voir annexe). Cette clause devient la première clause de référence.

Dans tous les autres cas, PARADISE recherche une faille dans la clause de référence située en sommet de pile. Pour cela, il va essayer d'appliquer les trois stratégies mentionnées plus haut "à l'envers" (*i.e.* de  $CR_k$  vers  $R$ ). Une réplique invalidante  $R$  est un bon candidat, et sera donc émise si sa valeur de vérité est vraie ou inconnue.

La stratégie ❷ est essayée en premier, lorsque la dernière réplique analysée  $R_0$  peut être infirmée : [ $R \ \& \ \langle \textit{autres termes vrais} \rangle$ ]  $\Rightarrow$  *non*  $R_0$ .  $R$  est alors une bonne tentative d'invalidation.

Si cela n'est pas possible, c'est la stratégie ❶ qui est essayée, si un terme de  $CR_k$  peut être mis en doute.

Si cela échoue aussi, PARADISE décide de remettre récursivement en question une connaissance apprise, et cherche à appliquer la stratégie ③. C'était le cas en B5: *travail-fastidieux* est remis en question dans la clause 1. Mais il est établi par la clause 10, qui est valide à ce point de la conversation. PARADISE remet *décapage* en question dans la clause 10. Ce terme semble provenir de la clause 5, et là PARADISE parvient à trouver une faille.

## 7. Limitations actuelles de PARADISE

Le but de PARADISE est de reproduire la performance des interlocuteurs en utilisant une connaissance pragmatiquement pertinente. Idéalement, le problème d'expliquer la dynamique de l'échange argumentatif est ramené, grâce à PARADISE, à celui d'isoler cette connaissance statique (*i.e.* déclarative). PARADISE constitue donc un premier pas dans la voie qui mène vers la modélisation et la synthèse des discours argumentés. Nous devons toutefois mentionner quelques limites actuelles de la performance de PARADISE .

Nous commencerons par commenter les répliques additionnelles B3' et A4' à propos de la tendreté du bois. Si nous ôtons de la connaissance toute mention sur la tendreté du bois, ou si nous ajoutons le fait que le bois est tendre, alors PARADISE ne suggérera pas *brosse-métallique*. La question est donc la suivante: pourquoi B, dans le dialogue réel, a-t-il émis la suggestion B3, mais pas B3' ? Nous pouvons donner deux types de réponse en nous fondant sur notre modèle de la connaissance de B. Soit B n'a pas la connaissance modélisée par la clause 12 présente à l'esprit au moment de B3, soit B "sait" avant de dire B3' que A ne peut pas oublier des prémisses factuelles (comme *plusieurs-couches*, *bois-tendre*, etc.). Ces deux hypothèses peuvent être implantés, par exemple grâce à une distinction entre connaissance opérationnelle et connaissance non opérationnelle, comme cela est fait en EBL (voir chap. II.3).

Il y a d'autres limites à la capacité de PARADISE d'imiter le comportement conversationnel. La plus évidente concerne l'ordre des répliques équivalentes. B2 (*chalumeau*) et B3 (*brosse-métallique*) sont produites grâce au même mécanisme et ne sont émises dans le bon ordre que par chance. On pourrait dire la même chose de B5 (*ponçage*) et de B7 (*enduit*), mais la clause 5 doit être absolument utilisée en premier car un seul de ses termes est inconnu, alors que la clause 6 comporte deux termes inconnus et est donc beaucoup plus "risquée".

Un autre problème vient de ce que PARADISE analyse les répliques en se référant à la clause  $CR_k$  située en sommet de la pile. Il ne peut donc pas suivre les retours en arrière (*backtrackings*), qui se traduisent par l'enlèvement d'une ou plusieurs clauses du sommet de la pile, sauf lorsqu'une acceptation explicite est donnée.

Nous pouvons encore trouver des limitations à PARADISE , notamment dans le fait qu'il n'utilise pas les conditions "suffisantes" (règles 4, 7, 8, 11) pour tester les suggestions. Il semble que les conditions "nécessaires" sont utilisées par les interlocuteurs pour découvrir les arguments pertinents, puisque PARADISE y parvient par ce mécanisme. Mais les interlocuteurs vérifient la suffisance des suggestions, comme B7 le montre clairement: "si la vieille peinture tient bien,..."(cf. clause 7).

## 8. PARADISE en tant que modèle cognitif

Les limitations mentionnées ci-dessus sont provisoires, et elles ne doivent pas masquer les progrès que PARADISE nous permet d'accomplir dans la compréhension de l'analyse et de la synthèse des arguments par les interlocuteurs. Chaque détail du fonctionnement de PARADISE est en lui-même un modèle testable du comportement conversationnel humain.

Considérons par exemple la capacité qu'a PARADISE de remettre en question des informations apprises. Cette capacité n'est déclenchée que lorsque les autres modes de réplique ont été explorés en vain. C'est ce mécanisme qui explique pourquoi les suggestions *chalumeau* et *brosse-métallique* ont été émises en premier, avant *ponçage* et *enduit*. Ceci constitue un modèle testable et prédictif d'un aspect de la génération d'arguments.

---

Si nous voulons que PARADISE constitue un modèle cognitif de la conversation, nous devons le faire reposer sur des mécanismes simples. C'est à peu près le cas puisque son comportement de base est dicté par seulement trois stratégies (malgré nos efforts pour le garder simple et concis, le module conversationnel contient tout de même à lui seul environ 700 clauses Prolog). Nous travaillons à simplifier encore davantage les mécanismes fondamentaux de PARADISE.

## 9. Perspectives

Si PARADISE, ou une de ses versions ultérieures, peut être considéré comme un modèle adéquat de la compétence conversationnelle, alors nous pouvons imaginer de l'utiliser différemment. Si nous sommes capables d'engendrer correctement les arguments lorsque la connaissance est disponible et non ambiguë, alors PARADISE peut être utilisé à l'inverse pour valider la connaissance dans des situations moins claires. Nous serions alors en mesure d'inférer, à partir de la performance conversationnelle, quelle est la connaissance pragmatiquement pertinente que nous devons prêter aux interlocuteurs.

L'élucidation de la compétence conversationnelle, pour laquelle PARADISE constitue un premier pas, ouvre la voie à de nombreuses applications. Nous en avons exploré certaines, en EIAO (chap. III.2) et pour la génération d'explications dans les systèmes à base de connaissance (chap. III.4). Plus généralement, on aurait de la peine à imaginer que les systèmes du futur, qui seront certainement capables d'argumenter leurs connaissances et leurs décisions d'une manière humainement acceptable, puissent le faire sans utiliser des principes qui ressemblent à ceux que nous avons commencé d'explorer avec PARADISE.



## Annexe : exemples de reconstruction de dialogues par PARADISE

Nous donnons tout d'abord la reconstruction effectuée par PARADISE de l'extrait traité dans le chapitre I.3. Cette fois, PARADISE joue le rôle de A.

<i>conversation d'origine</i>	<i>reconstitution par PARADISE</i>
<i>A1- Ben moi, j'en bave actuellement parce qu'il faut que je refasse mes portes, la peinture. Alors j'ai décapé à la chaleur. Ça part bien. Mais pas partout. C'est un travail dingue, hein?</i>	A1• C'est quand même embêtant que repeindre constitue un travail dingue
<i>B2- Quelle chaleur? La lampe à souder?</i>	B2• on utilise un chalumeau
<i>A3- Ouais, avec un truc spécial.</i>	A3• C'est quand même embêtant que repeindre constitue un travail dingue
[...]	B3• on utilise une brosse métallique
<i>B3- Faut une brosse, dure, une brosse métallique.</i>	A4• on n'utilise pas une brosse métallique sinon on attaque le bois puisque le bois des portes est tendre
<i>A4- Oui, mais j'attaque le bois.</i>	
<i>B4- T'attaques le bois.</i>	
[pause 5 secondes]	B4• oui
<i>A5- Enfin je sais pas. C'est un boulot dingue, hein? C'est plus de boulot que de racheter une porte, hein?</i>	A5• C'est quand même embêtant que repeindre constitue un travail dingue
<i>B5- Oh, c'est pour ça qu'il vaut mieux laiss... il vaut mieux simplement poncer, repeindre par dessus</i>	B5• on ne décape pas les portes • on ponce les portes
<i>A6- Ben oui, mais si on est les quinzièmes à se dire ça</i>	
<i>B6- Ah oui.</i>	
<i>A7- Y a déjà trois couches de peinture, hein, dessus.</i>	A7• on décape les portes sinon l'état de surface n'est pas joli puisque il y a plusieurs couches superposées
<i>B7- Remarque, si elle tient bien, la peinture, là où elle est écaillée, on peut enduire. De l'enduit à l'eau, ou</i>	B7• on passe de l'enduit
<i>A8- Oui, mais l'état de surface est pas joli, quoi, ça fait laque, tu sais, ça fait vieille porte.</i>	A8• on décape les portes sinon l'état de surface n'est pas joli puisque il y a plusieurs couches superposées
	B8• OK
	A9• C'est quand même embêtant que repeindre constitue un travail dingue

L'extrait suivant est relatif au comportement des consommateurs qui achètent préférentiellement des poulets portant la mention "garanti poulet fermier". Pour *A*, ce comportement est incohérent, dans la mesure où la mention ne signifie rien, puisqu'elle est apposée sans aucun contrôle.

*A1- Alors ce qui impressionne les populations, c'est le mot garantie . Un gars qui fait des poulets d'élevage, il met garanti fermier, alors tout le monde se précipite.*  
*B1- Mais il y a des labels*  
*A2- Oui, mais garantis par qui?*  
*B2- Mais ils le disent, par qui, parfois... par la chambre syndicale des machins...*

Il s'agit d'une conversation sur le mode paradoxal. Pour *A*, tous les termes de la première clause ci-dessous sont simultanément vérifiés, ce qui produit une situation incohérente. Nous donnons l'ensemble de la connaissance qui permettra la reconstruction :

connaissance commune

- [ achat-dû-à(**Mention**) & non garanti(**Mention**) ]  $\Rightarrow$  **F**
- [ label(**L**) & accompagne(**L**,**Mention**) & garanti(**L**) & non garanti(**Mention**) ]  $\Rightarrow$  **F**
- [ garanti(**X**,**LM**) & non garanti(**LM**) ]  $\Rightarrow$  **F**
- [ **non** achat-dû-à(**garanti-fermier**) ]  $\Rightarrow$  **F**

connaissance de A

- [ garanti(*garanti-fermier*) ]  $\Rightarrow$  **F**

connaissance de B

- [ **non** label(*certain-labels*) ]  $\Rightarrow$  **F**
- [ **non** accompagne(*certain-label*, *garanti-fermier*) ]  $\Rightarrow$  **F**
- [ **non** garanti(*certain-label*,*syndicat-des-machins*) ]  $\Rightarrow$  **F**

Voici la reconstruction de PARADISE, d'abord dans le rôle de *B*, puis dans celui de *A*.

<i>conversation d'origine</i>	<i>reconstitution par PARADISE (dans le rôle de B)</i>
<i>A1- Alors ce qui impressionne les populations, c'est le mot garantie . Un gars qui fait des poulets d'élevage, il met garanti fermier, alors tout le monde se précipite.</i>	<i>A1</i> • <i>l'achat est provoqué par la mention Garanti-Poulet-Fermier</i> • <i>la mention Garanti-Poulet-Fermier n'est pas garanti(e)</i>
<i>B1- Mais il y a des labels</i>	<i>B1</i> • <i>la mention Garanti-Poulet-Fermier est garanti(e) puisque Certain-Label accompagne la mention Garanti-Poulet-Fermier et que Certain-Label est garanti(e)</i>
<i>A2- Oui, mais garantis par qui?</i>	<i>A2</i> • <i>Certain-Label n'est pas garanti(e)</i>
<i>B2- Mais ils le disent, par qui, parfois... par la chambre syndicale des machins...</i>	<i>B2</i> • <i>Certain-Label est garanti(e) puisque Certain-Label est garanti par le syndicat des machins</i>

On constate que PARADISE utilise les possibilités du moteur d'ordre 1 (voir annexe chap. III.2). Cela apparaît encore plus clairement dans la réplique *A2'* ci-dessous où PARADISE tente une invalidation en suggérant un terme non instancié (*pas garanti par Quelqu'un*).

<i>conversation d'origine</i>	<i>reconstitution par PARADISE (dans le rôle de A)</i>
<i>A1- Alors ce qui impressionne les populations, c'est le mot garantie . Un gars qui fait des poulets d'élevage, il met garanti fermier, alors tout le monde se précipite.</i>	A1• C'est étonnant que l'achat est provoqué par la mention Garanti-Poulet-Fermier puisque la mention Garanti-Poulet-Fermier n'est pas garanti(e)
<i>B1- Mais il y a des labels</i>	B1• <i>Certain-Label est un label</i> • <i>Certain-Label accompagne la mention Garanti-Poulet-Fermier</i>
<i>A2- Oui, mais garantis par qui?</i>	A2• Mais il est possible que Certain-Label n'est pas garanti(e) B1'• <i>Certain-Label est garanti(e)</i>
	A2'• On peut éviter que Certain-Label est garanti(e) ; il n'y a qu'à considérer que Certain-Label n'est pas garanti par Quelqu'un
<i>B2- Mais ils le disent, par qui, parfois... par la chambre syndicale des machins...</i>	B2• <i>Certain-Label est garanti par le syndicat des machins</i>



## **II- Apprentissage des concepts**

<b>II.1</b>	<b>Modèles de l'apprentissage</b>	<b>87</b>
<b>II.2</b>	<b>Limites de l'apprentissage indifférent</b>	<b>105</b>
<b>II.3</b>	<b>L'apprentissage des connaissances conceptuelles dirigé par l'expérience</b>	<b>137</b>
<b>II.4</b>	<b>Plausibilité cognitive d'un appareil conceptuel autonome</b>	<b>167</b>



## Introduction à la partie II

### Apprendre pour converser

#### *L'apprentissage des connaissances conversationnelles révèle une spécialisation cognitive*

Notre objectif, que nous réaliserons dans la partie III, est d'utiliser une interaction de type conversationnel pour enseigner des connaissances à des étudiants. Mais quel type de connaissances peut-on enseigner ainsi ? Pour y répondre, nous devons savoir quel est le type des connaissances qui sont manipulées dans les conversations, et comment les interlocuteurs les ont acquises.

Une réponse qui vient à l'esprit concernant cette dernière question consiste à penser que les individus *extraient* ces connaissances directement, à partir de leur expérience de nombreuses situations semblables. Ils parviendraient ainsi à juger de l'incohérence d'une situation, comme ils le font couramment dans les conversations, en la comparant à l'ensemble des situations normales qu'ils ont vécues dans les mêmes circonstances. Mais nous montrons qu'un tel mécanisme d'extraction ne peut pas expliquer la performance conversationnelle.

Nous proposons une solution différente : l'incohérence d'une situation est manipulée en tant que telle, comme une relation logique, par l'individu qui la vit. Dans cette façon de voir, les connaissances sont acquises par des mécanismes symboliques (logiques) opérant sur des *concepts*.

Dans ce cas, il reste à expliquer d'où viennent les concepts. Comment sont-ils formés, ou comme on dit dans le cas de symboles, comment sont-ils *ancrés* ? Là encore, nous examinons une alternative.

Une première hypothèse serait de considérer les concepts comme des perceptions intégrées, *extraites* des perceptions directes. Leur représentation physique dans le cerveau ressembleraient à celle des perceptions qui les évoquent une fois qu'ils sont acquis. C'est ce que nous appelons l'ancrage *analogique* des concepts. Malheureusement, l'invocation d'un mécanisme d'extraction se heurte encore une fois à un obstacle majeur, qui réside cette fois dans la systématisme de certains processus conversationnels.

L'autre hypothèse consiste à envisager un ancrage des concepts par *appariement*. Elle nous conduit à postuler l'existence d'un appareil conceptuel autonome qui aurait une représentation matérielle dans le cerveau.

Cet appareil aurait donc pour principale fonction d'opérer des manipulations symboliques sur les concepts. Or on reproche souvent à ce genre de modèle de n'être pas plausible : le cerveau renferme des neurones, pas des symboles ! Pour contrer cette critique, nous proposons un modèle neuronal biologiquement plausible d'un traitement conceptuel particulièrement important pour expliquer la compétence conversationnelle : la détection d'incohérence.

- *Le chapitre II.1 passe en revue un certain nombre de modèles de l'apprentissage, et en propose une classification.*

- *Le chapitre II.2 introduit la notion d'indifférence de certains mécanismes d'apprentissage. Ces mécanismes ne sont sensibles qu'aux propriétés relatives des objets qu'ils traitent, ils ne les comparent jamais à des valeurs fixes. Nous montrons que les mécanismes indifférents ne permettent pas d'expliquer l'apprentissage des connaissances conversationnelles.*
- *Dans le chapitre II.3, nous défendons l'idée qu'il existe dans le cerveau un appareil spécialisé dans le traitement des concepts. C'est grâce à cet appareil que nous pouvons converser et apprendre de nouvelles connaissances conceptuelles.*
- *Le chapitre II.4 est consacré à l'exposé d'un modèle neuronal plausible de cet appareil. Nous expliquons comment les relations logiques peuvent être encodées spatialement dans des ensembles de neurones, et comment l'incohérence peut être détectée par la synchronisation de tous les neurones de l'appareil.*

# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

Notre étude des aspects argumentatifs du langage nous conduit à constater que des connaissances peuvent être *appries* grâce au langage. Pourtant, ce processus d'acquisition est, de notre point de vue, très mal décrit par les principales théories de l'apprentissage. Comme notre objectif est d'utiliser les principes argumentatifs pour enseigner des connaissances, nous allons étudier, dans les chapitres de cette partie, comment peut s'opérer la formation de nouvelles connaissances, et quel rôle l'argumentation peut jouer dans ce processus.

Dans ce chapitre, nous passons en revue les principaux modèles de l'apprentissage qui sont couramment invoqués même lorsqu'il s'agit de rendre compte de l'acquisition de connaissances conceptuelles. Malgré la diversité des modèles, nous verrons qu'il est possible de distinguer deux classes fondamentales de mécanismes pour rendre compte de l'apprentissage. Cette distinction, qui est fondamentale lorsqu'il s'agit d'expliquer la formation de nouvelles connaissances, sera traitée en profondeur dans les chapitres II.2 et II.3.

## 2. Empirisme ou Instruction

La théorie empiriste est très bien définie par la citation suivante de John Locke [1690] pour qui toute idée provient soit de la *sensation*, soit de la *réflexion*:

"Let us then suppose the mind to be, as we say, white paper, void of all characters, without any ideas; how comes it to be furnished? [...] To this I answer, in one word, from *experience*. In that all our knowledge is founded, and from that it ultimately derives itself. [...]

First, our senses, conversant about particular sensible objects, do convey into the mind several distinct perceptions of things, according to those various ways wherein those objects do affect them. And thus we come by those *ideas* we have of *yellow, white, heat, cold, soft, hard, bitter, sweet*, and all those which we call sensible qualities. [...] This great source of most of the ideas we have, depending wholly upon our senses, and derived by them to the understanding, I call SENSATION.

Secondly, the other fountain from which experience furnisheth the understanding with ideas is the perception of the operations of our own mind within us, as it is employed about the ideas it has got; which operations, when the soul comes to reflect on and consider, do furnish the understanding with another set of ideas, which could not be had from things without. And such are *perception, thinking, doubting, believing, reasoning, knowing, willing*, and all the different actings of our own minds; [...] I call this REFLECTION."

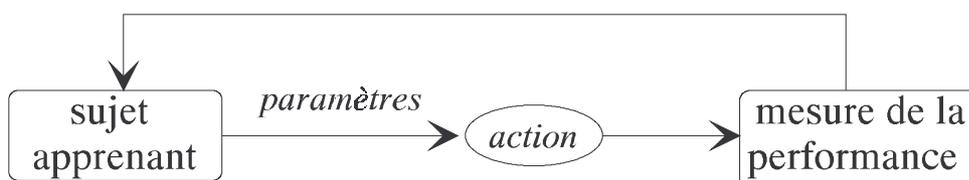
Exemples : ici, Locke parle bien de l'acquisition d'idées et de concepts. Notre esprit les reçoit comme une pellicule photographique reçoit des images, puis il les combine. Les théories instructives ont aussi été

invoquées dans d'autres contextes, par ex. aux débuts de l'immunologie pour rendre compte de la reconnaissance de n'importe quel corps étranger (y compris des molécules artificielles) par les anticorps.

Commentaire sur la théorie empirique : la théorie empirique prétend rendre compte de l'acquisition des connaissances. Si l'on accepte la thèse que nous développons dans les chapitres ultérieurs, la description donnée par Locke est auto-contradictoire : le fait que les "idées" puissent être combinées (notamment par les mécanismes conversationnels) exclut que certaines d'entre elles puissent être des sensations.

### 3. Apprentissage par optimisation

On présente souvent l'apprentissage comme une adaptation à une situation ou à un problème. A ce titre, les méthodes d'optimisation peuvent prétendre à être des modèles de certains apprentissages, notamment dans les problèmes à environnement constant. Nous ne mentionnerons pas les optimisations nécessitant une connaissance globale du problème (par ex. équation de la fonction à optimiser). Les méthodes d'optimisation sont généralement adaptatives : des valeurs de paramètres sont essayées, et la mesure de la performance donne une indication sur la façon de les modifier.



#### 3.1. Recherche aléatoire

La recherche aléatoire est une technique d'optimisation qui consiste à tirer à chaque fois au sort une combinaison entièrement nouvelle de paramètres, puis à mesurer la performance. Le système apprenant mémorise la meilleure combinaison obtenue jusque-là.

#### 3.2. Méthodes de descente

Les méthodes de descente (*hill climbing*) semblent pertinentes pour modéliser certains apprentissages (penser à un tir balistique):

- a- positionnement initial des paramètres, essai, mesure de la performance
- b- sélection d'un ou plusieurs paramètres qui sont modifiés d'une faible valeur (descente aveugle: variation aléatoire; descente de gradient: choix d'une variation "prometteuse" (*i.e.* qui donnerait une amélioration si elle était infinitésimale))
- c- essai avec les nouvelles valeurs, mesure de la performance
- d- adoption des nouvelles valeurs si la performance est augmentée
- e- retour en -b-

Exemple : tir à l'arc pour un novice

#### 3.3. Optimisation par mutation-sélection

Les modèles de mutation/sélection ressemblent à des descentes aveugles dotées d'un certain degré de parallélisme. Le principe de base équivaut à la description suivante:

- a- positionnement initial des paramètres d'une population de solutions, essais, mesures de performance
- b- variations aléatoires affectant (légèrement le plus souvent) un ou quelques paramètres de certaines solutions
- c- essai avec les nouvelles valeurs, mesure des performances
- d- sélection des meilleures solutions obtenues (anciennes ou nouvelles)
- e- retour en -b-

La recherche aléatoire peut être vue comme un système de mutation-sélection sans itération: l'ensemble des essais correspond à la population initiale.

Le modèle de mutation/sélection est souvent couplé à un système d'hybridation de solutions, comme celui qui prévaut dans les systèmes génétiques des populations naturelles. Cela revient à ajouter une étape -b'-:

- b'- hybridation par couples des solutions ainsi modifiées

Les algorithmes génétiques tels qu'ils sont utilisés en optimisation font jouer un rôle essentiel à l'hybridation, alors que les mutations n'ont qu'un rôle secondaire. Ils utilisent ce que J. Holland a baptisé le parallélisme implicite [Goldberg 1989]. Dans ce cas le rôle de l'hybridation est simplement combinatoire: il est de regrouper des solutions partielles préexistantes et indépendantes. Une telle approche ne semble convenir que pour des problèmes "simples" [Dessalles 1992a]. En revanche, on peut concevoir des algorithmes plus "biomimétiques" qui, certes, utilisent l'hybridation, mais simplement pour "sauver" les solutions découvertes par mutation [Dessalles 1992a].

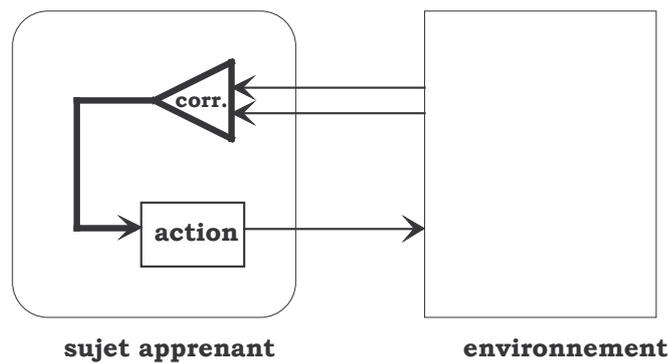
Exemples (systèmes de mutation-sélection): l'apprentissage d'un antigène par le système immunitaire, selon la théorie (généralement admise) de la sélection clonale que M. Burnet a établie en 1955. Le mécanisme de mutation (génération aveugle de diversité) suivi d'une sélection a été invoqué par Edelman et J-P. Changeux pour expliquer l'adaptation de certains aspects de la connectivité neuronale. Ces auteurs suggèrent même le fait que des mécanismes sélectifs peuvent être à la base de la cognition (cf. [Changeux & Dehaene 1989]). On peut même concevoir le progrès scientifique comme mû par un système de mutation/sélection (cf. [Walbridge 1989]).

Exemples (systèmes de mutation-hybridation-sélection): modèles génétiques des populations naturelles; algorithmes biomimétiques [Dessalles 1992a].

Commentaire sur l'optimisation : les techniques d'optimisation ont été conçues dans des contextes fort éloignés de la question de l'apprentissage des connaissances. Toutefois, plusieurs auteurs sont disposés, comme nous l'avons vu, à suggérer que les connaissances adéquates résultent d'un processus adaptatif. Si une telle théorie est séduisante au niveau socioculturel (cf. la théorie des "mêmes" de R. Dawkins [1982]), elle nous semble beaucoup plus discutable au niveau cognitif.

## 4. Apprentissage par extraction de régularités

Dans de nombreuses situations, l'entité apprenante doit adapter son comportement à un environnement changeant. L'apprentissage consiste alors à extraire les propriétés constantes et pertinentes de cet environnement :



Le résultat de l'apprentissage est présenté comme un "corrélateur" qui permet ensuite au sujet de reconnaître instantanément, au milieu de toutes les situations vécues, un environnement ayant les propriétés que les environnements, auxquels il a été exposé au cours de l'apprentissage, avaient en commun. Le corrélateur est éventuellement construit pour sélectionner une action.

Des schémas comme celui-ci ou équivalents à celui-ci sont souvent utilisés pour décrire des apprentissages humains, notamment en E.A.O. Il est donc nécessaire de décrire de tels mécanismes, car nous voulons montrer qu'ils ne peuvent pas suffire à rendre compte de l'acquisition de connaissances conceptuelles.

#### 4.1. *Le behaviorisme: apprentissage de régularités par conditionnement*

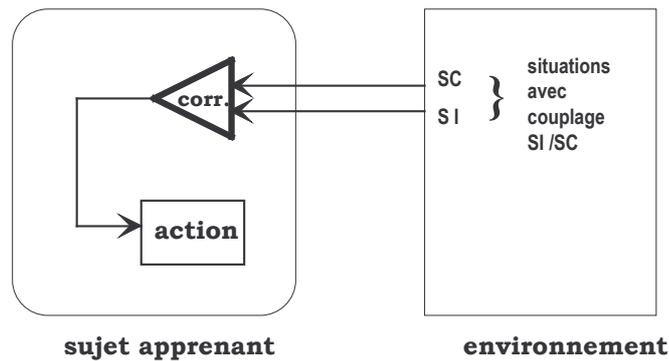
L'Ecole de pensée behavioriste est née à la suite des travaux de John Broadus Watson [Watson 1913]. Le Behaviorisme se préoccupe avant tout de psychologie animale. Son objectif était de rendre compte du comportement animal (puis humain), notamment de la faculté d'apprentissage dont les organismes vivants sont dotés, en rejetant tout "mentalisme". Le Behaviorisme va adopter en fait une position extrémiste: il s'interdit *a priori* toute hypothèse sur le fonctionnement interne de l'organisme étudié. On ne prend en compte que ce qui est "visible" objectivement, c'est-à-dire les stimuli (les entrées du système) et les réponses (les sorties du système).

La thèse centrale du behaviorisme est que tout comportement est la conséquence d'un stimulus, et que l'apprentissage d'un comportement, qu'il soit le fait d'un invertébré ou d'un mammifère supérieur, procède d'un conditionnement qui peut prendre deux formes.

##### 4.1.1. **Le conditionnement classique**

C'est l'association, mise en évidence dans l'expérience de Pavlov, entre deux stimulus: le stimulus inconditionnel (SI) et le stimulus conditionnel (SC). Cette association due à la répétition permet à un nouveau stimulus (par ex. SC = un son de cloche) de remplacer le stimulus "naturel" (SI = boulette de viande) et de provoquer le même comportement (salivation). Ce type d'apprentissage permet d'expliquer une certaine souplesse dans les conditions de production d'un comportement<sup>(1)</sup>.

<sup>1</sup> Le conditionnement classique est relativement limité. Il ne crée pas de nouveau comportement. L'association entre les deux stimulus est rapidement oubliée, et nécessite donc une exposition fréquente. De plus, le conditionnement de deuxième ordre, c.à.d. l'association avec un stimulus appris, fonctionne déjà beaucoup plus mal (habituation plus longue). Enfin, les associations ne sont pas toutes possibles [Eibl-Eibesfeldt 1967 p.260].

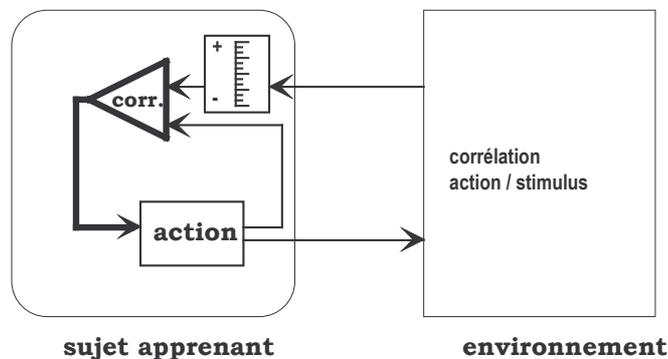


Exemple: dans l'expérience célèbre de Pavlov, un chien apprend à associer un son de cloche (SC) à la boulette de viande (SI) qui provoque une salivation. Après avoir été exposé à maintes reprises à des situations où les deux stimuli sont couplés, le chien salive en entendant la cloche.

#### 4.1.2. Le conditionnement opérant

C'est la version scientifique du dressage. La théorie élaborée par Burrhus F. Skinner montre qu'un stimulus agréable ou déplaisant peut augmenter ou diminuer respectivement la fréquence de production spontanée du comportement qui l'a précédé :

"Dès que nous avons établi le type particulier de conséquence que nous appelons un renforcement, nous sommes en mesure de modeler le comportement d'un organisme pour ainsi dire à notre gré. [...] Si nous prenons soin de présenter de la nourriture à un pigeon affamé à certains moments précis, [...] nous pouvons lui apprendre à tourner en rond, à décrire des huit sur le sol de sa cage, à se tenir immobile dans un coin, à tendre le cou ou à frapper du pied." [Skinner 1969].



Le stimulus provenant de l'environnement est évalué sur une échelle linéaire (plaisir / déplaisir), puis corrélé à l'action qui vient d'être produite<sup>(2)</sup>.

Exemple: le pigeon de Skinner est récompensé chaque fois qu'il décrit un rond en marchant. Il corréle ainsi le stimulus positif avec l'action qu'il a produite.

Le conditionnement, qu'il soit classique ou opérant, peut être présenté comme la construction d'un "corrélateur" qui va pouvoir repérer, parmi toutes les situations vécues, une corrélation temporelle telle qu'elle existe dans l'environnement.

<sup>2</sup> Certains aspects du dressage semblent échapper aux lois du conditionnement opérant. Par exemple, un cavalier frappe son cheval de sa cravache lorsque le cheval n'a pas produit le comportement souhaité. Dans ce cas, la punition doit être corrélée par l'animal avec une action qu'il n'a pas produite ! Cette opération, impossible dans le cadre du conditionnement opérant, ne peut être envisagée que si l'animal possède une représentation du comportement qu'il aurait dû produire.

#### 4.2. *Le connexionnisme : extraction de régularités par des neurones artificiels*

Les réseaux connexionnistes, dans leurs implémentations classiques (réseaux de Hopfield, réseaux perceptrons multicouches, cartes de Kohonen, réseaux à retour adaptatif, etc.) fonctionnent en extrayant des régularités dans les motifs présentés en entrée pendant la phase d'apprentissage.

Cette propriété est la conséquence d'une propriété plus générale de ces réseaux. Si l'on prend l'exemple des réseaux de neurones à états binaires, ils effectuent une application de l'espace des entrées dans l'hypercube. Chaque point de l'hypercube unité de dimension  $n$  représente un état d'un réseau de  $n$  neurones binaires. Le fonctionnement du réseau à un moment donné (pour un réseau déterministe) associe un état mémorisé à chaque configuration des entrées. L'effet de l'apprentissage est de rendre cette application continue: deux configurations voisines en entrée provoqueront des états voisins du réseau.

On comprend ainsi qu'un tel réseau soit capable d'extraire des régularités : une régularité rendra voisines certaines configurations d'entrée (par ex. en reconnaissance de caractères, une ligne horizontale à la base) et le réseau répondra de manière proche à ces configurations, et de manière significativement différente aux autres configurations (par ex. les formes sans ligne horizontale, si cette ligne est la seule régularité objectivement présente dans les données).

Exemple: Un réseau de Kohonen regroupe sur une couche carrée des phonèmes représentés en entrée par leur valeur sur 15 fréquences (d'après [T. Kohonen 1984]). Les phonèmes similaires se retrouvent proches dans la couche de Kohonen (voir figure).

o	o	a	a	æ	æ	e	e	e	e	e	e	e	e	e
u	o	a	a	æ	ø	ø	e	i	i	i	i	i	i	i
u	u	h	h	r	ø	ø	e	i	i	i	i	i	i	i
v	v	v	r	r	ø	y	j	i	i	i	i	i	i	i
v	v	d	r	r	y	l	j	i	i	i	i	i	i	i
n	v	n	d	n	η	m	η	s	s	s	s	s	s	s
n	n	n	n	n	m	m	m	s	s	s	s	s	s	s

#### 4.3. *Extraction de régularités par des systèmes sélectifs*

Les systèmes de mutations-sélection sont capables d'extraire des régularités d'un environnement changeant. Par exemple, le système immunitaire peut découvrir une configuration constante à la surface d'un virus doué d'une forte variabilité.

De même, le pool génétique d'une population animale peut enregistrer la corrélation entre le cri d'un prédateur et le danger qu'il représente, en augmentant la proportion des individus ayant une peur innée du cri en question.

#### 4.4. *Extraction de régularités par l'analyse de données*

L'analyse de données regroupe plusieurs méthodes permettant d'extraire des régularités. L'analyse de données *exploratoire* contient trois grands groupes de méthodes : analyse factorielle, classification hiérarchique, partitionnement. L'analyse de données *confirmatoire*, par exemple l'analyse discriminante, consiste à trouver les variables qui séparent correctement des classes données au départ.

Nous mentionnons ci-dessous une méthode d'analyse factorielle (analyse en composantes principales) et une méthode de partitionnement.

##### 4.4.1. *L'analyse en composantes principales (ACP)*

L'objectif est de rechercher les facteurs qui représentent le mieux les individus et permettent éventuellement de les classer.

$$\begin{array}{c}
 \text{individu2} \\
 \downarrow \\
 \left( \begin{array}{cccc}
 x_{11} & x_{12} & \cdots & x_{1m} \\
 x_{21} & x_{22} & \cdots & x_{2m} \\
 x_{31} & x_{32} & \cdots & x_{3m} \\
 x_{41} & x_{42} & \cdots & x_{4m} \\
 \cdots & \cdots & \cdots & \cdots \\
 x_{n1} & x_{n2} & \cdots & x_{nm}
 \end{array} \right) \leftarrow \text{critère 4}
 \end{array}
 \quad x_{ij}: \text{valeur de la } i^{\text{ème}} \text{ caractéristique du } j^{\text{ème}} \text{ individu}$$

Les  $m$  individus peuvent être représentés dans un espace de dimension  $n$  (on parle de nuage d'individus). L'objectif de l'analyse factorielle (ici, l'ACP) va être de réduire la dimension de cet espace en vue de résumer l'information [Choukair 1993].

L'ACP s'intéresse à des tableaux de caractéristiques quantitatives concernant des individus. Si  $X$  est la matrice  $n \times m$  correspondant à ce tableau, on forme:  $X \cdot {}^t X$  ou  ${}^t X \cdot X$ , selon que  $n < m$  ou que  $n > m$ .

Par ex. l'élément  $(i, j)$  de  $X \cdot {}^t X$  est  $\sum_k x_{ik} \cdot x_{jk}$ . Il correspond (à  $1/m$  près) à l'estimation de la covariance entre les critères  $i$  et  $j$ , lorsque l'on a, au préalable, centré tous les critères. Cela donne une matrice carrée symétrique que l'on diagonalise. Les vecteurs propres sont orthogonaux et les valeurs propres positives. Par exemple pour  $X \cdot {}^t X$ , la valeur propre  $\lambda_i$  est égale à  $\sum_k x'_{ik}{}^2$ , où  $x'_{ik}$  désigne la valeur de l'individu  $k$  selon le  $i^{\text{ème}}$  vecteur propre, et elle mesure donc une estimation de la variance des individus selon cet axe. On projette ensuite tant les individus que les points représentatifs des critères (extrémités des vecteurs de la base initiale) sur le sous-espace engendré par les premiers vecteurs propres rangés par valeurs propres décroissantes. Le premier vecteur propre indique l'allongement maximum du nuage (variance maximum), le second vecteur l'allongement maximum du nuage projeté dans le sous-espace supplémentaire orthogonal au premier vecteur, et ainsi de suite. On peut donc assurer que la distorsion due à la projection finale (si on ne garde que  $p$  vecteurs propres) est minimale. Si  $p < 3$ , les "distances" entre individus et critères peuvent être interprétées à l'oeil par l'observateur.

Si l'ensemble des individus possède une régularité (par ex. une combinaison de caractéristiques ayant la même valeur pour la plupart des individus), l'ACP va extraire cette régularité comme un vecteur propre de valeur propre minimale.

#### 4.4.2. Algorithmes de classification par partitionnement

Les algorithmes de classification consistent à fabriquer une partition de manière à minimiser la distance moyenne intra-classe tout en maximisant la distance inter-classes.

Le choix d'une partition convenable n'est pas *a priori* évident : s'il y a  $K$  objets, il faut choisir parmi les  $B_K$  partitions possibles, où  $B_K$  est le nombre de Bell (cf. [Favaron 1990]) donné par:

$$B_K = \frac{1}{e} \sum_{i>0} \frac{i^K}{i!}$$

Ce nombre atteint rapidement des valeurs gigantesques (il y a 1,4 milliard de façons de classer 15 objets). Les mécanismes permettant de classer des éléments en tenant compte des relations qui les lient ont été étudiées sur le plan mathématique (cf. [de Amorim 1990]). Par exemple, supposons que l'on dispose, dans l'ensemble  $V$  des objets à classer, d'une mesure de similarité graduelle  $s_{ij}$  entre tous objets  $i$  et  $j$  telle que:

$$\forall i \in V ; \quad s_{ii} = 0$$

$$\forall i, j \in V ; \quad s_{ij} = s_{ji}$$

(la première contrainte peut être satisfaite par une renormalisation). Le problème consistant à trouver une partition (ou, ce qui revient au même, une relation d'équivalence) qui respecte les liens de similarité peut être formulé comme suit :

$$\text{trouver la relation d'équivalence qui maximise } \sum_{i,j \in [1,K]} s_{ij} \cdot x_{ij}$$

où  $x_{ij}$  est la fonction caractéristique du graphe de la relation d'équivalence cherchée. Une solution de ce problème rend maximale la somme des ressemblances entre éléments classés ensemble pris deux à deux. Trouver une telle solution est de complexité NP-difficile. Il en est de même lorsqu'on dispose d'une mesure de similarité qualitative.

La méthode de partitionnement dite des centres mobiles consiste à rechercher par améliorations successives une partition du nuage en  $n$  classes ( $n$  étant donné par avance). La méthode consiste à partir de  $n$  individus "typiques" choisis au hasard, puis de regrouper les autres individus autour d'eux selon leur distance. On change ensuite d'individus typiques en les remplaçant par un point situé sur les (ou à proximité des) barycentres des groupes ainsi obtenus.

Si une partie des individus présente une propriété en commun (combinaison de caractéristiques), elle est isolée comme une classe par l'algorithme de partitionnement.

Commentaire sur l'extraction de régularités : beaucoup de théories psychologiques tiennent pour acquis le fait que les concepts sont formés par extraction de régularités à partir d'objets perçus hétéroclites. A part le behaviorisme qui nie l'existence même des concepts en tant qu'objets scientifiques, les autres mécanismes d'extractions de régularités que nous venons de citer (connexionnisme, sélection, analyse de données) peuvent être invoqués comme modèles pour l'extraction des connaissances conceptuelles. Nous nous opposerons à ce point de vue.

#### 4.5. *Apprentissage par compilation*

Nous regroupons sous le terme de *compilation* un certain nombre de mécanismes qui ont été proposés pour expliquer certains apprentissages. La théorie ACT\* d'Anderson [1983] compte parmi ses principes de base la possibilité que certaines connaissances (il s'agit chez Anderson d'enchaînements de règles de production permettant des actions) deviennent plus compactes et plus accessibles du fait de leur usage fréquent. La théorie des scripts [Schank 1984] suggère elle aussi que les unités de connaissances qui seront utilisées dans la plupart des actions sont issues d'un compactage dû à la répétition.

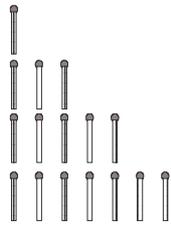
Commentaire sur la compilation : il est indéniable que les connaissances souvent utilisées deviennent plus accessibles, et que si elles résultent d'une combinaison, la forme combinée pourra être mémorisée en tant que telle. Nous pensons néanmoins que, contrairement aux prétentions souvent assignées à la théorie de la compilation, cet effet n'est pas le plus fondamental lorsqu'il s'agit d'expliquer la formation des connaissances conceptuelles. La plupart des concepts que nous mémorisons le sont à la suite d'une seule utilisation. Si la répétition joue un rôle, nous pensons qu'il est marginal par rapport à celui que jouent les relations logiques de type conversationnel.

### 5. **Innéisme et maturation**

Certaines facultés qui sont acquises au cours du développement de l'individu semblent apprises, mais peuvent n'être que l'effet du développement lui-même, sans que l'environnement joue de rôle dans la forme particulière des nouvelles facultés. Il ne s'agit donc pas dans ce cas d'apprentissage à proprement parler, et de

telles facultés seraient innées malgré leur apparition progressive. Ces questions étant polémiques lorsqu'il s'agit du développement des individus humains, nous préférons prendre un exemple sur une entité artificielle.

En 1980, nous avons développé un petit programme pour étudier l'apprentissage dans une version du jeu de Nim, popularisée par le film "L'année dernière à Marienbad" <sup>(3)</sup>.



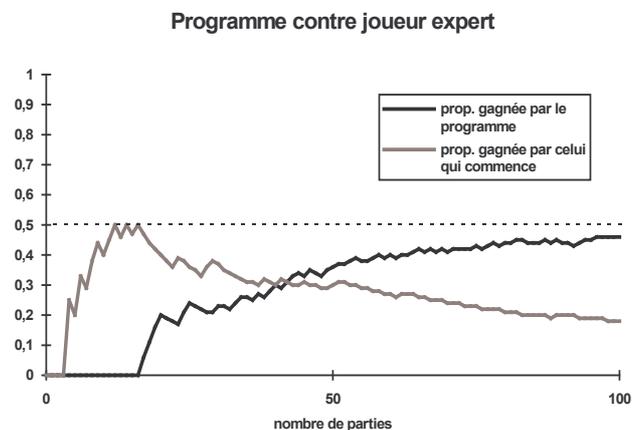
configuration de départ

Chaque joueur retire à son tour des allumettes en nombre quelconque, mais dans une seule rangée. Le perdant est celui qui prend la dernière allumette.

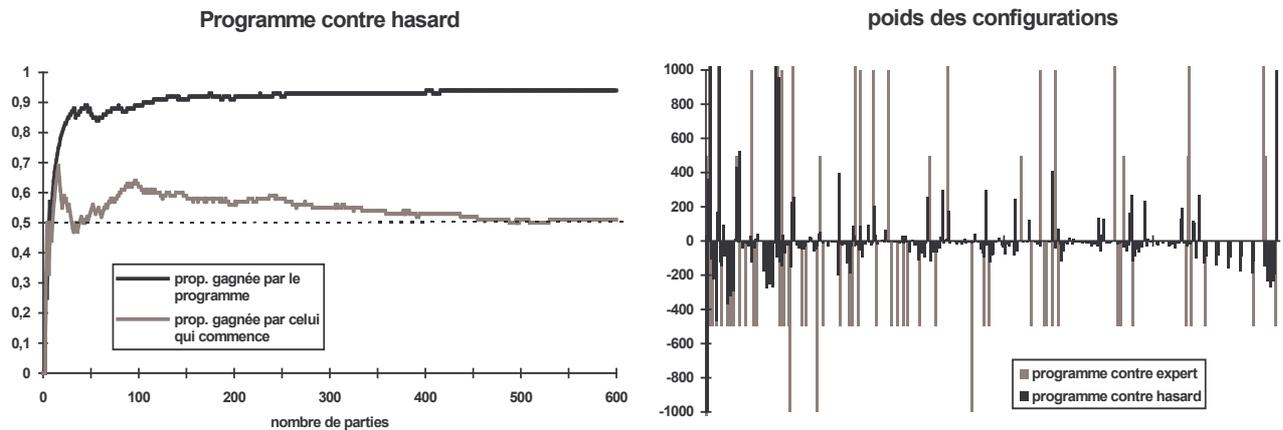
Ce jeu est biaisé, dans la mesure où celui qui commence est assuré de perdre si l'adversaire joue correctement. Notre programme fonctionne de la manière suivante : il a une représentation en mémoire de toutes les configurations possibles (il y en a 384). A chaque configuration est associé un poids. Quand il doit jouer, le programme choisit la configuration accessible ayant le poids maximum. A l'issue de chaque partie, le poids des différentes configurations que le gagnant a laissées à son adversaire pendant la partie est incrémenté, alors qu'il est décrémenté pour celles que le perdant a laissées.

On constate sur la figure ci-contre que le programme apprend rapidement contre un joueur expert. En quelques dizaines de parties, il "connaît" les bonnes configurations qui vont lui permettre de battre son professeur chaque fois que celui-ci commence (*i.e.* une fois sur deux en moyenne).

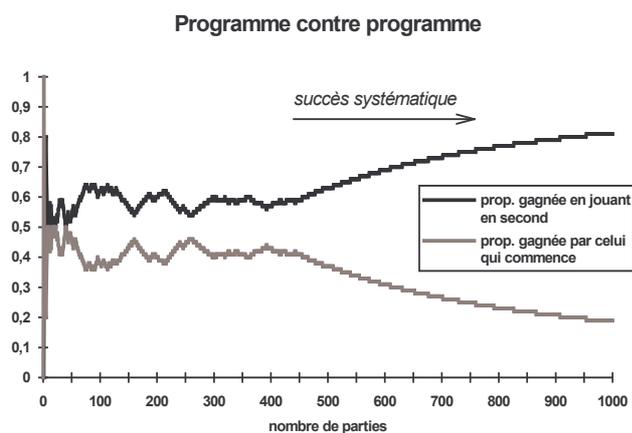
Il semble à ce stade que l'environnement joue un rôle déterminant dans l'apprentissage du programme. Pour tester ce point, on peut imaginer de faire jouer le programme contre le hasard. Effectivement, les résultats sont décevants. Le programme parvient facilement à vaincre un adversaire qui joue ses coups aléatoirement (voir figure ci-dessous), mais il n'apprend pas les configurations qui lui permettraient ensuite de jouer contre un expert : on constate sur le schéma des poids que les configurations retenues (celles qui ont un poids élevé en valeur absolue) diffèrent sensiblement des configurations correctes qui ont été apprises contre l'expert, ce qui est rédhibitoire pour l'efficacité. On en retire l'impression que le programme acquiert une information structurée seulement lorsqu'elle lui est fournie par l'environnement : dans le cas du joueur expert, il acquiert rapidement une information parfaitement structurée, alors que dans le cas du hasard, l'information extraite est très pauvre (il n'apprend à reconnaître que quelques configurations terminales).



<sup>3</sup> Nous remercions Cédric Chappelier et Pascal Héry pour leur aide dans la caractérisation et la présentation de ces résultats.



En réalité cette impression, selon laquelle un apprentissage correct ne peut provenir que d'un environnement "professeur" correctement structuré, est trompeuse, on peut même dire erronée. Pour le montrer, il suffit de faire jouer le programme contre lui-même. Dans ce cas, aucune "information" ne provient d'un quelconque professeur externe au programme. Au début, le programme joue aléatoirement, il repère quelques configurations favorables, mais chaque fois que son *alter ego* les occupe avant lui, il est contraint d'explorer de nouvelles configurations. Ceci le mène à un apprentissage parfait des configurations favorables. Il est alors imbattable si on le confronte pour la première fois à un joueur externe.



Y a-t-il apprentissage dans le cas où l'ordinateur joue contre lui-même? Après tout, aucune information ne provient de l'extérieur. Certes, l'état du système se modifie, il découvre les "bonnes" configurations bien que celles-ci n'aient pas été mentionnées dans le programme (et fussent même inconnues du programmeur) et il acquiert ainsi une compétence qu'il n'avait pas au départ. Pourtant, nous préférons réserver le mot apprentissage pour décrire des situations où l'environnement joue un rôle structurant. Ce n'est pas le cas ici. Les bonnes configurations ne sont pas données *explicitement* au programme, mais elles le sont d'une manière implicite : celui-ci, en jouant seul tour à tour les deux rôles, a un moyen systématique de les

Commentaire sur la maturation : il est vraisemblable que certains types de connaissances conceptuelles soient inaccessibles aux enfants trop jeunes, quel que soit la richesse de l'environnement dans lequel ils sont plongés. Comme l'a montré Piaget, la réversibilité systématique de certaines actions ne peut pas être conçue ni utilisée par des enfants de moins de 7 ans. On peut facilement imaginer que cette carence résulte d'un développement inachevé (au même titre que les caractères physiques comme la mue de la voix), autrement dit d'une maturation qui reste à venir, et non d'une construction inachevée par manque de temps, comme semble le suggérer Piaget.

## 6. Apprentissage par appariement

Certaines connaissances sont acquises par des mécanismes qui diffèrent fondamentalement de tous ceux que nous venons de passer en revue. Ces nouveaux mécanismes sont caractérisés par leur efficacité

(apprentissage "en un coup") et leur sélectivité (sélection des conditions de mémorisation). Ils reposent sur une *appariement* entre une forme perçue et un assemblage d'éléments manipulés par le système apprenant.

### 6.1. *Instanciation, détection, empreinte*

Considérons le programme suivant, que nous avons baptisé "caractron" :

```

program caractron;
var M, C: char;
begin
  read(M);
  repeat
    read(C);
    if C = M then beep
  until C = '$'
end.

```

Ce programme lit un caractère donné au clavier, puis parvient à signaler la présence de ce caractère dans les entrées ultérieures de l'utilisateur. A la suite de la première lecture, il a donc *appris* à reconnaître un caractère. Le caractère ainsi appris est certes fourni par l'environnement, et cela modifie le comportement ultérieur du programme. Toutefois, la forme apprise était déjà connue ! Tous les caractères que le programme peut apprendre ainsi sont disponibles à l'avance. Ce sont les valeurs que la fonction *read* peut retourner, ou que la comparaison "=" accepte en deuxième argument. Cela

n'enlève rien au fait qu'un apprentissage a eu lieu.

**Empreinte** : le programme *caractron* est un modèle très simplifié du phénomène d'*empreinte* mis en évidence par K. Lorenz chez les jeunes oiseaux [Lorenz 1974]. Le jeune canard qui vient d'éclore prend pour sa "mère" toute forme de taille adéquate capable de se mouvoir (par ex. une caisse à roulette tirée par un fil), et se met à la suivre pendant plusieurs semaines. Lorenz montre, en modifiant les caractéristiques de l'objet d'attachement, comment l'oisillon est sensible de manière sélective à quelques caractéristiques de la forme perçue (taille, mouvements, pépiement). La présence de ces caractéristiques au cours de la période sensible (env. un jour après l'éclosion) sert de signal. Les autres caractéristiques de la "mère" (couleur, forme, etc.) sont alors instanciées, un peu comme pour le caractron.

**Détection** : le phénomène d'empreinte repose sur la détection, pendant la période sensible, de caractéristiques comme le mouvement d'un objet et le son qu'il émet. Ce phénomène de *détection* nous semble fondamental. Dans le caractron, la décision de terminer est prise lorsque le caractère \$ est détecté.

Un détecteur est un dispositif capable de reconnaître une forme. L'existence même du détecteur indique que la forme est en quelque sorte connue à l'avance. Le rôle du détecteur est simplement de repérer la présence de cette forme dans l'environnement. Dans [Dessalles 1992c] nous avons insisté sur l'importance de la détection dans les processus perceptifs. Nous suggérons ici que l'apprentissage des connaissances résulte de phénomènes de détection, au même titre que l'empreinte.

**Instanciation** : l'apprentissage des connaissances commence, en effet, par l'apprentissage des faits. Avant même d'*apprendre* que tel personnage public est décédé, nous savons qu'en tant qu'être vivant, il possède une caractéristique, son état de "fonctionnement", dont nous attendons qu'il prenne ses valeurs dans un ensemble discret (e.g. {*en-bonne-santé, malade, mort*}) pour une personne que nous connaissons peu). Lorsque nous apprenons qu'il est décédé, nous ne faisons qu'instancier cette caractéristique. Nous développerons en détail, dans les chapitres ultérieurs, le point de vue selon lequel l'apprentissage des connaissances conceptuelles utilise ce phénomène d'instanciation.

### 6.2. *Apprentissage par imitation*

L'imitation est souvent invoquée pour rendre compte de l'acquisition des informations. Nous voulons simplement rappeler ici que l'imitation est un processus complexe, et non un principe explicatif de base. En particulier, l'imitation semble comporter nécessairement une phase de décodage (et donc de détection), comme le montrent les observations de Visalberghi [1990]. Les singes capucins apprennent plus vite que les chimpanzés à atteindre une friandise placée dans un tube transparent, grâce à des baguettes trop courtes qu'ils parviennent à mettre bout à bout. Il apparaît en fait que leur succès relatif provienne d'une activité plus

intense conduisant à un nombre d'essais plus grand. Ces singes semblent, en revanche, incapables d'imiter : ils ne profitent pas de l'observation de leurs congénères experts, et font beaucoup d'erreurs après un succès. Les chimpanzés, à l'inverse, même s'ils mettent davantage de temps à découvrir la solution, non seulement sont capables de la mémoriser en une fois, mais surtout parviennent à imiter : ils copient la solution à partir de l'observation de leurs congénères plus habiles.

On peut penser que cette capacité d'imitation provient d'une détection. Ils détectent apparemment l'opération *mettre bout à bout*, et ils semblent reconnaître dans les deux baguettes mises bout à bout la forme attendue *baguette assez longue*. Il serait intéressant de proposer des modèles informatiques de cette performance basés sur l'appariement d'éléments comportementaux avec un plan d'action.

### 6.3. Théorie des "fusibles"

Le phénomène d'appariement peut être complexe et peut conduire à l'apprentissage de formes complexes. Noam Chomsky invoque un mécanisme de ce genre pour rendre compte de l'apprentissage par l'enfant de la compétence grammaticale dans telle ou telle langue [Chomsky 1975 ; Piatelli-Palmarini 1988 ; Berwick 1991 ; Crain 1991]. L'enfant disposerait de la capacité innée d'apprendre un certain ensemble de systèmes syntaxiques. L'exposition à quelques exemples de sa langue lui permettrait de fixer les valeurs de certains paramètres qualitatifs parmi les valeurs qui lui sont accessibles de manière innée.

Ceci revient à considérer que l'enfant possède de manière innée une structure très riche, qui se trouve en quelque sorte "appauvrie" par l'exposition à langue maternelle. L'apprentissage se traduit dans ce cas par une élimination de potentialités [Mehler 1974]. Comment peut-on alors expliquer que les langues réelles diffèrent tant, si elles dérivent toutes d'une même structure innée ?

La réponse réside en partie dans la combinatoire des paramètres susceptibles d'être fixés par l'enfant qui apprend sa langue maternelle. Pour rendre ce phénomène plus compréhensible, nous avons développé un exemple "jouet" dans lequel une grammaire éventuellement complexe est apprise à partir de quelques exemples de la langue à apprendre et d'une méta-grammaire innée.

Nous considérons la langue formée par l'ensemble des mots :

$$a^{k.n} b^{k.m}$$

où les valeurs de  $m$  et  $n$  sont fixes et déterminent un dialecte particulier, et  $k$  varie pour engendrer les mots de ce dialecte. Il n'est pas possible d'engendrer cette langue à l'aide d'une grammaire finie. En revanche, chaque dialecte peut être engendré par la grammaire très simple :

$$S \rightarrow a^n S b^m$$

$$S \rightarrow .$$

De plus, chacune de ces grammaires peut être engendrée par une méta-grammaire (les signes et symboles non terminaux de cette méta-grammaire sont soulignés) :

$$\underline{\mathbf{R}} \Rightarrow S \rightarrow \underline{\mathbf{G}} S \underline{\mathbf{D}}$$

$$\underline{\mathbf{R}} \Rightarrow S \rightarrow .$$

$$\underline{\mathbf{G}} \Rightarrow a \underline{\mathbf{G}}$$

$$\underline{\mathbf{G}} \Rightarrow .$$

$$\underline{\mathbf{D}} \Rightarrow b \underline{\mathbf{D}}$$

$$\underline{\mathbf{D}} \Rightarrow .$$

Cette méta-grammaire engendre les grammaires qui reconnaissent les dialectes et les réunions de dialectes<sup>(4)</sup>. Le principe de la théorie des "fusibles" consiste à appliquer la méta-grammaire pour reconnaître indirectement des exemples : en présence d'un ensemble de quelques mots du dialecte à apprendre, la méta-grammaire est mise en oeuvre, elle engendre des grammaires jusqu'à qu'une grammaire reconnaisse correctement les mots présents.

Dans le contexte de l'apprentissage par "fusibles", la méta-grammaire correspond à une connaissance *innée*. L'application de certaines règles de la méta-grammaire, dans un certain ordre et un certain nombre de fois, correspond à un ensemble de choix qualitatifs. Si on considère ces choix comme *irréversibles* (une fois qu'ils ont conduit à une grammaire efficace qui reconnaît les exemples fournis, la grammaire ne changera plus), la situation rappelle celle de la fusion de connexions dans un circuit matriciel. Nous préférons cette analogie avec des fusibles, plutôt que la dénomination habituelle de "fixation de paramètres", afin de bien rappeler que les paramètres en question sont qualitatifs.

Par exemple, on peut imaginer que la présentation des mots  $a^4b^8$ ,  $a^6b^{12}$  déclenche, éventuellement à la suite de quelques essais, la création de la grammaire correspondant à  $(m,n) = (2,4)$ . Effectivement, cette grammaire reconnaît les exemples présentés. Le système de méta-grammaire est ainsi capable d'apprendre le dialecte infini  $\{a^{2k}b^{4k} ; k \in \mathbb{N}\}$  à partir de seulement deux exemples, et ceci parmi une infinité de dialectes possibles !

Malheureusement, l'apprentissage n'est pas unique. Le problème vient ici de ce que le couple (1,2) convient aussi pour rendre compte des deux exemples présentés. Il en est de même des grammaires qui résulteraient de l'adjonction de règles inutiles. On peut imaginer deux solutions :

- une méta-grammaire moins générale (par ex. avec la règle  $\underline{\mathbf{G}} \rightarrow a\underline{\mathbf{G}}$  à la place de la 3<sup>ème</sup> règle) qui exclut le cas (1,2)
- une phase de génération, au cours de laquelle le système apprenant tente de produire des mots du dialecte tel qu'il l'a analysé. Si le dialecte correct, dans l'exemple ci-dessus, correspond à (2,4), le système qui a engendré la grammaire (1,2) produira des mots erronés comme  $ab^2$  ou  $a^3b^6$ , et sera amené à rechercher une grammaire plus restrictive.

On comprend ainsi comment l'existence d'une structure innée confère une efficacité impressionnante à l'apprentissage. Quelques exemples suffisent à déterminer toute une langue. Ces exemples sont en quelque sorte "attendus" (par opposition à des chaînes comme *bbabbaa* qui ne peuvent qu'être ignorées), ils sont *appariés* avec la structure innée, et de cet appariement résulte la fixation d'un certain nombre de paramètres qualitatifs (ici  $m$  et  $n$ ) qui caractérisent après coup la langue apprise.

Ce modèle permet aussi de comprendre comment la combinatoire des options successives (dans notre exemple, il s'agit des règles de la méta-grammaire qui ont été déclenchées) peut expliquer non seulement la diversité des langues naturelles, mais aussi le caractère fortement typé et non aléatoire des *erreurs* commises par les enfants qui apprennent leur langue maternelle.

Dans notre exemple, la grammaire incorrecte (1,2) produira des erreurs comme  $ab^2$ ,  $a^3b^6$ ,  $a^5b^{10}$ . L'ensemble de ces erreurs est fortement structuré, même si sa structure est erronée. Cette structure des erreurs est caractéristique d'un apprentissage par "fusibles", par opposition à des erreurs qui seraient produites par d'autres mécanismes d'apprentissage comme une recherche aléatoire ou une optimisation (ici, il n'y a pas de topologie naturelle dans l'espace des performances).

On observe le même phénomène pour les erreurs linguistiques des enfants. Par exemple les enfants surgénéralisent des règles. Certes, Plunkett et Marchman [1990] suggèrent que, dans certains cas, une extraction de régularités abusive (qu'ils modélisent à l'aide d'un réseau connexionniste) peut rendre compte de ce type d'erreur. Pourtant, Piatelli-Palmarini [1988] note que la typologie des erreurs issues de mécanismes statistiques ou d'une technique de résolution de problème diffère fondamentalement de ce que

---

<sup>4</sup> Il faut s'assurer que la deuxième règle de la méta-grammaire est déclenchée, sinon on obtient des grammaires non fonctionnelles. Il serait facile de remédier à ce défaut, en introduisant un séparateur de règles dans les métarègles.

l'on peut observer chez les enfants. Les erreurs observées seraient beaucoup mieux expliquées, selon lui, par un mécanisme d'apprentissage par "fusibles".

L'apprentissage par "fusibles" est donc une version complexe de l'apprentissage par appariement, au même titre que l'apprentissage par détection, empreinte ou imitation. Les possibilités d'appariement, dans le cas de l'apprentissage des connaissances conceptuelles, seront discutées dans le chapitre II.3.

Commentaire sur l'apprentissage par appariement : Le mécanisme d'appariement rend parfaitement compte de la relative diversité des connaissances apprises, mais son intérêt principal est d'expliquer l'étonnante efficacité de certains cas d'apprentissage (apprentissages en un coup). Cet aspect, qui est fondamental pour notre approche, fera l'objet des chapitres II.2 et II.3.

## 7. Constructivisme de Jean Piaget

Jean Piaget, à l'occasion de ses études du développement cognitif du jeune enfant, s'est opposé à la vision behavioriste de l'apprentissage. Il a montré par de nombreuses expériences à quel point rien n'était passif dans l'acquisition de nouvelles aptitudes, et que l'enfant confronté à une situation nouvelle formait des hypothèses qu'il confrontait ensuite plus ou moins bien avec la réalité.

L'apport principal de Piaget se situe sans doute dans la mise en évidence de stades bien identifiés par lesquels tout enfant passe successivement et dans l'ordre. Selon Piaget, l'enfant *construit* sa compétence, chaque stade correspondant à des états fonctionnels de cette compétence. Le stade final, qui est stable, doit sa stabilité aux contraintes que nous impose la réalité, et notamment certaines symétries inhérentes au domaine considéré.

Nous aurons l'occasion de revenir sur le point de vue Piagétien dans le chapitre II.2.

Exemple : L'enfant comprend pourquoi l'eau versée du verre large dans le verre étroit a gardé le même volume dès qu'il perçoit la possibilité d'inverser l'action [Piaget & Inhelder 1947]. Le concept élémentaire de justice (et d'injustice) n'est correctement "construit" que vers l'âge de 10 ans, avec la découverte de la réciprocité [Piaget 1932].

Commentaire sur le constructivisme : La théorie de Piaget, appliquée à l'apprentissage des connaissances, concerne surtout les grandes acquisitions comme les principes de conservation ou de réversibilité. Elle ne rend pas compte des connaissances mémorisées quotidiennement.

## 8. Analogie, instanciation, généralisation

L'apprentissage des connaissances conceptuelles peut s'opérer par des opérations sur des connaissances déjà possédées. Les mécanismes que l'on peut imaginer immédiatement sont ceux prévus par un formalisme logique : conjonction, disjonction, négation, instanciation, généralisation, analogie, etc. Ces mécanismes ont été bien étudiés en psychologie et en intelligence artificielle. Ils constituent des moyens élémentaires pour former de nouvelles connaissances, mais nous ne connaissons pas de théorie utilisant ces éléments qui permette de prévoir quelle connaissance sera formée et mémorisée dans un contexte concret particulier. Notre objectif est d'aborder cette question avec l'éclairage que nous fournit l'étude des conversations.

## 9. Classification des modes d'apprentissage

Nous venons de passer en revue un certain nombre de modèles d'apprentissage qui peuvent être invoqués pour rendre compte de l'apprentissage des connaissances. Il en existe bien entendu d'autres, mais nous avons voulu mentionner les plus caractéristiques de notre point de vue.

Nous allons proposer ici deux critères pour distinguer les différents mécanismes d'apprentissage. Le premier concerne un aspect que nous n'avons presque pas évoqué. Il s'agit des conditions de l'apprentissage, de ce qui permet au mécanisme de déclencher la mémorisation d'une forme.

Le deuxième critère sera largement développé dans le chapitre II.2. Il s'agit du critère d'*indifférence*, qui permet de quantifier la connaissance *a priori* que le système possède sur ce qu'il doit apprendre.

### 9.1. Extraction des formes à apprendre

Le problème fondamental qui se pose à tout mécanisme d'apprentissage consiste à extraire, des situations rencontrées, la forme qui sera finalement mémorisée. Quatre solutions semblent être employées :

- ❶ mémoriser les formes fréquentes
- ❷ utiliser une information de supervision
- ❸ mémoriser les formes symétriques
- ❹ mémoriser, parmi les formes attendues, toutes celles qui se présentent

Ces différentes solutions semblent être combinées dans les mécanismes d'apprentissage que nous avons évoqués. Voyons comment.

#### 9.1.1. Extraction par seuil de corrélation

Nous pouvons tout d'abord constater qu'un certain nombre de mécanismes effectuent l'équivalent d'un comptage des situations pour retenir celles qui sont les plus fréquentes. Nous pouvons les considérer comme des mécanismes statistiques, dans la mesure où une configuration doit être répétée pour être apprise. Ce sont les mécanismes que nous avons rangés sous la rubrique "Extraction de régularités" : conditionnement classique, conditionnement opérant, connexionnisme, mécanismes sélectifs, ACP et nuées dynamiques, ainsi que les mécanismes de compilation.

Le nombre de mesures nécessaire pour que la régularité soit extraite sera d'autant plus grand que la régularité sera faiblement marquée. Pour caractériser ce phénomène, supposons que le système apprenant dispose de  $n$  entrées  $x_1, x_2, \dots, x_n$ . Supposons aussi qu'une régularité existe entre  $p$  de ces entrées :  $x_{k_1}, x_{k_2}, \dots, x_{k_p}$ . La régularité sera extraite si l'estimation de la "corrélation"  $\rho$  entre ces variables dépasse un seuil  $\xi$  fixé à l'avance, avec une erreur relative  $e_r$  faible par rapport à  $\rho$ . On montre dans l'annexe 3 du chapitre II.2 que  $e_r$  est de l'ordre de  $1/(\rho \cdot \sqrt{m})$ , où  $m$  est le nombre de mesures effectuées. On comprend ainsi pourquoi, lorsque le seuil  $\xi$  est faible,  $m$  doit être grand lorsque  $\rho$  est de l'ordre de  $\xi$ .

Si la corrélation est très faible et que la fréquence des mesures est faible elle aussi, alors le temps nécessaire pour extraire la régularité peut être très long.

Dans le cas des organismes vivants, il peut dépasser la durée de la vie ! Dans ce cas, c'est au niveau du pool génétique que l'extraction s'opère. Ce n'est plus l'individu qui apprend, mais l'ensemble des gènes de la population<sup>(5)</sup>.

Les systèmes sélectifs, les systèmes connexionnistes non supervisés et les algorithmes exploratoires de classification reposent sur ce principe de seuil de corrélation.

#### 9.1.2. L'information de supervision

La durée nécessaire pour extraire une corrélation peut se révéler prohibitive, par exemple dans les situations où la corrélation est très pertinente (e.g. pour la survie du système apprenant) ou a une existence

---

<sup>5</sup> Un herbivore vivant en troupeau a peu d'occasions pour mesurer la corrélation entre l'attaque d'un prédateur et sa propre survie. Dans ce cas extrême, le nombre de mesures "positives" au cours de la vie de l'individu est inférieur à un. Seul le pool génétique peut extraire une corrélation entre l'attaque du prédateur et la fréquence, dans la population d'herbivores, de certains gènes comme ceux qui gouvernent les conditions de la fuite.

limitée dans le temps. Le système peut utiliser dans ce cas une information de *supervision* : il accorde (de manière innée) à certaines de ses entrées un statut spécial. La mémorisation n'a lieu que si ces entrées de supervision appartiennent à un certain domaine.

Par exemple, le phénomène d'*empreinte* (voir p. 97) subordonne la mémorisation de l'aspect de la "mère" à la présence de certains stimulus particuliers. Dans ce cas extrême, tout se passe comme si l'oisillon faisait l'hypothèse *a priori* de la corrélation entre la présence de sa mère et celle de ces stimulus. Il n'a pas besoin d'extraire cette corrélation. Il n'en aurait pas le temps<sup>(6)</sup>.

Dans le conditionnement, nous avons affaire à une situation intermédiaire. Dans le conditionnement "opérant", par exemple, seules les situations associées à une sensation marquée de plaisir / déplaisir sont prises en considération. Cette sensation constitue donc bien une entrée de supervision, mais la mémorisation n'est pas pour autant immédiate. Elle nécessite qu'un certain seuil de corrélation soit atteint, sauf quand le (dé)plaisir est intense. Dans le cas du conditionnement opérant, l'information de supervision sert donc à abaisser le seuil de corrélation à partir duquel la mémorisation se produit.

### 9.1.3. Mémorisation des formes symétriques

Le constructivisme, mais aussi la théorie de la Gestalt, invoquent un autre critère qu'un seuil de corrélation pour décider qu'une forme sera mémorisée. Dans ces modèles, les formes apprises jouissent nécessairement de propriétés de symétrie. Nous aurons l'occasion de revenir sur cet aspect dans le chapitre II.2, en montrant comment, dans le cadre du constructivisme, les propriétés de fermeture opérationnelle des structures construites peuvent être traduites en termes de symétrie.

### 9.1.4. Répertoire limité et appariement

La dernière possibilité que nous avons envisagée pour guider la mémorisation est caractéristique de l'apprentissage par "fusibles", et plus généralement de l'apprentissage par appariement.

Dans le contexte de l'apprentissage par "fusibles", le système apprenant ne peut reconnaître qu'un nombre relativement limité de formes. On peut considérer que la mémorisation est déclenchée chaque fois que l'appariement a pu se produire.

Dans le cas de l'apprentissage des connaissances, nous verrons au chapitre II.3 que l'appariement est presque toujours possible. La mémorisation n'est alors déclenchée, dans le modèle que nous proposons au chapitre II.4, que lorsqu'une situation problématique est rencontrée.

## 9.2. Indifférence des mécanismes d'apprentissage

Dans les chapitres II.2 et II.3, nous ferons une autre distinction, essentielle à nos yeux. Nous opposerons les mécanismes *indifférents* aux mécanismes faisant appel à un appariement structuré.

Un mécanisme est indifférent si son fonctionnement est indépendant d'un changement de repère dans l'espace de ses entrées (voir définition chap. II.2). Par exemple, une ACP fonctionnera de manière totalement identique si l'on opère une rotation de l'ensemble des individus dans l'espace des caractéristiques ; les axes principaux auront simplement tourné.

Les mécanismes d'appariement sont généralement non-indifférents. Dans l'apprentissage "par fusibles", par exemple, un changement dans la représentation des exemples risque de diminuer ou d'augmenter la qualité de l'appariement, jusqu'à le supprimer ou créer des sauts qualitatifs. Dans le système décrit p.98, une permutation systématique de *a* et de *b* dans tous les mots-exemples va totalement supprimer l'appariement, car le système ne peut engendrer aucune grammaire qui reconnaîtra un mot contenant des *b* suivis de *a*.

Nous utiliserons cette propriété d'indifférence pour qualifier la connaissance *a priori* que le mécanisme possède de son environnement.

Nous pouvons tenter de classer selon ce critère *indifférent / non-indifférent* les mécanismes mentionnés dans ce chapitre. Le tableau suivant est donné à titre simplement indicatif, car sa justification détaillée nous ferait sortir du cadre de cet exposé. Il pourra être utile au lecteur du chapitre II.2 qui voudrait estimer par lui-même le caractère indifférent des modèles d'apprentissage.

<sup>6</sup> En fait, cette corrélation a été extraite au niveau phylogénétique.

<i>modèle d'apprentissage</i>	<i>élément appris</i>	<i>entrée</i>	<i>indifférence</i>	<i>non-indifférence</i>
Empirisme ou Instruction	Idées	Sensations	Oui	
Recherche aléatoire	Extremum	Couples (x,f(x))	Translation	
Méthodes de descente	Extremum	Couples (x,f(x))	Translation	
Optim. par mutation-sélection	Extremum	Couples (x,f(x))	Translation	
Conditionnement classique	Couplage SC-Action	SI, SC, Situation	Changement de SC (en théorie)	SI non permutable avec SC
Conditionnement opérant	Couplage Situation- Récompense- Action	Situation, Récompense	Changement systématique des situations récompensées (en théorie)	Récompense non permutable avec un élément quelconque de la situation. Translation sur Réc
Connexionnisme (rétrop.)	Couplage $E_i/S_j$	Patterns d'entrée Patterns d'erreur	Permutation dans les patterns d'entrée et d'erreur	Permutation entre des entrées et des signaux d'erreur
Connexionnisme (Kohonen)	Couplage topologique	Patterns d'entrée	Permutation dans les patterns d'entrée	
Extraction de régularités par des systèmes sélectifs (alg. gén.)	Combinaison de gènes	Evaluation des phénotypes	Certains changements systématiques de l'évaluation	Dépendance par rapport au codage
Partitionnement par nuées dynamiques	Classes, Représentants typiques	Individus	Translation, rotation des individus	
Analyse en composantes principales (ACP)	Axes principaux	Individus	Translation, rotation des individus	
Partitionnement par similarités	Classes	Individus	Translation, rotation des individus	
Apprentissage par compilation	Action			
Innéisme - maturation	Nouvelles facultés			
Marienbad (contre joueur externe)	Configurations favorables	Actions du joueur		Qualité du jeu adverse
Caractron	Caractère	Caractère en binaire	Changement du caractère modèle	
Empreinte (oie)	Forme de la mère	Images mobiles		Détecteur "accordé" sur certaines situations
Détection - appariement (pattern-matching)	Activité du détecteur	Situation		Détecteur "accordé" sur certaines situations
Imitation	Comportement complexe	Comportement observé		Comportement observé analysable
Théorie "des fusibles"	Valeurs de paramètres qualitatifs	Exemples		Exemples corrects et discriminants
Constructivisme	Structure opératoire fermée	Situations opératoires	Changements d'échelle (translation log.)	

On constate que les mécanismes statistiques, qui extraient des régularités dans l'environnement, sont généralement de type indifférent. A cet égard, la distinction entre les mécanismes indifférents et l'apprentissage par appariement rejoint une distinction faite en statistique, entre les méthodes non-paramétriques et les méthodes paramétriques. Par exemple, l'ajustement (appariement) d'une série chronologique à une fonction polynomiale de degré donné est une méthode paramétrique, alors que l'ACP est une méthode non-paramétrique.

## **10. Conclusion**

Nous montrerons dans le chapitre II.2 que les mécanismes indifférents ne peuvent pas rendre compte de l'apprentissage des connaissances comme celles qui sont manipulées dans les conversations. Notre objectif est de suggérer que l'apprentissage des connaissances conceptuelles, comme celui qui opère dans les conversations, s'effectue par des mécanismes d'appariement, que le déclenchement de ces mécanismes ne doit rien à la répétition des situations.

# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

Dans le chapitre II.1, nous avons décrit un certain nombre de modèles d'apprentissage, en rappelant que la plupart d'entre eux ont été invoqués par divers auteurs pour expliquer l'acquisition des connaissances du type de celles que les interlocuteurs manipulent au cours des conversations. Notre objectif ici est de montrer qu'un certain nombre de ces modèles de l'apprentissage possèdent une propriété, que nous appelons propriété d'*indifférence*, qui certes est désirable dans certains contextes, mais qui les disqualifie lorsqu'il s'agit d'apprentissage de connaissances de type conversationnel.

L'indifférence caractérise les systèmes qui fonctionnent indépendamment des caractéristiques *absolues* de leurs entrées. Les mécanismes d'extraction de régularités sont des mécanismes indifférents : ils n'utilisent que les propriétés *relatives* des objets (ressemblances). En revanche, un détecteur, qui repère une configuration particulière de ses entrées, est par essence non-indifférent : le résultat de son fonctionnement dépend de manière cruciale d'une configuration absolue, avec laquelle il compare ses entrées.

Les aspects non-indifférents d'un système doivent nécessairement exister *a priori*, c.à.d. être antérieurs à toute expérience. C'est l'une des raisons pour lesquelles on donne la préférence aux systèmes indifférents dans toute modélisation. Pourtant, nous montrons que les systèmes indifférents ne permettent pas de modéliser l'apprentissage des connaissances conversationnelles.

Dans un premier temps, nous définissons l'indifférence, et nous l'illustrons par des exemples. Puis nous montrons que les mécanismes indifférents sont efficaces pour l'apprentissage de formes fortement "symétriques", et inefficaces dans les autres cas.

Pour cela, nous estimons le nombre minimal de données dont un système indifférent a besoin pour parvenir à un apprentissage correct dans une situation donnée. Ce nombre devient d'autant plus grand que la forme à apprendre manque de "symétrie".

Ayant constaté cela, nous montrons que les connaissances nécessaires pour soutenir une conversation sur un sujet donné (par ex. pour juger de la cohérence d'une situation) correspondent à des formes totalement dépourvues de symétrie. Nous en concluons que les mécanismes d'apprentissage indifférents sont de mauvais modèles de l'apprentissage des connaissances.

Nous en tirons argument pour suggérer, dans le chapitre II.3, un mode d'apprentissage non-indifférent des connaissances utilisées dans les conversations.

## 2. Indifférence d'un mécanisme d'apprentissage

Le concept central, qui permet de caractériser l'absence de "connaissance" spécifique *a priori* dans la constitution du système apprenant, est le concept d'*indifférence*. Pour savoir si un système est indifférent, on opère des changements systématiques dans les entrées du système (permutations, changements d'origine) de manière à observer si le système se comporte différemment.

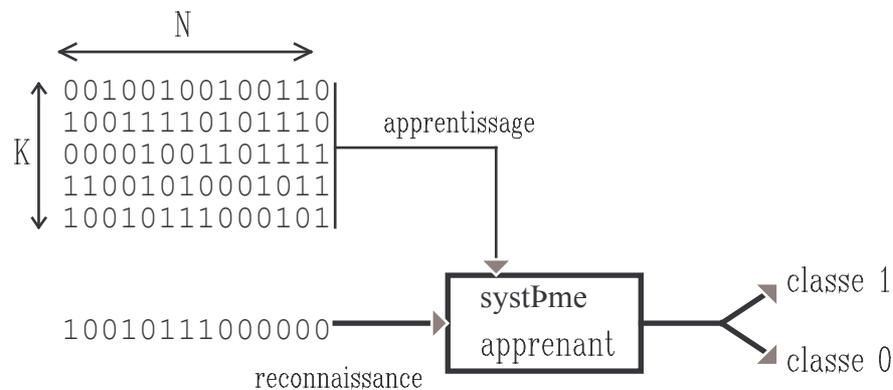
Si, par exemple, un système qui comporte deux capteurs n'a pas le même comportement lorsqu'on effectue (après réinitialisation) une permutation systématique des signaux qui parviennent à ces capteurs, c'est que ce système, de par sa constitution même, fait une différence *a priori* entre ces deux capteurs.

Imaginons maintenant un système muni d'une caméra, capable d'apprendre à reconnaître des chiffres manuscrits. Si l'on observe que l'apprentissage est plus difficile quand la caméra est tournée de 30°, on peut en conclure que la verticalité est connue *a priori* du système, et que ce système n'est pas indifférent en raison de son anisotropie.

Le concept d'indifférence que nous allons définir est valable pour une classe très large de systèmes de traitement de l'information, c.à.d. de systèmes soumis à des entrées et dont on peut observer un état. Un tel traitement peut être analogique ou digital. Toutefois, afin de pouvoir quantifier nos résultats, nous donnerons toutes nos définitions dans le cas d'un apprentissage de classification binaire. Les résultats que nous obtiendrons pourront être transposés sans difficultés à des situations réelles.

### 2.1. Exemple d'apprentissage : une classification binaire

Pour donner un sens précis à nos définitions et illustrer la différence entre certaines théories de l'apprentissage, nous nous placerons dans le cadre suivant, où il s'agit d'apprendre à distinguer deux classes à partir d'un jeu d'exemples :



Le système apprenant est en contact avec l'environnement par  $N$  bascules à deux états. Son expérience consiste en une présentation d'un *jeu de  $K$  exemples* de taille  $N$  bits. Ces exemples contiennent une éventuelle information de supervision. Grâce à eux, le système construit une classification qui lui permet de ranger tout  $N$ -uplet binaire dans la classe notée 1 ou dans la classe notée 0.

Nous nous plaçons dans une situation qui comporte deux phases. La phase d'apprentissage, durant laquelle le système est exposé aux  $K$  exemples, puis la phase de reconnaissance. Lors de l'opération de reconnaissance, le système est capable de distinguer les  $2^N$  entrées différentes (ensemble  $E$  des exemples ou données), et il peut assigner une classe à chacune d'entre elles. Le résultat de l'apprentissage est donc une *classification*.

Chaque classification est une partition de l'hypercube  $E$  de dimension  $N$  en deux classes, et elle peut être décrite par une application de l'ensemble  $E$  dans  $\{0, 1\}$ . Il y a donc  $2^{2^N}$  manières de classifier les éléments de  $E$ . C'est l'expérience des  $K$  exemples (équivalents à  $K.N$  bits) et le fonctionnement interne du système qui lui permettent de se fixer sur l'une de ces classifications. Le *mécanisme d'apprentissage*  $\mathcal{A}$  est ainsi une application de l'ensemble des jeux d'exemples dans l'ensemble des classifications.

Nous supposons ici, pour simplifier, qu'à la suite de son apprentissage, le système a correctement appris, c'est-à-dire qu'il classe sans erreur. Mais les raisonnements restent valables si l'on tolère un taux d'échec donné.

Nous allons chercher à caractériser un tel système en fonction de l'ensemble des classifications qu'il peut atteindre, et de la taille  $K.N$  de l'ensemble d'apprentissage que cela nécessite. L'objectif est de montrer que certains apprentissages sont inaccessibles à certains mécanismes, car ils nécessiteraient trop d'exemples.

## 2.2. Définition de l'indifférence

**Mécanisme d'apprentissage indifférent** : un mécanisme d'apprentissage est indifférent (aux isométries) si le fait de faire subir un déplacement isométrique aux exemples (changement de repère) conduit à une classification qui reconnaît les données déplacées.

Cela signifie que le traitement n'utilise au plus que les distances relatives entre les données. Dans notre exemple, l'isométrie dans l'hypercube est à prendre au sens de la distance de Hamming.

Plus précisément, appelons  $\mathcal{A}$  le mécanisme d'apprentissage.  $\mathcal{A}$  est *indifférent* si on peut écrire pour toute isométrie  $\rho$  de  $E$  dans  $E$  et pour tout jeu d'exemples  $J$  :

$$\mathcal{A}(J) = [\mathcal{A} \circ \rho(J)] \circ \rho$$

équation d'indifférence

Ici,  $\rho$  (en car. gras) désigne l'opération qui à un jeu d'exemples associe le jeu des exemples ayant subi le déplacement  $\rho$  :  $\rho(J) = \{\rho(x) \mid x \in J\}$ . Le tableau suivant rappelle les ensembles sur lesquels opèrent les fonctions mentionnées ici.

$\rho$ :	{exemple}	→	{exemple}	(isométrie de $E$ dans $E$ )
$\rho$ :	{jeu d'exemples}	→	{jeu d'exemples}	(isométrie $\rho$ , mais de $\rho(E)$ dans $\rho(E)$ )
$\mathcal{A}$ :	{jeu d'exemples}	→	{classification}	(mécanisme d'apprentissage)
$\mathcal{A}(J)$ :	{exemple}	→	{0,1}	(classification)

La signification de cette équation d'indifférence correspond à celle d'un changement de repère. Un système indifférent se comporte de façon rigoureusement identique si on applique le changement de repère  $\rho$  (aussi bien aux exemples qui servent à l'apprentissage qu'aux données à reconnaître). Noter que la formule précédente peut être aussi écrite de la manière suivante :

$$\forall x \in E ; \quad \mathcal{A}(\rho(J))(x) = \mathcal{A}(J)(\rho^{-1}(x))$$

Toute isométrie au sens de la distance de Hamming dans l'hypercube est la composée d'une translation et d'une permutation, et réciproquement (voir annexe n°1). Comme il y a  $2^N$  translations possibles et  $N!$  permutations, on en déduit qu'il y a  $2^N \cdot N!$  isométries différentes de  $E$  dans  $E$ . Nous pouvons donner les définitions suivantes :

**Traitement isotrope** : Nous dirons qu'un système est *isotrope* s'il est indifférent aux permutations des coordonnées.

**Traitement relatif** : Nous dirons de même qu'un système est *relatif* s'il est indifférent aux translations.

En particulier, un système relatif est indifférent à une permutation dans l'alphabet binaire (complémentation), car celle-ci correspond à une translation selon le vecteur "diagonal" (1 1 ... 1). Toute translation revient à inverser les valeurs 0 et 1 sur certaines coordonnées des exemples.

Grâce au résultat de l'annexe n°1, on voit aisément qu'un système isotrope et relatif est indifférent aux isométries.

Il est possible de définir la propriété d'indifférence pour d'autres classes de transformations que celle des isométries. Sans mention explicite, "indifférent" signifiera "indifférent aux isométries".

### 2.3. Exemples de mécanismes d'apprentissage indifférents

Nous pouvons donner facilement des exemples et des contre-exemples de mécanismes indifférents en restant dans le cadre de l'apprentissage d'une classification binaire défini p. 106.

Un mécanisme d'apprentissage par coeur est un mécanisme indifférent.

Pour un tel mécanisme, la classe n°1 est exactement le jeu d'exemples  $J$ . Pour appliquer l'équation d'indifférence, on applique une transformation systématique  $\rho$  à chaque élément de  $J$  (permutation des coordonnées ou complémentations de certains bits), ce qui donne  $J_I$ . L'apprentissage par coeur à partir de  $J_I$  définit  $J_I$  comme la nouvelle classe n°1. On constate qu'une donnée transformée  $\rho(x)$  est dans cette nouvelle classe n°1 (i.e.  $\rho(x) \in J_I$ ) si et seulement si la donnée initiale  $x$  était la classe n°1 de départ (i.e.  $x \in J$ ).

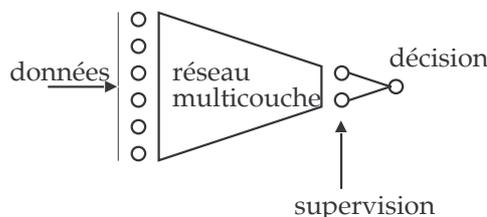
Un autre mécanisme manifestement indifférent est le mécanisme trivial tel que  $\mathcal{A}(J) = \mathbf{1}$ , où  $\mathbf{1}$  désigne la classification indépendante du jeu d'exemples  $J$  qui classe toute donnée dans la classe n°1.

Parmi les exemples plus intéressants, on peut mentionner les mécanismes d'apprentissage comme les algorithmes de classification statistiques courants ou les architectures connexionnistes courantes.

Si l'on applique aux données du jeu d'exemples un algorithme de nuées dynamiques avec deux centres mobiles qui utilise la distance de Hamming dans l'hypercube, on obtient une règle de classification. Comme cet algorithme ne fait jamais référence aux coordonnées absolues des points et qu'une isométrie dans l'hypercube conserve les distances relatives, on conclut aisément que ce mécanisme est indifférent.

L'architecture d'un réseau de Kohonen [Kohonen 1984] rend indiscernables les unités d'entrée. En effet, pour une cellule donnée, le câblage afférent et le câblage interne à la couche sont totalement indépendants des coordonnées de la cellule<sup>1</sup>. De même, l'algorithme d'apprentissage ne mentionne jamais les coordonnées absolues des cellules ni les valeurs absolues des excitations. Il ne fait que des comparaisons de produits scalaires, et comme une isométrie conserve le produit scalaire, l'algorithme est indifférent aux isométries.

Il en est à peu près de même pour les architectures multi-couches généralement utilisées. Pour rester dans le cadre de notre exemple d'une classification binaire, la dernière couche pourra par exemple comporter deux cellules, plus un mécanisme de décision qui détermine la cellule la plus active.



Un réseau multi-couches supervisé peut être considéré comme un système quasi indifférent.

<sup>1</sup> Ceci n'est rigoureusement vrai que si l'on ne tient pas compte des effets de bords (les unités au bord de la couche de Kohonen ayant moins de voisins) ou si l'on travaille avec une couche torique.

En effet, le câblage des unités ne dépend pas, généralement, de leur position au sein de la couche d'entrée ou au sein de la couche de supervision. Cela est vrai aussi de l'algorithme de rétro-propagation. Le mécanisme est en ce sens isotrope. Toutefois, dans notre contexte, les unités qui reçoivent l'information de supervision sont à considérer comme des unités d'entrée au même titre que celles qui reçoivent les données proprement dites, alors qu'elles ne sont pas permutables avec ces dernières. Le comportement du système n'est donc pas isotrope par rapport à une permutation entre la couche d'entrée et la couche de supervision. En revanche, l'architecture et l'algorithme d'apprentissage sont totalement relatifs: rien ne dépend des valeurs absolues des exemples. Nous dirons d'un tel système qu'il est quasi-indifférent.

#### 2.4. Exemples de mécanismes d'apprentissage anisotropes ou non relatifs

Nous avons indiqué dans le chapitre II.1 que certains mécanismes importants ne possédaient pas la propriété d'indifférence. Nous pouvons donner ici quelques exemples de mécanismes d'acquisition de classification binaire qui soient anisotropes.

Un premier exemple nous est fourni par le mécanisme trivial  $\mathcal{A}$  tel que  $\mathcal{A}(J) = C$ , où  $C$  désigne une classification particulière, indépendante du jeu d'exemples  $J$ . Si les classes définies par  $C$  dans l'hypercube ne sont pas invariantes pour toutes les permutations des coordonnées, alors le mécanisme n'est pas isotrope.

D'une manière générale, tout mécanisme capable de discerner les bits d'entrée sera généralement anisotrope. Considérons par exemple le mécanisme qui considère ses entrées comme des entiers codés sur  $N$  bits, puis calcule leur centre de gravité  $G$  et classe les données selon leur position par rapport à  $G$ . La permutation de deux bits en entrée va modifier sensiblement la position relative des points, et la nouvelle classification obtenue ne classera pas les données permutées comme la classification initiale classait les données non permutées.

jeu d'exemples	entiers associés	centre de gravité	classes	exemples permutés	entiers associés	centre de gravité	classes
0101	5	7.2	0	0011	3	6.8	0
1100	12	7.2	1	1010	10	6.8	1
0011	3	7.2	0	0101	5	6.8	0
1001	9	7.2	1	1001	9	6.8	1
0111	7	7.2	0	0111	7	6.8	1

Dans le tableau ci-dessus, on constate que la permutation du deuxième et du troisième bit modifie le classement de 0111 (qui se trouve être invariant dans cette permutation). Le mécanisme n'est donc pas isotrope, et ceci vient ici du rôle différent assigné aux bits d'entrée dans la détermination de la classification.

Un mécanisme d'apprentissage peut être isotrope sans être relatif. Considérons le mécanisme qui produit le type de classification suivant : les données (mots de 32 bits) de la classe n°1 contiennent à peu près  $\alpha$  fois plus de 0 que de 1,  $\alpha$  étant un entier déterminé à partir du jeu d'exemples. Un tel mécanisme peut fonctionner sur un jeu d'exemples de taille 1. Il détermine le ratio du nombre de 0 sur le nombre de 1 dans l'exemple, il l'arrondit, ce qui détermine  $\alpha$  de manière irréversible :  $\alpha = \text{round}(N_0/N_1)$ . Il calcule le même ratio arrondi pour les données, et classe la donnée en classe n°1 si la valeur obtenue est égale à  $\alpha$ . Ce mécanisme est manifestement isotrope (une permutation des coordonnées est même sans effet), mais il n'est pas relatif.

Si l'apprentissage est effectué, par exemple, sur le mot suivant de 32 bits, 01001001000000010000110001000000, qui contient 7 bits à 1 et 25 à 0,  $\alpha$  se retrouve fixé à la valeur 4. Seront alors dans la classe n°1 les mots binaires qui comporteront à peu près 4 fois plus de 0 que de 1, c.à.d. les mots de 32 bits contenant 6 ou 7 bits à 1.

Ce mécanisme est sensible à la complémentation. Pour s'en convaincre, appliquons l'équation d'indifférence (p. 107). On commence par compléter l'unique exemple du jeu d'apprentissage, pour obtenir le vecteur 101101101111110111001110111111.  $\alpha$  prend alors la valeur 0, ce qui fournit une nouvelle classification. Or celle-ci devrait classer les mots binaires complémentés comme la classification  $\alpha = 4$  classait les mots avant complémentation. Il n'en est rien, comme le montre le tableau ci-dessous. Par exemple, un mot de 32 bits avec le

ratio 24/8 est dans la classe n°0 pour  $\alpha = 4$ , alors que son complément, de ratio 8/24, se retrouve dans la classe n°1 pour  $\alpha = 0$ .

$\alpha$	$(N_0/N_1)$	classe	$\alpha$	$(N_0/N_1)$	classe
4	25/7	1	0	0-15/17-32	1
4	26/6	1	0	16-32/0-16	0
4	0-24/8-32	0			
4	27-32/0-5	0			

Nous avons pu ainsi donner des exemples de mécanismes indifférents, et de mécanismes qui ne l'étaient pas, parce qu'ils étaient anisotropes ou non relatifs. Il existe toutefois des mécanismes qui sont "presque" indifférents.

### 2.5. Estimation du degré d'indifférence

Notre objectif dans la suite est d'opposer les mécanismes d'apprentissages indifférents et ceux qui ne le sont pas. En fait, de nombreux mécanismes, notamment les mécanismes supervisés (voir ci-dessus l'exemple des réseaux multi-couches) sont "presque" indifférents. Nous voulons tout de même les distinguer des mécanismes qui ne sont radicalement pas indifférents.

Il y a sans doute plusieurs moyens pour cela. Nous pouvons estimer la proportion des cas où les classifications  $\mathcal{A}(J)$  et  $[\mathcal{A} \circ \rho(J)] \circ \rho$  diffèrent. Ou plus simplement se contenter de compter parmi les  $2^{N_x N}$  isométries de  $E$  dans  $E$  (dénombrées dans l'annexe 1) la proportion de celles qui laissent le mécanisme indifférent.

Les exemples que nous considérerons dans la suite se rangeront clairement dans la catégorie des systèmes (quasi-)indifférents ou, à l'inverse, dans la catégorie de ceux qui sont très sensibles à une permutation ou à un changement des valeurs binaires.

Notons que le caractère stochastique de certains systèmes ne doit pas être confondu avec un manque d'indifférence. Si la détermination de la règle de classification par apprentissage à partir du jeu d'exemples  $J$  comporte une part d'aléatoire, il suffit d'écrire l'équation d'indifférence pour une même réalisation  $\omega$ :

$$\mathcal{A}(J, \omega) = [\mathcal{A}(\rho(J), \omega)] \circ \rho$$

Nous avons donné ces définitions dans le cadre assez restreint de l'apprentissage d'une classification binaire à partir d'un jeu d'exemples exprimés sur  $N$  bits. Nous sommes cependant convaincus que ces définitions peuvent être facilement étendues à des classes très larges de systèmes, y compris des systèmes analogiques.

## 3. Difficulté de l'apprentissage indifférent

### 3.1. Intérêt des systèmes apprenants indifférents

L'indifférence est généralement, et souvent à juste titre, considérée comme une qualité d'un mécanisme d'apprentissage. On préfère un système qui fait jouer *a priori* un rôle équivalent à ses entrées et qui ne tient compte que des valeurs relatives de ce qu'il perçoit.

Dans la conception de nombreux systèmes, on essaie de ne retenir que l'information de ressemblance entre les données, estimée par la mesure d'une distance (dans notre exemple, nous avons pris la distance de Hamming). Toute caractéristique de la constitution du système ou de son fonctionnement qui vient modifier la prise en compte des ressemblances est alors considérée comme un biais parasite. Par exemple un système qui ne peut traiter que des nombres (penser à une calculatrice de poche programmable) est au départ un mauvais système pour classer des données codées arbitrairement sous forme binaire, car il va assigner  $a$

*priori* un statut différent aux différents bits (penser au calcul du produit scalaire entre deux vecteurs binaires stockés sous forme de nombres dans la calculatrice).

De plus, l'objectif poursuivi dans la réalisation ou la modélisation de systèmes apprenants est souvent de tenter d'apprendre à partir de "rien", c'est-à-dire sans aucune information *a priori* sur les données autre qu'un moyen d'appréhender leur ressemblance. Toute information qui vise, par exemple, à privilégier certaines données, est considérée comme une sorte de tricherie. Tout ceci conduit à rechercher des systèmes apprenants indifférents.

Nous constaterons toutefois que les systèmes indifférents ont des limitations fondamentales. Ils n'apprennent facilement que les classifications fortement "symétriques", à moins d'utiliser des jeux d'apprentissage de taille très importante.

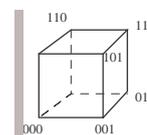
### 3.2. Apprentissages difficiles

Notre objectif est de montrer que certains apprentissages sont intrinsèquement *difficiles* pour les systèmes indifférents. Ce problème est bien connu dans le domaine du connexionnisme.

En effet, certaines tâches d'apprentissage sont réputées plus difficiles que d'autres pour les réseaux connexionnistes. [Ackley et al. 1985] mentionnent le problème du codage/décodage ou celui de la parité comme des problèmes difficiles du point de vue d'un réseau de Boltzman. Les réseaux connexionnistes sont aptes à repérer des "régularités" statistiques, c'est-à-dire des corrélations entre les entrées. Ils ont en revanche des difficultés à classer correctement les exceptions, qu'ils doivent, comme on dit, apprendre "par coeur".

Les réseaux multicouches, par exemple, isolent plus facilement des régions connexes de l'hypercube. L'activité d'une unité connectée aux neurones de la couche d'entrée (nous raisonnons pour des neurones à deux états) définit une région de l'hypercube d'entrée limitée par un hyperplan. L'activité combinée de plusieurs unités de ce type, détectée par des neurones des couches cachées, va définir une réunion de régions limitées par des hyperplans. On comprend que l'apprentissage de zones connexes de l'hypercube est plus facile, alors que l'apprentissage de données isolées va nécessiter l'intervention de nombreux neurones.

Par exemple, l'apprentissage de la parité (somme modulo 2 des bits d'entrée), qui consiste à séparer les données de parité paire et impaire, est difficile pour les réseaux multicouches, car deux données de la même classe ne sont jamais voisines. Chaque donnée est entourée à distance 1 de données appartenant à l'autre classe.



Pourtant, l'apprentissage de la parité peut être très facilement réalisé par un mécanisme indifférent. Il suffit d'imaginer un mécanisme qui apprend sur un exemple, en faisant la somme modulo 2 de ses bits (carré scalaire), puis qui classe en classe  $n^{\circ}1$  toute donnée dont le carré scalaire donne un résultat identique. Ce mécanisme est insensible à toute permutation des bits, et le classement de données translâtées restera identique si l'on fait apprendre le système sur l'exemple translâté.

Nous allons toutefois montrer que certains apprentissages sont difficiles pour les mécanismes indifférents (et donc *a fortiori* pour les réseaux connexionnistes habituels). Parmi eux nous trouverons un apprentissage qui nous concerne particulièrement, l'apprentissage des relations logiques.

### 3.3. Définir la difficulté d'apprentissage dans le cas général

Pour pouvoir estimer la difficulté d'une tâche d'apprentissage de classification pour un mécanisme donné, nous devons estimer le nombre  $\text{card}(C(K))$  de classifications différentes accessibles à ce mécanisme, avec :

$$C(K) = \{A(J) \mid J \in E^K\}$$

Le mécanisme que nous avons imaginé précédemment pour apprendre la parité ne peut accéder qu'à deux classifications différentes. Il apprenait avec un seul exemple. Un apprentissage "par coeur" peut en revanche accéder, au moins en principe, aux  $2^{2^N}$  classifications possibles. Mais il exige pour cela un ensemble d'exemples très vaste, puisque  $K$  (qui est supposé fixé avant tout apprentissage) doit alors valoir  $2^N$ .

**Difficulté d'apprentissage:** Nous estimerons la difficulté d'un apprentissage, pour un mécanisme  $\mathcal{A}$  donné, par la valeur minimale  $K_{min}$  de  $K$  (si elle existe) telle que  $C(K_{min})$  contienne la classification à apprendre.

En d'autre termes, si  $K_{min}$  existe, alors il existe un jeu de  $K_{min}$  exemples qui permet d'apprendre la classification demandée. La classification demandée est difficile à apprendre si  $K_{min}$  est grand.

Cette définition nous permet de donner une borne inférieure à la taille du jeu d'exemples. Celui-ci, qui regroupe  $K.N$  bits, doit être suffisamment grand pour au moins permettre de discriminer entre les  $C(K)$  classifications accessibles. On a donc au minimum:

$$K \geq \log_2(\text{card}(C(K)) / N)$$

En effet, le nombre  $2^{K.N}$  de jeux d'exemples différents doit être supérieur au nombre  $\text{card}(C(K))$  de classifications accessibles par le mécanisme d'apprentissage.

Noter que  $2^{K.N}$  est, dans la pratique, très supérieur à  $\text{card}(C(K))$ . En effet, la situation "optimale" où chaque jeu d'exemple conduirait à une classification différente est plutôt invraisemblable. Elle suppose que *chaque* bit de *chaque* exemple est nécessaire, lors de l'apprentissage, pour déterminer la classification (élément de  $C(K)$ ) qui doit être apprise.

Il est possible d'affiner la valeur de cette borne inférieure. Si nous supposons que l'ordre des exemples est sans effet sur le résultat de l'apprentissage, alors on peut montrer (voir annexe 2) que :

$$K.N - \log_2 K! + \frac{K^2}{2^{N-1}} \geq \log_2(\text{card}(C(K)))$$

Nous utilisons cette formule pour calculer la valeur maximale de  $\log_2(\text{card}(C(K)))$  (log. du nombre des classifications qui peuvent être apprises) en fonction de  $N$  (taille d'une donnée) et de  $K$  (nombre d'exemples). Ce nombre est à comparer à  $2^N$  qui est le logarithme du nombre total de classifications binaires (voir tableau). Comme on peut le constater, le fait que l'ordre des exemples soit non pertinent n'apporte pas de modification sensible du nombre minimal d'exemples pour les faibles valeurs de  $\text{card}(C(K))$ .

$N$	$K$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$2^N$
		$\log_2(\text{card}(C(K)))$																
8		8	15	21	27	33	39	44	49	54	58	63	67	71	76	80	84	256
12		12	23	33	43	53	63	72	81	90	98	107	115	123	132	140	148	4096
16		16	31	45	59	73	87	100	113	126	138	151	163	175	188	200	212	65536
20		20	39	57	75	93	111	128	145	162	178	195	211	227	244	260	276	$10^6$
24		24	47	69	91	113	135	156	177	198	218	239	259	279	300	320	340	$10^7$
28		28	55	81	107	133	159	184	209	234	258	283	307	331	356	380	404	$10^8$
32		32	63	93	123	153	183	212	241	270	298	327	355	383	412	440	468	$10^{10}$
36		36	71	105	139	173	207	240	273	306	338	371	403	435	468	500	532	$10^{11}$
40		40	79	117	155	193	231	268	305	342	378	415	451	487	524	560	596	$10^{12}$
44		44	87	129	171	213	255	296	337	378	418	459	499	539	580	620	660	$10^{13}$
48		48	95	141	187	233	279	324	369	414	458	503	547	591	636	680	724	$10^{14}$
52		52	103	153	203	253	303	352	401	450	498	547	595	643	692	740	788	$10^{16}$
56		56	111	165	219	273	327	380	433	486	538	591	643	695	748	800	852	$10^{17}$
60		60	119	177	235	293	351	408	465	522	578	635	691	747	804	860	916	$10^{18}$
64		64	127	189	251	313	375	436	497	558	618	679	739	799	860	920	980	$10^{19}$

Grâce à ce tableau, on peut estimer la borne inférieure  $K_{min}$  de la taille du jeu d'exemple pour réaliser un apprentissage donné. En effet, dans chaque situation d'apprentissage, on est généralement capable d'estimer  $\text{card}(C(K))$ . Grâce à ce tableau, on peut donc prévoir dans certains cas que l'apprentissage sera difficile.

Il faut noter que ce tableau est extrêmement "optimiste". Comme nous l'avons déjà noté, ce calcul suppose que chaque jeu d'exemples (ici, à l'ordre près) conduise à un apprentissage différent !

Dans tous les cas pratiques, la valeur minimale de  $K$  est donc bien supérieure à la borne fournie par ce tableau. Nous l'avons constaté pour l'apprentissage "par coeur", pour lequel  $K_{min}=2^N$ . Cela est vrai aussi pour les apprentissages de type statistique, pour lesquels on peut estimer, sous certaines hypothèses, que  $K_{min} \geq N^4$ .

Les systèmes statistiques sont efficaces pour isoler une régularité dont le jeu d'exemple  $J$  est un échantillonnage, par exemple une "boule" de rayon  $\sigma$  dans l'hypercube (*i.e.* la classe  $n^o I$  est constituée de tous les points à distance inférieure à  $\sigma$  d'un point donné). Le système le plus naturel est celui qui calcule la moyenne (barycentre) des exemples pour estimer le centre de la boule, et la moyenne des carrés des distances à ce centre pour estimer le carré du rayon. Pour réaliser ces estimations, ce système aura besoin d'un échantillon d'au moins  $K = \sigma^4$  exemples (voir annexe 3). Or  $\sigma$  peut prendre une valeur allant jusqu'à  $N$  (diamètre de l'hypercube), et donc il faut prévoir  $K_{min} \geq N^4$ . Noter que cette taille minimale de  $K$  est très supérieure à la borne calculée dans le tableau précédent :  $card(C(K))$  vaut ici moins de  $2^N \cdot N$  (il y a  $2^N$  centres possibles et au plus  $N$  valeurs pour  $\sigma$ ), et le tableau prévoit un minimum supérieur ou égal à 2 exemples !

Grâce à ces deux exemples, nous vérifions que la borne inférieure pour la taille du jeu d'exemples peut être revue très nettement à la hausse par rapport à l'estimation théorique que nous venons d'effectuer dans le cas le plus général (p. 112). Le nombre minimal d'exemples dépend en effet du type de classification apprenable et de la capacité du mécanisme d'apprentissage étudié à exploiter l'information fournie par les exemples. Il dépendra aussi du fait que le jeu d'exemples risque d'être bruité, comme nous allons le vérifier maintenant.

### 3.4. Difficulté d'apprentissage en présence de bruit

Le mécanisme d'apprentissage défini p. 106 suppose l'existence de capteurs qui permettent de "lire" les exemples et les données (nous avons supposé que ces capteurs étaient des balances à deux états). L'apprentissage peut-il rester correct lorsqu'une partie des données fournies par ces capteurs est erronée ?

On peut distinguer deux types de perturbations du jeu d'exemples. Soit le bruit affecte légèrement chaque exemple (probabilité d'erreur  $p_b$  sur les bits, due par ex. à l'imperfection des capteurs), soit il provoque l'insertion d'exemples aberrants (probabilité  $p_e$  qu'un exemple soit aberrant, due à des erreurs dans la sélection des exemples fournis au système). Nous nous intéressons à des mécanismes d'apprentissage robustes, ainsi définis :

**Mécanisme robuste:** un mécanisme d'apprentissage robuste, par rapport à une probabilité d'erreur donnée sur les bits ( $p_b$ ) ou sur les exemples ( $p_e$ ), sera tel que  $\mathcal{A}(J) = \mathcal{A}(n(J))$ , si  $n(J)$  est le jeu d'exemples bruité correspondant à  $J$ .

Par exemple, le mécanisme statistique évoqué plus haut (voir aussi l'annexe 3) est robuste pour des erreurs de faible amplitude sur les bits, car elles ne vont pas affecter le calcul de la moyenne ou de l'écart type. Les effets d'un bruit plus important peut être neutralisés par une légère augmentation de la taille du jeu d'exemple.

En revanche, le mécanisme indifférent donné précédemment pour apprendre la parité (par somme des bits d'un exemple unique) n'est absolument pas robuste.

Dans le cas d'une utilisation optimale du jeu d'exemples, une probabilité d'erreur  $p_b$  sur les  $K \cdot N$  bits du jeu d'exemples nous permet d'écrire la contrainte suivante pour le nombre  $K$  d'exemples :

$$K \cdot N(1 - p_b \cdot \log_2 K \cdot N) - \log_2 K! \geq \log_2(\text{card}(C(K)))$$

Une probabilité d'erreur  $p_b$  sur les  $K \cdot N$  bits du jeu d'exemples va provoquer environ  $p_b \cdot K \cdot N$  erreurs. Un apprentissage correct exige donc que les jeux  $J$  distants (au sens de Hamming) de moins de  $p_b \cdot K \cdot N$  conduisent au même apprentissage  $\mathcal{A}(J)$ . Or une "boule" de rayon  $p_b \cdot K \cdot N$  dans le  $K \cdot N$ -hypercube contient...

$$\sum_{p=0}^{p_b \cdot K \cdot N} C_{K \cdot N}^p \gg (K \cdot N)^{p_b \cdot K \cdot N}$$

éléments (il y a  $C_{K \cdot N}^p$  façon de changer  $p$  bits). Le nombre de telles boules est donc au plus  $2^{K \cdot N} / (K \cdot N)^{p_b \cdot K \cdot N}$  ("volume" du  $K \cdot N$ -hypercube,  $2^{K \cdot N}$ , divisé par le "volume" d'une boule). Ce nombre de "boules" est nécessairement supérieur au nombre  $\text{card}(C(K))$  de classifications pouvant être apprises correctement, ce qui conduit à la formule annoncée (on remplace  $2^{K \cdot N}$  par le nombre de boules dans la formule de la page 112).

Rappelons que dans cette majoration du nombre de classifications accessibles à partir d'un jeu de taille  $K$ , chaque bit de chaque exemple est utilisé pour déterminer la classification qui doit être apprise. Une erreur (de prob.  $p_e$ ) qui porterait cette fois sur un exemple peut se traduire par  $N$  bits erronés, et a le même effet ici qu'une erreur binaire  $p_b = 1/K$ . Si plusieurs exemples peuvent être erronés, l'accroissement du nombre minimal d'exemples nécessaires peut devenir significatif.

Nous venons de montrer qu'un mécanisme d'apprentissage, quel qu'il soit, nécessitait un nombre minimal d'exemples ( $K_{min}$ ) pour fonctionner, et que  $K_{min}$  pouvait atteindre des valeurs importantes dans certains cas. Mais nous nous intéressons dans ce chapitre aux mécanismes d'apprentissage indifférents. En quoi l'hypothèse d'indifférence va-t-elle modifier la contrainte sur  $K_{min}$ ? En d'autres termes, le fait que le mécanisme soit indifférent rend-il l'apprentissage plus difficile?

Nous allons montrer qu'effectivement, certains apprentissages sont plus difficiles que d'autres (au sens du nombre d'exemples nécessaires) lorsque le mécanisme est supposé indifférent. Ce sera un argument fort pour suggérer que l'apprentissage des connaissances utilisées dans les conversations ne peut pas résulter d'un mécanisme indifférent.

### 3.5. Classifications accessibles aux mécanismes indifférents

Dans certaines situations réelles, on observe que l'apprentissage présente une grande stabilité (*i.e.* le résultat est le même pour un grand nombre de jeux d'exemples différents). Il est alors naturel de supposer que le mécanisme, de par sa constitution même, ne peut apprendre qu'un nombre limité de formes différentes. Mais une telle hypothèse n'est possible, lorsque le mécanisme est de plus considéré comme *indifférent*, que lorsque les formes qu'il peut apprendre sont fortement "symétriques". C'est ce que nous allons nous attacher à montrer ici.

Supposons que  $\mathcal{A}$  soit un mécanisme indifférent, et que  $C(K)$  soit de cardinalité limitée, très inférieure au nombre  $2^{N \times N}$  d'isométries de l'hypercube. Alors nous nous proposons de montrer que les classifications accessibles (les éléments de  $C(K)$ ) seront *harmonieuses*, *i.e.* reconnaîtront des ensembles fortement "harmonieux" :

**Ensemble harmonieux** : un sous-ensemble de l'hypercube sera dit harmonieux s'il est invariant pour de nombreuses isométries. L'**harmonie** d'un tel ensemble peut être estimée par le nombre d'isométries qui laissent ce sous-ensemble invariant.

**Classification harmonieuse** : une classification sera dite harmonieuse si les classes qu'elle définit sont harmonieuses.

Nous allons essayer de calculer l'harmonie des classifications qui peuvent être apprises par un mécanisme indifférent. Cela semble *a priori* compliqué, dans la mesure où l'ensemble  $C(K)$  de ces classifications accessibles contient sans doute des classifications d'harmonies fort différentes. Pour clarifier cette situation, considérons la relation d'équivalence dans  $C(K)$ , notée  $\leftrightarrow$  :

$$\mathcal{A}(J_1) \leftrightarrow \mathcal{A}(J_2) \text{ ssi } \exists \rho \text{ isométrie, } J_2 = \rho(J_1) = \{\rho(x) \mid x \in J_1\}$$

Si  $\mathcal{A}$  est un mécanisme indifférent, alors les classes définies par  $\mathcal{A}(J_1)$  et  $\mathcal{A}(J_2)$  dans  $E$  sont en bijection par  $\rho$ .

En effet, si  $x$  est dans la classe  $n^{\circ}i$  pour  $\mathcal{A}(J_1)$ , alors  $\rho(x)$  est dans la classe  $n^{\circ}i$  pour  $\mathcal{A}(J_2)$  :

$$\mathcal{A}(J_2) (\rho(x)) = \mathcal{A}(\rho(J_1)) (\rho(x)) = \mathcal{A}(J_1) (x) = i$$

Réciproquement, si  $y$  est classé en  $i$  pour  $\mathcal{A}(J_2)$ , son image par  $\rho^{-1}$  sera classée en  $i$  par  $\mathcal{A}(J_1)$  puisque la relation d'équivalence  $\mathcal{A}(J_2) \leftrightarrow \mathcal{A}(J_1)$  s'écrit de même avec  $\rho^{-1}$ .

La relation  $\leftrightarrow$  va nous permettre de structurer  $C(K)$  en classes dont les éléments ont la même harmonie. En effet, si une isométrie  $\sigma$  laisse la classe  $n^{\circ}i$  de  $\mathcal{A}(J_1)$  invariante, alors l'isométrie  $\rho \circ \sigma \circ \rho^{-1}$  laissera invariante la classe  $n^{\circ}i$  de  $\mathcal{A}(J_2)$ .

En effet, soit  $y$  un élément tel que  $\mathcal{A}(J_2) (y) = i$ . Du fait de la définition de l'indifférence :

$$\mathcal{A}(J_2) (\rho(\sigma(\rho^{-1}(y)))) = \mathcal{A}(\rho(J_1)) (\rho(\sigma(\rho^{-1}(y)))) = \mathcal{A}(J_1) (\sigma(\rho^{-1}(y)))$$

Or  $\rho^{-1}(y)$  appartient, comme on l'a vu, à la classe  $n^{\circ}i$  pour  $\mathcal{A}(J_1)$ , classe que  $\sigma$  laisse invariante :

$$\mathcal{A}(J_1) (\sigma(\rho^{-1}(y))) = \mathcal{A}(J_1) (\rho^{-1}(y))$$

et on a bien :

$$\mathcal{A}(J_2) (\rho(\sigma(\rho^{-1}(y)))) = \mathcal{A}(J_1) (\rho^{-1}(y)) = \mathcal{A}(\rho(J_1)) (\rho(\rho^{-1}(y))) = \mathcal{A}(J_2) (y)$$

Ce résultat nous permet de constater que les éléments de  $C(K)$  équivalents pour  $\leftrightarrow$  reconnaissent des ensembles de même harmonie.

En effet, l'application de l'ensemble des isométries de l'hypercube dans lui-même:  $\sigma \rightarrow \rho \circ \sigma \circ \rho^{-1}$  est une bijection. Cette bijection applique le groupe des isométries qui conservent une classe de  $\mathcal{A}(J_1)$  sur le groupe des isométries qui conservent la classe correspondante de  $\mathcal{A}(J_2)$ . Ces deux groupes ont donc le même nombre d'éléments.

L'ensemble  $C(K)$  des classifications que le système peut apprendre est donc constitué d'un certain nombre de sous-ensembles  $S_p$  (les classes de  $\leftrightarrow$ ) dont les éléments ont la même harmonie. Essayons d'aller plus loin, en estimant l'harmonie  $H(S_p)$  caractérisant l'un de ces sous-ensembles de  $C(K)$ . Nous allons montrer le résultat suivant, qui sera très important pour la suite :

$$H(S_p) \cdot \text{card}(S_p) = 2^N \cdot N!$$

Nous allons estimer l'harmonie  $\text{Harm}(\mathcal{A}(J_o))$  d'un des éléments de  $S_p$ , noté  $\mathcal{A}(J_o)$ . En effet, par définition de  $S_p$ ,  $H(S_p) = \text{Harm}(\mathcal{A}(J_o))$ . Parmi les  $2^N x N!$  isométries de l'hypercube, il y a le sous-groupe  $\text{Inv}(\mathcal{A}(J_o))$  de celles qui conservent les classes de  $\mathcal{A}(J_o)$ , et les autres. On a par définition  $\text{Harm}(\mathcal{A}(J_o)) = \text{card}(\text{Inv}(\mathcal{A}(J_o)))$ . Soit la relation d'équivalence notée  $\mathfrak{H}$ , qui relie les isométries conduisant aux mêmes classifications:

$$\rho \mathfrak{H} \tau \text{ ssi } \mathcal{A}(\rho(J_o)) = \mathcal{A}(\tau(J_o))$$

avec toujours la notation  $\rho(J_o) = \{\rho(x) \mid x \in J_o\}$ . Il est facile de voir que :

$$\rho \mathfrak{H} \tau \text{ ssi } \rho^{-1} \circ \tau \in \text{Inv}(\mathcal{A}(J_o))$$

En effet,  $\rho^{-1} \circ \tau(x)$  est classé comme  $x$  par  $\mathcal{A}(J_o)$ , pour tout  $x$  :

$$\mathcal{A}(J_o) (\rho^{-1}(\tau(x))) = \mathcal{A}(\rho(J_o)) (\tau(x)) = \mathcal{A}(\tau(J_o)) (\tau(x)) = \mathcal{A}(J_o) (\tau^{-1}(\tau(x))) = \mathcal{A}(J_o) (x)$$

La relation  $\cong$  définit donc des classes modulo le sous-groupe  $Inv(\mathcal{A}(J_\rho))$  : chaque classe peut s'écrire  $\rho \square o Inv(\mathcal{A}(J_\rho))$ . En particulier, les classes de  $\cong$  ont toutes le même cardinal, égal à celui de  $Inv(\mathcal{A}(J_\rho))$ .

Combien y a-t-il de classes d'équivalence pour  $\cong$  ? Pour répondre, il suffit de remarquer que  $\rho \square o Inv(\mathcal{A}(J_\rho)) = Inv(\mathcal{A}(\rho(J_\rho)))$ . En effet, pour  $\sigma \in Inv(\mathcal{A}(J_\rho))$ ,  $\rho \square o \sigma \square \in Inv(\mathcal{A}(\rho(J_\rho)))$  (et réciproquement) :

$$\mathcal{A}(\rho(J_\rho))(\rho(\sigma(x))) = \mathcal{A}(J_\rho)(\sigma(x)) = \mathcal{A}(J_\rho)(x)$$

Chaque classe modulo  $\cong$  peut donc s'écrire  $Inv(\mathcal{A}(\rho(J_\rho)))$ . Or  $\mathcal{A}(\rho(J_\rho))$  et  $\mathcal{A}(J_\rho)$  sont équivalentes pour la relation  $\leftrightarrow$ . Chaque classe modulo  $\cong$  correspond donc à un élément différent de  $S_p : \mathcal{A}(\rho(J_\rho))$ . Les classes modulo  $\cong$  sont donc au nombre de  $card(S_p)$ . Comme elles ont toutes le même cardinal  $card(Inv(\mathcal{A}(\rho(J_\rho)))) = H(S_p)$  et comme il y a  $2^N \times N!$  isométries au total, la relation est démontrée.

Pour un jeu d'exemples  $J$ , si  $Var(\mathcal{A}(J))$  désigne la variété de  $\mathcal{A}(J)$ , i.e. le nombre de classifications différentes  $\mathcal{A}(\rho(J))$ , alors on peut alors écrire :

$$\text{Harm}(\mathcal{A}(J)) \cdot \text{Var}(\mathcal{A}(J)) = 2^N \cdot N!$$

Cette relation est la même que la précédente :  $Var(\mathcal{A}(J)) = card(S_p)$ , où  $S_p$  est la classe modulo  $\leftrightarrow$  contenant  $\mathcal{A}(J)$ .

Comme  $Var(\mathcal{A}(J))$  est inférieur au nombre  $card(C(K))$  de classifications accessibles au mécanisme d'apprentissage, nous obtenons le résultat annoncé au début de cette section :

*Si un mécanisme d'apprentissage  $\mathcal{A}$  indifférent ne peut atteindre qu'un nombre restreint ( $\ll 2^N \times N!$ ) de classifications, alors ces classifications sont nécessairement harmonieuses.*

Ce résultat a des conséquences qui nous semblent fondamentales pour le problème de l'apprentissage des connaissances (et pour l'apprentissage en général). Dans la suite de ce chapitre, nous allons l'appliquer à différentes situations d'apprentissage.

Auparavant, nous allons essayer d'estimer le nombre minimal d'exemples qu'un mécanisme indifférent nécessite lorsqu'il peut atteindre une classification inharmonieuse.

### 3.6. Difficulté d'apprentissage pour un mécanisme indifférent

Si l'on rapproche la relation précédente de la taille minimum du jeu d'apprentissage nécessaire pour discriminer les classifications accessibles (cf. formule p. 112), on arrive à (en négligeant le terme en  $K^2$ ) :

$$\begin{aligned} K \cdot N - \log_2 K! &\geq \log_2(\text{card}(C(K))) \\ &\geq \log_2(\text{Var}(J)) \\ &\geq \log_2[2^N \times N! / \text{Harm}(J)] \end{aligned}$$

En particulier, si le mécanisme  $\mathcal{A}$  est capable d'atteindre une classification  $\mathcal{A}(J)$  qui soit (totalement) inharmonieuse, c.à.d. telle que les classes qu'elle définit ne soient invariantes par aucune isométrie autre que l'identité ( $Harm(J) = I$ ), alors :

$$K \cdot N - \log_2 K! \geq N + \log_2 (N!)$$

Cette formule donne une borne inférieure<sup>(2)</sup> à la taille  $K$  d'un jeu d'exemples qui permet à un mécanisme d'apprentissage indifférent d'apprendre une classification dépourvue d'harmonie. Le tableau ci-dessous donne  $N + \log_2 N!$  pour quelques valeurs de  $N$ , à comparer avec les valeurs du tableau de la page 112.

N	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64
$N + \log_2 N!$	23	41	60	81	103	126	150	174	199	225	251	278	305	332	360

Cette comparaison montre que pour qu'un mécanisme indifférent, susceptible d'atteindre des classifications inharmonieuses, puisse fonctionner, le nombre d'exemples doit nécessairement excéder une valeur de 4 ou 6 (suivant la valeur de  $N$ ).

Ce résultat est important : il signifie qu'un mécanisme indifférent ne peut en aucun cas apprendre une classification inharmonieuse avec moins de 4 ou 6 exemples.

Il existe une autre contrainte sur le nombre des exemples avec lesquels un mécanisme indifférent d'apprendre une classification inharmonieuse. En effet, l'harmonie d'un jeu  $J$  est inférieure ou égale à celle de la classification apprise  $\mathcal{A}(J)$  :

$$Harm(J) \leq Harm(\mathcal{A}(J))$$

En effet, toute isométrie laissant  $J$  invariant laisse les classes de  $\mathcal{A}(J)$  invariantes :

$$\forall \rho \text{ isométrie } \rho(J) = J \Rightarrow \mathcal{A}(J) \circ \rho = \mathcal{A}(\rho^{-1}(J)) = \mathcal{A}(J)$$

En particulier, si  $\mathcal{A}(J)$  est inharmonieuse, alors  $Harm(J) = I$ . Ceci a des conséquences sur la taille  $K$  de  $J$ . Un ensemble de  $K$  vecteurs binaires de dimension  $N$  est nécessairement harmonieux si  $K$  est trop faible. On doit nécessairement avoir :

$$K \geq 1 + \log_2 N$$

Si  $N > 2^K$ , les exemples auront tous au moins deux composantes égales (aux mêmes positions) et le jeu sera invariant pour la permutation de ces deux composantes. Par exemple, les trois vecteurs suivants n'ont pas de composantes systématiquement égales, alors que les deux premiers ont par ex. les positions 1 et 4 égales.

$$\begin{array}{l} \left[ \begin{array}{cc|cc} \boxed{0} & 1 & 1 & \boxed{0} & 1 & 0 \end{array} \right] \leftarrow \text{vecteur n}^\circ 1 \\ \left[ \begin{array}{cc|cc} \boxed{1} & 0 & 0 & \boxed{1} & 1 & 0 \end{array} \right] \\ \left[ \begin{array}{cc|cc} 0 & 0 & 1 & 1 & 0 & 1 \end{array} \right] \leftarrow \text{vecteur n}^\circ K \\ \leftarrow \quad N \quad \rightarrow \end{array}$$

Il y a  $2^K$  "colonnes" possibles différentes, et  $N$  colonnes effectives. D'où la condition  $N \leq 2^K$  pour avoir un jeu qui ait des chances de ne pas être invariant.

Cela ne suffit malheureusement pas. Dans cet ensemble de 3 vecteurs, les positions 1 et 3, par exemple, sont complémentaires, et la permutation de ces positions, venant après la translation par le vecteur  $[1,0,1,0,0,0]$ , laisse l'ensemble de trois vecteurs invariant. Pour éviter cela, il faut que  $N \leq 2^{K-1}$  (il y a  $2^{K/2}$  "colonnes" possibles à la complémentation près).

<sup>2</sup> En effet, le calcul montre que le membre de gauche est une fonction croissante de  $K$  pour les valeurs suffisamment inférieures à  $2^N$ .

Toutefois, la non-harmonie ne sera toujours pas garantie pour autant. L'ensemble suivant ne comporte pas de positions complémentaires, mais il est pourtant invariant pour cinq permutations, par exemple la permutation  $\langle 2,1,4,3,5,6 \rangle$  qui laisse invariants les 3<sup>ème</sup> et 4<sup>ème</sup> vecteurs, et échange les deux premiers.

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Il faut adjoindre un autre vecteur, par exemple :

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

pour obtenir un ensemble inharmonieux, comme le montre une recherche exhaustive.

La condition  $K \geq I + \log_2 N$  que nous venons d'obtenir est à peu près équivalente à la condition obtenue au début de cette section (p.112). Il suffit pour s'en convaincre d'appliquer la formule de Stirling dans le membre  $I + \log_2 N! / N$ .

L'exigence de non-harmonie pour  $J$  peut se révéler contraignante lorsque  $J$  est bruité. Il doit se trouver (en tant que vecteur binaire de dimension  $K.N$ ) à une distance supérieure à  $p_b.K.N$  de tout jeu harmonieux (ce n'était pas le cas pour le jeu de cinq vecteurs de l'exemple précédent : une erreur sur le dernier bit du dernier exemple rend ce jeu invariant pour la permutation  $\langle 1,6,3,5,4,2 \rangle$ ).

Rappelons que cette borne inférieure du nombre  $K$  d'exemples nécessaires à l'apprentissage d'une classification qui peut être inharmonieuse est très sous-estimée dans la plupart des cas pratiques. Elle suppose une constitution optimale du jeu d'exemples. De plus notre estimation de  $C(K)$  ne prend en compte qu'une seule classification inharmonieuse et les classifications obtenues par transformation isométrique.

Cette minoration se révélera pourtant utile, soit pour contribuer à un calcul plus complet du nombre minimal d'exemples dans tel ou tel cas particulier, soit pour montrer qu'un système apprenant des classifications inharmonieuses avec un seul exemple ne peut pas être indifférent.

Le résultat que nous avons obtenu p. 116 a de nombreuses conséquences. Nous venons d'en observer la première : l'apprentissage d'une classification inharmonieuse exige un nombre minimum d'exemples qui est supérieur à 4 ou 6. En conséquence, l'apprentissage "en un coup" d'une forme inharmonieuse ne peut être que le fait d'un mécanisme non indifférent, qui utilise une connaissance *a priori*.

De même, nous allons constater maintenant qu'un apprentissage stable de formes inharmonieuses ne peut pas être obtenu avec un mécanisme indifférent. Ceci nous permettra d'exclure ces mécanismes lorsqu'il s'agira d'expliquer l'apprentissage des connaissances.

## 4. Convergence, indifférence et théories de l'apprentissage

### 4.1. Théories de l'apprentissage et indifférence

Dans le chapitre II.1, nous avons indiqué succinctement les aspects indifférents ou non indifférents des différents modèles d'apprentissage. Typiquement, les mécanismes d'appariement ont peu de chance d'être indifférents, alors que les mécanismes d'extraction de régularités (connexionnisme, systèmes sélectifs, algorithmes statistiques) sont généralement indifférents.

Le critère d'indifférence devient alors un outil particulièrement intéressant pour juger de l'adéquation d'un de ces modèles à une situation d'apprentissage réellement observée. En particulier, certaines caractéristiques du cas observé, comme l'harmonie des formes apprises, peut donner des indications qui permettront éventuellement d'exclure certains modèles. C'est ce que nous allons montrer en considérant les apprentissages convergents.

## 4.2. Convergence de certains apprentissages

Dans certaines situations, le résultat de l'apprentissage est à peu près identique pour des sujets apprenants différents. Nous caractérisons un tel phénomène en disant qu'il s'agit d'un apprentissage convergent :

**Apprentissage stable** : un apprentissage avec le mécanisme  $\mathcal{A}$  est stable par rapport à un ensemble  $\mathcal{E}$  de jeux d'exemples si  $\mathcal{A}(J)$  a une valeur constante pour tout  $J$  appartenant à  $\mathcal{E}$ .

**Apprentissage convergent** : un apprentissage est convergent si plusieurs mécanismes (individus), confrontés à des jeux d'exemples éventuellement différents, donnent le même résultat.

Face à un apprentissage convergent entre individus similaires, on peut dire que le *mécanisme* est stable (si les expériences des individus sont différentes), ou que les *données* sont stables (si tous les individus sont confrontés à la même expérience).

Par exemple, dans certaines conditions, des algorithmes de partitionnement équivalents (*e.g.* nuées dynamiques avec des choix de "semences" différents) pourront partitionner certaines données de la même façon. On observe de même une convergence dans de nombreux apprentissages humains. Les enfants parviennent à tenir en équilibre sur un vélo en tournant le guidon du côté où ils tombent; ils apprennent à parler leur langue maternelle d'une façon peu discernable de celle des enfants de la même communauté linguistique (phonèmes prononcés, accent, tournures syntaxiques employées, etc.); ils acquièrent des connaissances très voisines sur un domaine donné (par ex. le code de la route); en dessous d'un certain âge, tous dessinent les arbres perpendiculaires à une pente, pour les dessiner verticaux plus tard [Piaget & Inhelder 1947:444]. Jean Piaget est l'un de ceux qui a le mieux montré, et dans le détail, l'étonnante ressemblance entre les apprentissages qui permettent à des enfants différents d'atteindre l'état "stable" des compétences adultes.

Certes, l'apprentissage est souvent vu comme un système permettant de faire l'économie des particularités. Contrairement à de nombreux animaux capables de communiquer (*e.g.* les abeilles), nous ne naissons pas avec la capacité de parler notre langue maternelle. Nous naissons avec la capacité de l'acquérir. Les particularités de la langue n'ont pas à être codées dans notre ADN. L'apprentissage est ainsi décrit comme permettant d'accéder à des formes qui, du fait de leur variabilité potentielle, apparaissent comme plus riches, plus complexes et mieux adaptés que ce que permettent les systèmes figés.

Toutefois, pour les actes de communication comme pour beaucoup d'autres (notamment les comportements sociaux), l'apprentissage n'est tolérable que s'il conduit à des formes stables d'un individu à l'autre. La convergence est dans ces cas-là une contrainte fondamentale hors de laquelle l'apprentissage peut être considéré comme non viable.

On ne s'étonnera donc pas de ce que les cas de convergence soient très nombreux. La stabilité du résultat de l'apprentissage peut même apparaître comme une caractéristique du problème de l'apprentissage, dans la mesure où l'on définit le plus souvent "ce qui doit être appris" par rapport à la performance d'un autre système apprenant.

Or cette nécessité de la convergence est rarement prise en considération de manière explicite dans les théories de l'apprentissage, ce qui est étonnant dans la mesure où il peut constituer une difficulté pour certaines d'entre elles.

La stabilité du résultat d'un apprentissage peut être, selon les cas, soit niée ou relativisée, soit attribuée à la stabilité des données fournies au système apprenant, soit encore expliquée comme due à des contraintes internes du sujet apprenant. Nous allons voir comment les principales théories se positionnent par rapport au problème de la convergence, et dans quelle mesure l'indifférence d'un mécanisme est compatible avec sa stabilité.

### 4.3. Expliquer la convergence pour un mécanisme indifférent

Une stabilité réelle des apprentissages ne peut avoir que deux origines : la stabilité des jeux de données, ou le très faible nombre des formes apprenables. Lorsqu'on a affaire à un mécanisme indifférent, la deuxième solution exige, comme nous l'avons vu (p. 116), que les formes apprises soient très harmonieuses. En utilisant ces critères, nous pouvons classer les principales théories de l'apprentissage d'une manière éclairante, quoiqu'un peu réductrice.

mécanisme	exemples	indif.	origine de la convergence
apprentissage statistique	conditionnement opérant carte de Kohonen	oui	stabilité des données ou faible nombre et donc harmonie des formes apprenables
constructivisme (Piaget)	apprentissage de la conservation d'une quantité	oui	faible nombre et donc harmonie des formes apprenables
théorie des "fusibles"	apprentissage d'une grammaire	non	faible nombre des formes apprenables

En termes logiques, nous pouvons caractériser la situation grâce aux deux incompatibilités suivantes :

[ *mécanisme indifférent & formes apprenables peu nombreuses*  
& **non** *formes apprenables harmonieuses* ]  $\Rightarrow$  **F**

[ *apprentissage convergent & non stabilité des données*  
& **non** *formes apprenables peu nombreuses* ]  $\Rightarrow$  **F**

On constate que les différentes théories de l'apprentissage adoptent des hypothèses différentes pour échapper à ces deux contraintes.

#### 4.3.1. Convergence et apprentissage statistique

Les mécanismes d'apprentissage de type statistique évitent de faire intervenir toute contrainte interne à l'organisme apprenant qui puisse expliquer une stabilité éventuelle des apprentissages. En particulier, l'apprentissage *stable* de formes inharmonieuses ne peut alors être dû qu'à l'identité des expériences auxquelles sont exposés les individus. Cette identité est généralement peu crédible, d'autant qu'elle porte, pour les mécanismes statistiques, sur des échantillons de taille conséquente.

En conséquence, les mécanismes statistiques peuvent raisonnablement être invoqués seulement pour des apprentissages de formes harmonieuses. L'apprentissage de formes inharmonieuses sera difficile et non stable.

#### 4.3.2. Convergence et constructivisme

Jean Piaget s'oppose à l'idée selon laquelle les structures apprises sont le reflet de régularités de l'environnement. Il écrit à propos des opérations logico-mathématiques [Piaget 1967] :

"Nous serons, en effet, conduits à chercher le point de départ de la formation des opérations logico-mathématiques dans une abstraction à partir de la coordination générale des actions. D'une part, ces opérations ne peuvent pas être tirées des objets eux-mêmes, puisque l'abstraction à partir des objets ne donne lieu qu'à des constatations non nécessaires (au sens de la nécessité déductive) et plus précisément à des jugements simplement probables, tandis que les opérations logico-mathématiques sont caractérisées par une nécessité interne due à leur réversibilité entière (donc non physique)." (p.34)

La théorie constructiviste de Jean Piaget est fondée sur l'harmonie des formes apprenables. Piaget décrit l'ensemble des "formes" correctes que l'enfant possède, lorsqu'il maîtrise un stade donné, comme *invariant* pour un ensemble *fermé* de transformations qui lui sont accessibles. Autrement dit, l'enfant accepte comme correcte une situation  $s_1$  obtenue par une transformation  $\tau$  à partir d'une situation  $s_0$ , connue comme correcte, et ceci d'autant

plus facilement qu'il est capable d'inverser  $\tau$ . Par combinaison sur les transformations, l'enfant atteint rapidement l'ensemble des formes correctes. Les transformations  $\tau$  prises en considération par l'enfant correspondent généralement à des "schèmes d'action", et peuvent être modélisées dans notre contexte par des isométries.

Considérons par exemple la célèbre expérience des verres [Piaget 1967:610], expérience qui peut être très facilement reproduite avec des enfants de 6 à 8 ans. Lorsque l'on verse le contenu d'un verre large dans un verre étroit, l'enfant de 6 ans déclare qu'il y a davantage d'eau après, alors que l'enfant de 8 ans déclare qu'il y en a autant (parce que c'est "la même eau"). Selon Piaget, le jeune enfant ne prend en compte qu'une seule dimension, ici la hauteur, pour estimer la quantité relative de liquide. Par ailleurs, il n'est pas conscient de l'absurdité (d'où vient le liquide supplémentaire ?) de sa réponse. En revanche, son aîné, sensible au fait que la transformation (transvasement du liquide) est une opération réversible, prend conscience de la conservation de la quantité. Il résout alors l'apparente contradiction en prenant en compte simultanément la hauteur et la largeur du verre.

Essayons de décrire cette expérience dans notre contexte. Les transformations auxquelles l'enfant est sensible pendant l'expérience résultent des actions de transvasement. Nous pourrions projeter la théorie piagétienne dans notre contexte en disant que :

- l'enfant apprend selon un mécanisme *indifférent* (voir déf. p.107) aux transformations qui lui sont accessibles
- les formes apprises sont invariantes par toute transformation accessible à l'enfant (ces transformations constituent un groupe laissant la forme apprise invariante).

Cette description est parfaitement compatible avec la *convergence* des apprentissages observée par Piaget. Il va jusqu'à insister sur la *nécessité* de la forme apprise [Piaget 1979]. Du fait de leur propriété d'invariance, les formes apprenables sont, comme nous l'avons vu (p. 116), relativement peu nombreuses, et l'enfant détermine la bonne forme sans difficulté en étant exposé à seulement quelques situations. Mais nous pouvons pousser l'analogie un peu plus loin.

Lors de l'action de transvasement, l'enfant de six ans ne tient pas compte du rayon  $r$  du verre. Il apprend la forme suivante liant la quantité d'eau ( $v$ ) et la hauteur  $h$  atteinte par le liquide

$$\{ (\text{Log } v, \text{Log } h) \mid \text{Log } v = \text{Log } h + \text{Cte} \}$$

Nous supposons que l'enfant est sensible au logarithme des dimensions du verre (rayon  $r$  et hauteur  $h$  atteinte par le liquide) et de la quantité d'eau ( $v$ ). En effet, les capteurs humains semblent être logarithmiques, ce qui signifie ici que l'enfant est sensible aux variations relatives.

L'enfant de 6 ans semble appliquer au transvasement la "loi" qu'il a repérée dans les opérations de remplissage d'un verre unique :  $v = C.h$ . A partir de quelques exemples (en jouant par ex. avec le robinet), c.à.d. de quelques couples  $(v_p, h_p)$ , il a généralisé les situations "normales" à l'ensemble  $\{ (\text{Log } v, \text{Log } h) \mid \text{Log } v = \text{Log } h + \text{Log } C \}$ . Il semble assimiler le transvasement à un remplissage.

On peut décrire le résultat de l'apprentissage comme une classification binaire des situations en situations normales / situations anormales. Cet enfant de 6 ans serait certainement surpris si, dans une opération de remplissage, on rajoutait de l'eau et que (grâce à un dispositif approprié) le niveau de l'eau restait constant.

L'enfant plus âgé repère la constance du volume et l'influence de la section du verre, et apprend la forme "correcte" :

$$\{ (\text{Log } v, \text{Log } h, \text{Log } r) \mid \text{Log } v = \text{Log } h + 2.\text{Log } r + \text{Cte} \}$$

La loi "exacte" est, pour un verre cylindrique,  $v = \pi.r^2.h$ . L'enfant repère qualitativement l'influence de la section du verre.

Dans les deux cas, les transformations dues aux transvasement telles qu'elles sont perçues par l'enfant laissent invariante la forme apprise. Or elles se traduisent par des translations dans le repère logarithmique :  $(\text{Log } v, \text{Log } h) \rightarrow (\text{Log } v + \text{Log } k, \text{Log } h + \text{Log } k)$  dans le premier cas (transvasement assimilé à un remplissage), et  $(\text{Log } v, \text{Log } h, \text{Log } r) \rightarrow (\text{Log } v, \text{Log } h + \text{Log } k, \text{Log } r - 1/2 . \text{Log } k)$  dans le second cas. Les

formes apprises sont des droites dans ces repères. Ce sont les formes harmonieuses (invariantes pour un groupe de translations, celles qui correspondent aux transformations notées par l'enfant) que la théorie piagétienne désigne comme seules accessibles dans ce cas.

On comprend, grâce à Piaget, comment ces formes peuvent être facilement apprises à partir d'une expérience très limitée (quelques couples  $(v_i, h_i)$  ou quelques triplets  $(v_i, h_i, r_i)$ ) : l'enfant "extrapole" à une forme symétrique (harmonieuse) compatible avec les cas qu'il perçoit. Le mécanisme indifférent postulé par Piaget oblige en quelque sorte l'enfant à opérer cette extrapolation. La rareté des formes harmonieuses compatibles explique la convergence de ce type d'apprentissage (selon les contraintes explicitées p. 120).

### 4.3.3. Convergence et Gestaltisme

L'idée que les formes apprenables sont harmonieuses et peu nombreuses est à la base de la théorie de la Gestalt, qui privilégiait les "bonnes formes", celles qui sont simples, régulières et symétriques. Rock et Palmer [1991] rappellent les idées de Fritz Heider sur la vision gestaltiste des sciences sociales :

"Heider développa, en outre, le concept de l'équilibre : les individus préfèrent les relations cognitives harmonieuses. Par exemple, si Jean aime une personne X et pense que cette personne X aime une personne Y, alors le système des accords sera équilibré si Jean aime aussi Y - et sera déséquilibré s'il ne l'aime pas. Cette idée rappelle le principe de simplicité : la tendance à obtenir l'organisation la meilleure ou la plus simple."

La théorie de la Gestalt prête aux organismes des propriétés qui sont caractéristiques des systèmes *indifférents* : harmonie des formes apprises lorsqu'elles sont nécessairement peu nombreuses (cf. p. 116). Par exemple, le système visuel est censé préférer la perception la plus régulière et la plus symétrique, compatible avec l'information sensorielle. Les "bonnes" images, c.à.d. celles qui peuvent être décrites avec le moins d'information, sont reconnues plus vite et sont mieux mémorisées que les "mauvaises" [Rock & Palmer 1991]. C'est donc ici que l'on trouve la référence la plus explicite à l'harmonie des formes apprises, rendue nécessaire par l'observation d'une stabilité des formes apprises d'un sujet à l'autre, et par la diversité des expériences (cf. p. 120).

### 4.3.4. Convergence et théorie des "fusibles"

La théorie des "fusibles" (voir chap. II.1) n'est absolument pas liée à l'hypothèse d'indifférence. Elle postule l'existence d'un appareil dont la structure innée riche n'a aucune raison de posséder la propriété d'indifférence. En revanche, cette structure innée peut être suffisamment contraignante pour que le nombre des formes apprenables dans un environnement "normal" soit très fortement restreint, ce qui revient à dire que des apprentissages très stables d'un sujet à l'autre peuvent être atteints même lorsque l'environnement est assez variable (se reporter à l'exemple de l'apprentissage d'une grammaire simple, chap. II.1). Cette situation échappe à la première des deux contraintes explicitées p. 120, puisque le système n'est pas indifférent, ici.

Les théories que nous venons d'évoquer, et bien d'autres, ont été utilisées par certains pour rendre compte de l'apprentissage le plus caractéristique des êtres humains, celui des relations logiques qui lient les connaissances conceptuelles. Nous allons maintenant tenter d'indiquer que l'apprentissage des relations logiques ne peut pas être le fruit d'un apprentissage indifférent.

## 5. Indifférence et apprentissage de relations logiques

### 5.1. L'apprentissage des relations logiques

Certains auteurs défendent l'idée que les relations que nous modélisons par la logique sont acquises par des mécanismes statistiques. Par exemple P. Smolensky [1988] défend la version subsymbolique représentée par le connexionnisme en la disant capable d'effectuer les mêmes opérations que celles dont sont capables les systèmes symboliques:

"if an agent possesses knowledge in the symbolic form  $p \rightarrow q$  and additional knowledge  $p$ , then by syntactic operation the agent can produce  $q$ ; proof theory guarantees that the truth conditions of the

agent's knowledge (or beliefs) have not changed.

There are fairly direct subsymbolic counterparts to this proof theoretic account. The role of logical inference is played by statistical inference. By explicitly formalizing tasks like prediction as statistical inference tasks, it is possible to prove for appropriate systems that subsymbolic computation is valid in a sense directly comparable to symbolic proof."

Il s'agit bien ici de doter le système connexionniste des règles d'inférence que le système symbolique équivalent utilise. Simplement, ces règles doivent être acquises par un mécanisme indifférent, en l'occurrence un apprentissage statistique. Nous verrons qu'une telle position conduit à exiger que le sujet soit exposé à des jeux de données volumineux et fiables, ce qui est très loin d'être systématiquement le cas.

La position de Piaget est tout à fait différente. Les relations logiques ne peuvent pas, selon lui, résulter de la seule observation de régularités par le sujet apprenant.

"A considérer le développement de l'enfant de 2 à 15 ans il est évident que la logique devient <<nécessaire>> en ce sens qu'il est impossible à la pensée de l'adolescent de ne pas se servir d'un certain nombre de mécanismes inférentiels et de ne pas éprouver, en dégageant leurs conclusions, un sentiment particulier de <<nécessité>> déductive : par exemple si A implique B et si B implique C il est impossible ou absurde de ne pas admettre que A implique C. Il y a donc là un rapport *sui generis* qui déborde largement la constatation ou la simple régularité inductive et expérimentale, relevant de la probabilité ou d'un déterminisme de fait mais non pas de la nécessité intrinsèque." [Piaget 1967] (p.424)

L'établissement de ces liens logiques sont, pour Piaget, la conséquence de la fermeture par composition des actions du sujet, comme il l'explique à propos de la mise en série d'objets :

"Tant que la sériation, par exemple, soit  $A < B < C \dots$  ne donne lieu par l'enfant qu'à une construction par tâtonnements empiriques, la structure ne peut être dite fermée et par conséquent la transitivité appliquée aux objets ( $A < C$  si  $A < B$  et  $B < C$ ) n'apparaît pas comme nécessaire, mais comme possible, probable, etc. Dès que la sériation est construite de façon opératoire en choisissant chaque fois le plus petit élément donné ou restant, et en comprenant donc qu'un élément quelconque E est à la fois plus grand que les précédents A, B, C, D et plus petit que les suivants F, G, etc., alors la structure constitue un tout cohérent et fermé, c'est-à-dire dont les relations sont interdépendantes et composables entre elles sans sortir du système : en ce cas la transitivité apparaît comme <<nécessaire>> et cette <<nécessité>> logique ne se reconnaît pas simplement à un sentiment intérieur, impossible à vérifier, mais au comportement intellectuel du sujet qui utilise avec sécurité et rigueur le nouvel instrument déductif qu'il vient de conquérir." [Piaget 1967] (p. 437)

Nous avons essayé de montrer plus haut que le mécanisme d'apprentissage invoqué par Piaget était un mécanisme indifférent aux transformations perçues par le sujet (ces transformations pouvant souvent être modélisées comme des isométries). On devrait alors observer une "harmonie" (voir p. 114) dans les relations logiques apprises, or nous verrons que cela est loin d'être systématiquement le cas lorsqu'il s'agit des relations logiques portant sur des objets particuliers.

Rock et Palmer [1991] rappellent le point de vue gestaltiste sur l'apprentissage des liens logiques :

"L'un des moyens propres aux humains, contrairement aux animaux, pour parvenir à leurs fins, est celui de l'explication. Ecouter seulement ne suffit pas, bien sûr, car l'auditeur doit parvenir à la même structure cognitive que celui qui explique, s'il veut percevoir les liens essentiels entre les diverses données. Les auditeurs n'ont pas à passer par le processus créatif, comme celui qui a résolu le problème, mais l'état final du processus de compréhension doit être le même. [...] Non seulement il est satisfaisant d'obtenir ainsi la solution d'un problème, mais elle sera moins vite oubliée que lorsqu'on la mémorise mécaniquement et elle servira à résoudre de nouveaux problèmes."

Cet apprentissage des liens logiques par l'explication correspond assez à notre propre thèse. Toutefois rien n'est dit ici de la nature de l'explication et des contraintes qu'elle impose à celui qui explique. En particulier, compte tenu du contexte général de la théorie gestaltiste, on est en droit de s'attendre à ce que toute "bonne" explication présente des caractéristiques de "bonne forme", ce qui semble ne pas être le cas.

## 5.2. Logique et harmonie

### 5.2.1. Inharmonie des contextes conversationnels

Nous allons indiquer ici le fait que les relations logiques qui expriment les contraintes d'un contexte donné définissent un ensemble, l'ensemble des situations possibles, qui est inharmonieux. L'analyse des conversations montre que les interlocuteurs partagent un *contexte* dès le début de l'interaction, que ce contexte est clairement circonscrit (cf. discussion du chapitre I.2). Nous avons représenté par des règles d'incompatibilité (chapitre I.1). Nous défendrons le point de vue selon lequel ces règles d'incompatibilité sont connues et manipulées en tant que telles par les interlocuteurs (chapitre II.3).

On peut toutefois imaginer (c'est ce que font certains auteurs) que les interlocuteurs n'ont pas directement connaissance de ces incompatibilités. Si l'on accepte ce point de vue, comment font-ils pour décider que la situation à laquelle ils sont confrontés (ou qui est évoquée devant eux) est incohérente ? Une réponse souvent invoquée est que les individus ont réussi à circonscrire l'ensemble des possibles grâce à leur expérience. La connaissance des situations qu'ils ont vécues (ou qu'on leur a rapportées) leur permettait ainsi de décider si une nouvelle situation est acceptable ou non. Examinons cette question à partir d'un exemple.

Notre expérience dans la constitution de contextes logiques pour la conduite de dialogues de type SAVANT3 (cf. chap. III.2) nous a habitués à traiter des ensembles de 10 à 15 propositions sur lesquelles portent moins de 20 relations d'incompatibilité. Un de nos dialogues, conçu pour l'enseignement de la notion de filtrage, utilise 13 propositions, et les 14 relations d'incompatibilité suivantes (ci-dessous à gauche) :

	1	2	3	4	5	6	7	8	9	10	11	12	13
[ 1 2 ]	0	1	1	0	0	0	1	1	1	0	0	0	0
[ -3 ]	0	1	1	0	0	0	1	1	1	0	0	0	1
[ 4 5 ]	0	1	1	0	0	0	1	1	1	0	0	1	1
[ 6 ]	0	1	1	0	1	0	1	1	1	0	0	0	0
[ 3 -7 ]	0	1	1	0	1	0	1	1	1	0	0	0	1
[ -8 ]	0	1	1	0	1	0	1	1	1	0	0	1	1
[ 8 -9 ]	0	1	1	1	0	0	1	1	1	0	0	0	0
[ 9 10 ]	0	1	1	1	0	0	1	1	1	0	0	0	1
[ 3 11 ]	0	1	1	1	0	0	1	1	1	0	0	1	1
[ 2 10 ]	1	0	1	1	0	0	1	1	1	0	0	0	0
[ 7 -2 -10 -1 -10 -6 ]	1	0	1	1	0	0	1	1	1	0	0	0	0
[ 10 1 ]	1	0	1	1	0	0	1	1	1	0	0	0	1
[ 2 12 -13 ]	1	0	1	1	0	0	1	1	1	0	0	1	0
[ 1 -4 ]	1	0	1	1	0	0	1	1	1	0	0	1	1

Les nombres, dans la partie de gauche, indiquent les numéros des propositions logiques. Les propositions d'une même ligne (entre crochets) sont incompatibles. C'est le cas par ex. des propositions n°1 et n°2. Le signe "-" désigne la négation. Ces incompatibilités sont autant de contraintes qui limitent l'ensemble des possibilités.

Nous avons indiqué à droite l'ensemble des situations possibles. Chaque ligne de 1 et de 0 représente une situation possible, un 1 en colonne  $n$  signifiant que la  $n^{\text{ième}}$  proposition est vraie dans cette situation, un 0 qu'elle est fautive. On peut par exemple vérifier en parcourant les deux premières colonnes que les propositions n°1 et n°2 ne sont jamais simultanément vraies.

En effet, si l'on en croit la plupart des théories (dont la théorie statistique de P. Smolenski rappelée ci-dessus), le sujet apprenant acquiert la connaissance (que nous représentons par des incompatibilités logiques) à partir des situations "normales" auxquelles il est exposé. Les incompatibilités ci-dessus ne seraient ainsi reconnues par le sujet qu'indirectement, en tant que situations qui n'ont jamais été rencontrées auparavant. Dans l'exemple ci-dessus, les 14 relations logiques n'autorisent que quelques situations possibles. *C'est à certaines de ces situations que le sujet serait exposé, et non aux relations logiques.*

La question qui se pose est alors la suivante. L'individu apprend-il à discriminer les cas possibles des cas impossibles à partir d'exemples possibles, ou dérive-t-il cette capacité d'autres connaissances à l'aide de mécanismes à préciser ? Nous allons examiner ici la première de ces hypothèses (la deuxième sera largement explorée dans le chapitre II.3).

L'ensemble des situations possibles de notre exemple possède une harmonie relativement faible : seules 20160 isométries laissent cet ensemble invariant, à comparer avec les  $2^{13} \times 13!$  ( $= 5 \times 10^{13}$ ) transformations isométriques possibles.

Cet exemple a été choisi car l'harmonie de l'ensemble des possibilités est facile à calculer. En effet, les 7 propositions 3, 6, 7, 8, 9, 10 et 11 ont toujours la même valeur de vérité, et sont systématiquement permutable entre elles, et entre elles seulement (avec une complémentation associée lorsque les valeurs sont opposées). Elles contribuent donc pour  $7!$  à l'harmonie totale.

L'ensemble des situations possibles exprimées sur les six propositions restantes est invariant pour les 4 isométries suivantes, et celles-là seulement (chaque ligne indique la permutation puis la translation) :

$\langle 1,2,4,5,12,13 \rangle \rightarrow [0,0,0,0,0,0]$   
 $\langle 1,2,4,5,13,12 \rangle \rightarrow [0,0,0,0,1,1]$   
 $\langle 2,1,4,5,12,13 \rangle \rightarrow [1,1,0,0,0,0]$   
 $\langle 2,1,4,5,13,12 \rangle \rightarrow [1,1,0,0,1,1]$

L'harmonie de l'ensemble des possibilités (pour les 13 propositions) est donc seulement de  $4 \times 7! = 20160$ .

Cet exemple permet de montrer que des liens logiques issus de situations réelles conduisaient à des ensembles de possibilités qui n'ont aucune raison d'être harmonieux. Certes l'ensemble possède une certaine harmonie, mais elle est très faible en valeur relative, et provient de la constance de 7 propriétés sur 13 (caractéristique que nous avons sélectionnée pour faciliter les calculs). En fait, l'harmonie n'est pas facile à calculer dans le cas général. Pour montrer qu'il n'y a en général aucune raison pour que certaines propositions aient une valeur constante, nous donnons ci-dessous un deuxième exemple d'ensemble de possibilités pour un dialogue SAVANT3, relatif lui aussi au filtrage :

	1	2	3	4	5	6	7	8	9	10	11	12	13		
[ 1	-2	]	*	1	1	1	0	*	1	1	0	1	1	0	0
[ 3	-4	]	*	1	1	1	0	0	1	1	0	1	*	0	0
[ 5	3	]	*	1	1	1	0	0	1	1	0	1	*	*	0
[ 6	-5	-7	]	0	0	0	*	0	0	0	1	*	0	*	*
[ -8	]		0	0	0	*	0	0	0	1	0	*	*	*	1
[ 9	10	]	0	0	0	*	1	*	0	1	*	0	*	*	0
[ 11	12	1	]	0	0	0	*	1	0	0	1	0	*	*	*
[ 13	5	]	0	0	1	1	0	0	*	1	0	1	*	*	0
[ 13	3	]	0	0	1	1	0	0	0	1	*	0	*	*	0
[ -5	-3	-13	]	0	0	1	1	0	0	0	1	0	*	*	*
[ 7	-3	-6	]	0	1	1	1	0	*	1	1	0	1	1	*
[ 6	-5	-11	]	1	1	0	*	0	0	0	1	*	0	*	0
[ 2	5	]	1	1	0	*	0	0	0	1	*	0	0	*	1
[ 6	-5	-2	]	1	1	0	*	0	0	0	1	0	*	*	0
[ 7	-10	]	1	1	0	*	0	0	0	1	0	*	0	*	1
[ 10	6	-3	]	1	1	1	1	0	0	*	1	0	1	*	0
[ 2	-7	-1	]	1	1	1	1	0	0	*	1	0	1	0	*
				1	1	1	1	0	0	0	1	*	0	*	0
				1	1	1	1	0	0	0	1	*	0	0	*
				1	1	1	1	0	0	0	1	0	*	*	0
				1	1	1	1	0	0	0	1	0	*	0	*

Le signe \* signifie que la valeur de vérité de la proposition correspondante est quelconque. Il n'est pas possible d'estimer l'harmonie de cet ensemble par une génération exhaustive des  $2^{13} \times 13!$  ( $= 5 \times 10^{13}$ ) isométries possibles. Nous pouvons penser qu'elle est encore plus faible que celle du premier exemple.

Il semble donc que les contextes maniés couramment dans les conversations correspondent à des ensembles de possibilités non harmonieux. Pour citer encore un exemple, un de nos dialogues de démonstration (ce dialogue, que nous ne détaillons pas ici, parle d'un étudiant qui assiste à un séminaire) a un contexte de 6 règles portant sur 5 propositions, qui permet un ensemble de 11 possibilités. Or cet ensemble, pourtant restreint, est inharmonieux (harmonie égale à 1).

### 5.2.2. Apprentissage direct des situations possibles et compétence conversationnelle

La compétence conversationnelle des individus leur permet de juger, dans chaque contexte, si une situation est possible ou impossible.

Dans l'hypothèse où l'apprentissage consisterait à apprendre l'ensemble des situations possibles, on se doit d'expliquer comment ces contextes, qui sont donc des formes inharmonieuses, peuvent être appris de manière convergente (au sens de la définition p. 119). Car pour chaque contexte, nous sommes bien en présence d'un cas de convergence, dans la mesure où différents individus sont parvenus à la même connaissance.

Dans la situation présente, telle qu'elle est décrite par les contraintes explicitées en page 120, on peut exclure un apprentissage direct des situations possibles selon un mode qui serait gestaltiste ou piagétien<sup>1</sup>. En effet, comme nous l'avons vu, de tels modèles font l'hypothèse de l'harmonie des formes apprises pour expliquer la convergence.

En fait, nous allons donner des arguments qui indiquent qu'aucun mécanisme indifférent ne peut expliquer l'apprentissage convergent des situations possibles. Nous suggérerons que cet apprentissage est indirect, qu'il passe par une manipulation de *connaissances*, et que celles-ci sont acquises par des mécanismes non indifférents.

Nous allons donner deux arguments pour exclure l'hypothèse d'un apprentissage direct des situations possibles par un mécanisme indifférent. Le premier est relatif au nombre de situations normales qui doivent être vécues par le sujet afin qu'il puisse exclure les situations impossibles, pour un contexte donné. Le second argument est relatif au fait que les individus connaissent des expériences différentes, ce qui, dans ce cas, devrait rendre la convergence impossible.

### 5.2.3. L'apprentissage direct des situations possibles est trop "difficile"

Si un système d'apprentissage indifférent, quel qu'il soit, est capable d'apprendre des ensembles de possibilités comme ceux que nous avons estimés dans les exemples précédents, alors il est aussi capable d'atteindre les ensembles de possibilités différents obtenus par isométries ( $5 \times 10^{13} / 20160 = 2.530.344.960$  ensembles différents dans notre premier exemple p. 124). Le manque de régularité des formes atteignables peut rendre la discrimination problématique pour certains systèmes indifférents, notamment les systèmes statistiques.

Le tableau de la page 112 montre qu'un système indifférent optimal aurait besoin d'au moins 3 exemples pour apprendre l'ensemble des possibles de l'exemple p. 124. Avec moins de 3 exemples, il ne pourrait pas déterminer l'ensemble à apprendre parmi tous ceux qui sont apprenables, car parmi ceux-ci il y a les 2.5 milliards obtenus par isométrie (ici  $N=13$  et  $\log_2 C(K) > \log_2(2.5 \times 10^{13}) > 31$ ).

Cette estimation, qui suppose que chaque bit de chaque exemple est utilisé pour la discrimination, est largement sous-évaluée. Par exemple, si l'on suppose que tous les ensembles de situations concernant les 13 propositions sont potentiellement apprenables, alors  $\log_2 C(K) = 2^{13} = 8192$ , et la formule de la page 112 permet d'estimer qu'un système optimal nécessiterait au moins 2088 exemples pour apprendre un contexte

<sup>1</sup> Nous ne remettons pas en question, ici, le modèle invoqué par Piaget pour expliquer l'acquisition de la logique, puisque Piaget fait intervenir la manipulation des connaissances logiques. Nous parlons ici d'une acquisition de connaissances contextuelles qui se résumerait à un ensemble de situations possibles.

comme celui de l'exemple p. 124 ! Mais on peut refuser le fait que tous les ensembles de situations soient apprenables...

On peut de toutes façons noter que tout sous-ensemble de l'ensemble de possibilités de l'exemple p. 124 peut être obtenu par l'adjonction de contraintes supplémentaires. Si l'on accepte que ces nouveaux ensembles sont des contextes acceptables, autrement dit qu'ils sont apprenables, alors le nombre d'exemples ne peut être inférieur au nombre de possibilités à apprendre (13 dans l'exemple).

L'apprentissage direct des situations possibles d'un contexte donné exige donc un nombre significatif d'exemples, ce qui semble exclure que l'on puisse connaître avec précision des contextes (et converser à partir d'eux) qui n'ont jamais été rencontrés auparavant ! Même notre estimation minimale de 3 exemples exclut la possibilité que la seule situation du moment permette de circonscrire l'ensemble des situations possibles, si cette opération est due à un mécanisme indifférent. Or nous sommes capables de circonscrire les contextes de situations que nous rencontrons pour la première fois. Par exemple, le premier dialogue SAVANT3 que nous avons utilisé ci-dessus (p. 124) portait sur des notions de filtrage qui sont nouvelles pour l'étudiant. Celui-ci est pourtant capable, dans la plupart des cas, de maîtriser le contexte du dialogue.

Une objection serait de dire que les interlocuteurs jugent de la possibilité de la situation présente  $s_o$  par analogie avec des situations rencontrées dans le passé. Ces autres situations ne seraient pas à compter parmi les exemples, puisqu'elles sont relatives à d'autres contextes (*i.e.* elles portent sur d'autres propositions), mais elle aideraient néanmoins à juger de la possibilité de la situation  $s_o$ .

Cette objection ne tient pas. L'hypothèse que nous sommes en train d'examiner, et qui est celle qui est généralement acceptée par ceux qui proposent un apprentissage statistique des situations, précise que les situations rencontrées sont par définition des situations possibles. Les situations passées, vécues par le sujet, sont donc des situations possibles. Le fait que certaines d'entre elles ressemblent à  $s_o$  ne donne aucune indication fiable sur la possibilité de  $s_o$ . En effet, deux situations sont analogues si une partie de leurs caractéristiques diffèrent (et une partie seulement) :

	1	2	3	4	5	6	7	8	9	10	11	12	13
situation $s_o$ :	0	1	1	0	0	0	0	1	-	-	-	-	-
situation $s_I$ :	-	-	-	-	1	0	0	1	1	0	0	0	0

Dans cet exemple, les situations  $s_o$  et  $s_I$  sont analogues parce que  $p_5, p_6, p_7, p_8$  sont des caractéristiques communes aux deux situations, et que trois d'entre elles ont la même valeur dans les deux cas. Peut-on dire que  $s_o$  est possible parce que  $s_I$  l'est ? Non, car l'impossibilité de  $s_o$  proviendra justement de telles caractéristiques divergentes<sup>(2)</sup>. Autrement dit, le fait qu'une situation soit analogue à une situation possible ne donne aucune indication sur sa propre possibilité<sup>(3)</sup>.

#### 5.2.4. La diversité des expériences individuelles

L'apprentissage indifférent des connaissances intervenant dans des contextes, directement à partir de situations rencontrées, pose bien d'autres problèmes. La convergence qu'il est facile de constater (stabilité des formes apprises d'un individu à l'autre) et la non-harmonie des formes apprises obligent à postuler une identité des expériences vécues par les individus (cf. les contraintes explicitées p. 120). Or il est manifeste que les individus parviennent à des connaissances très similaires sur un sujet donné tout en ne pouvant faire référence qu'à des expériences différentes. De plus, notre estimation du nombre d'exemples suppose que le contexte ne fasse toujours intervenir que les mêmes propositions. Or il est impossible que les situations similaires vécues qui ont permis l'apprentissage n'aient pas fait intervenir d'autres caractéristiques (lieu, saison, etc.) qui ne sont pas pertinentes dans le contexte présent.

<sup>2</sup> Dans cet exemple, l'analogie permet juste de dire que l'impossibilité de  $s_o$  ne viendra pas de la combinaison 0 0 1 pour  $p_6, p_7, p_8$ , ce qui n'est pas grand'chose.

<sup>3</sup> Il est clair qu'une situation analogue à une situation impossible a des chances d'être impossible. C'est sans doute cela qui donne le sentiment que l'analogie joue un rôle dans ce cadre. Mais cela revient à apprendre directement les incompatibilités. Or nous sommes en train d'examiner l'hypothèse alternative, selon laquelle les individus apprendraient l'ensemble des possibles, sans avoir accès directement aux incompatibilités.

Pour toutes ces raisons, il semble difficile de croire que l'apprentissage des situations possibles puisse être réalisé directement (*i.e.* à partir de la seule donnée des situations elles-mêmes), surtout par un mécanisme indifférent. En particulier l'apprentissage purement statistique des inférences ne peut pas mener à la connaissance des contextes des situations quotidiennes.

## 6. Conclusion

Nous avons consacré ce chapitre à l'étude des mécanismes d'apprentissage que nous avons définis comme *indifférents*, car ceux-ci sont souvent avancés pour rendre compte de l'apprentissage des connaissances. Nous avons montré que ces mécanismes avaient comme propriété principale d'apprendre facilement les ensembles *harmonieux*, et difficilement les ensembles *inharmonieux*. Nous avons ainsi pu interpréter certains modèles, notamment le modèle piagétien et le modèle gestaltiste, comme des mécanismes indifférents conduisant nécessairement à des formes harmonieuses.

Nous avons appliqué ce résultat à l'apprentissage des contextes conversationnels. L'apprentissage d'un contexte conversationnel se traduit par la capacité de juger si une situation est possible. Nous avons examiné l'hypothèse selon laquelle l'individu apprend, par un mécanisme indifférent, l'ensemble total des possibilités pour être en mesure d'effectuer ce jugement. Or nous avons trouvé qu'un tel ensemble de possibilités est *inharmonieux*. Il s'agit donc d'un apprentissage "trop difficile" pour un système indifférent.

L'abandon de cette hypothèse d'apprentissage direct des possibles a des conséquences qui nous paraissent très importantes. Toute théorie de l'apprentissage des situations possibles doit donc faire intervenir un mécanisme non indifférent. Dans le prochain chapitre, nous proposerons de tels mécanismes, par lesquels les individus manipulent des *connaissances explicites*, et qui leur permettent de *déterminer des connaissances à partir d'autres connaissances*. Pour déterminer l'ensemble des situations possibles dans un contexte donné, un individu utilise ainsi des connaissances portant sur des éléments non présents dans le contexte. Nous verrons que l'existence de tels mécanismes a des conséquences sur la forme des connaissances apprises, et elle suggère que certains d'entre eux, qui font intervenir les processus logiques observés dans les conversations, jouent un rôle prépondérant dans l'acquisition de ces connaissances.

**remerciements** : je remercie tout particulièrement Eric Bonabeau, Gérard Cohen et Olivier Hudry pour leur aide.

## Annexe 1 : Les isométries de l'hypercube

Il s'agit ici de déterminer l'ensemble des transformations de l'hypercube  $E$  qui conservent la distance de Hamming. Si  $\sigma$  est une telle transformation (isométrie),  $x$  et  $y$  deux éléments quelconques de l'hypercube de dimension  $N$ , et  $H(\cdot)$  la distance de Hamming, on a :

$$H(\sigma(x) - \sigma(y)) = H(x - y)$$

Notons que si l'on plonge l'hypercube dans  $\mathbb{R}^N$ , la distance de Hamming est égale à la distance euclidienne. Or toute transformation linéaire de  $\mathbb{R}^N$  qui applique l'hypercube sur lui-même est une isométrie euclidienne (et sera donc une isométrie pour la distance de Hamming). En effet, une telle transformation amène l'origine sur un point de l'hypercube (translation), et applique les vecteurs de la base de départ sur toutes les arêtes partant de ce point.

Ainsi, toute translation dans l'hypercube combinée avec une permutation des vecteurs de base est une isométrie au sens de la distance de Hamming. Ces transformations sont bien sûr toutes distinctes, et sont au nombre de  $2^N \times N!$ . Montrons qu'il n'y a pas d'autre isométrie, pour la distance de Hamming.

Soit  $\sigma$  une isométrie de l'hypercube. Nous allons montrer que si l'on choisit une image pour l'origine et les points adjacents, alors les images par  $\sigma$  de tous les points de l'hypercube sont automatiquement déterminées. Nous allons montrer par récurrence que l'image par  $\sigma$  de tout point à distance  $k$  de l'origine est déterminée univoquement.

Nous supposons que les images par l'isométrie  $\sigma$  des points à distance  $\leq k-1$  de l'origine sont déterminées univoquement. Soit  $x$  un point de l'hypercube à distance  $k$  de l'origine  $O$  :  $H(x) = k$ . Un chemin minimal  $\{y_i ; 0 \leq i \leq k\}$  est tel que :

$$y_0 = O ; y_k = x ; \forall i ; H(y_{i+1} - y_i) = 1$$

Parmi les  $k!$  chemins minimaux qui mènent de  $O$  à  $x$ , il y en a  $k$  d'extrémités  $y_{k-1}$  distinctes, car il y a  $k$  façons distinctes de choisir la dernière coordonnée qui doit passer de 0 à 1. Nous pouvons choisir trois chemins de ce type dès que  $k \geq 3$ .

Les images de ces trois chemins minimaux par l'isométrie  $\sigma$  sont des chemins minimaux menant de  $\sigma(O)$  à  $\sigma(x)$ . Leurs extrémités  $\sigma(y_{k-1}^j)$  sont déterminées de manière univoque par hypothèse de récurrence. Or ces trois points sont à distance 1 de  $\sigma(x)$  et à distance 2 les uns des autres. Ceci suffit à déterminer  $\sigma(x)$ . Les tableaux suivants montrent en effet que  $\sigma(x)$  est entièrement déterminé dans les deux seules configurations possibles qui lient les valeurs binaires  $a, b, c$  sur lesquelles les trois points retenus peuvent différer.

$\sigma(y_{k-1}^1)$	$\sigma(y_{k-1}^2)$	$\sigma(y_{k-1}^3)$	$\sigma(x)$	$\sigma(y_{k-1}^1)$	$\sigma(y_{k-1}^2)$	$\sigma(y_{k-1}^3)$	$\sigma(x)$
a	1-a	a	a	a	1-a	1-a	1-a
b	1-b	1-b	1-b	b	1-b	b	b
c	c	1-c	c	c	c	1-c	c

Il reste pour conclure à amorcer la récurrence en vérifiant que les points à distance 2 de l'origine ont des images bien déterminées par  $\sigma$ . Les points  $\sigma(y_1^1)$  et  $\sigma(y_2^2)$  sont connus, car images de points adjacents à l'origine. Le tableau suivant montre qu'il n'y a que deux possibilités pour  $\sigma(x)$ . Or l'une de ces possibilités est nécessairement l'image de l'origine. Il ne reste donc qu'une possibilité pour  $\sigma(x)$ .

$\sigma(y_1^1)$	$\sigma(y_2^2)$	$\sigma(x)$	$\sigma(x)$
a	1-a	1-a	a
b	1-b	b	1-b

Nous pouvons donc conclure que les isométries de l'hypercube, pour la distance de Hamming, sont les composées d'une translation et d'une permutation, et sont donc exactement au nombre de  $2^N \times N!$ .

## Annexe 2 : taille minimum d'un jeu d'exemples non ordonné

Pour donner une borne inférieure au nombre  $K$  d'exemples qui sont nécessaires pour qu'un apprentissage ait lieu, nous faisons la constatation très simple suivante : il doit y avoir plus de jeux de  $K$  exemples que formes apprenables. Si  $K$  est trop petit, un jeu d'exemples ne peut suffire à désigner, parmi les formes accessibles, celle qui doit être apprise.

Si  $C(K)$  est l'ensemble des formes apprenables, et s'il y a  $M$  exemples possibles, alors il y a  $M^K$  jeux distincts, et  $K$  vérifie :

$$M^K \geq \text{card}(C(K))$$

Dans le cas d'une représentation des exemples sur  $N$  bits, on a  $M = 2^N$ , et donc :

$$K.N \geq \log_2(\text{card}(C(K)))$$

Nous allons affiner cette minoration de  $K$  lorsque l'ordre des exemples est quelconque et ne peut pas servir à déterminer la forme à apprendre. Nous dirons que deux jeux sont équivalents s'ils diffèrent par une permutation de leurs exemples.

Intuitivement, on peut prévoir que le nombre de jeux non équivalents est le nombre total de jeux divisé par  $K!$ , ce qui conduit à :

$$M^K / K! \geq \text{card}(C(K))$$

Cela suppose toutefois qu'il y ait  $K!$  jeux dans chaque classe d'équivalence. Or ceci est faux, puisque certains jeux sont insensibles à certaines permutations, dans la mesure où ils contiennent des exemples répétés. Cependant, si  $K \ll M$ , on peut espérer que la plupart des jeux comportent peu d'exemples répétés, et que les classes contiennent près de  $K!$  éléments. Examinons cela d'une manière plus précise.

La constitution d'un jeu d'exemples correspond à un tirage (avec remise) de  $K$  objets parmi  $M$ . Lorsque l'ordre importe, il y a bien sûr  $M^K$  jeux possibles. Si, comme nous en faisons l'hypothèse, l'ordre est non significatif, le tirage équivaut à dire combien de fois on prend le premier exemple, combien de fois on prend le second, etc. jusqu'au nombre du  $M^{\text{ième}}$  exemple, ceci à concurrence de  $K$  éléments choisis.

Autrement dit, le nombre de jeux non ordonnés est égal au nombre de  $M$ -uplets  $\{m_i\}$  tels que :

$$\sum_{i=1}^M m_i = K$$

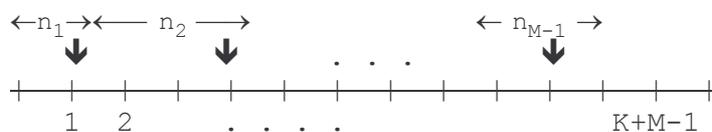
avec  $m_i \geq 0$ . Montrons que ce nombre est égal à :

$$\text{nombre de jeux non ordonnés} : \binom{M-1}{K+M-1}$$

Pour cela, posons  $n_i = m_i + 1$ . Il nous faut déterminer le nombre de  $M$ -uplets  $\{n_i\}$ , avec  $n_i > 0$ , tels que :

$$\sum_{i=1}^M n_i = K + M$$

Ce nombre est égal au nombre de façons de placer  $M-1$  "flèches" sur  $K+M-1$  positions :



On obtient bien le nombre annoncé. Ceci nous permet d'écrire la contrainte suivante sur  $K$  :

$$C_{K+M-1}^{M-1} \geq \text{card}(C(K))$$

Si  $M$  est suffisamment grand (supérieur à 10 ; or dans notre application,  $M=2^N$  avec  $N \geq 8$ ), on peut appliquer la formule de Stirling, ce qui permet d'écrire :

$$\begin{aligned} C_{K+M-1}^{M-1} &\cong \sqrt{\frac{M}{M+K}} \cdot \frac{(M+K)^{M+K}}{M^M K!} e^{-K} \\ &\cong \sqrt{\frac{M}{M+K}} \cdot \frac{M^K}{K!} e^{-K} \left[ e^{(K+K^2/M)} \right] \end{aligned}$$

Nous avons donc la condition :

$$\frac{M^K}{K!} e^{\frac{K^2}{M}} \geq \text{card}(C(K))$$

soit, si  $M=2^N$  :

$$\boxed{K \cdot N - \log_2 K! + \frac{K^2}{2^{N-1}} \geq \log_2 \text{card}(C(K))}$$

Cette formule est très proche de celle que nous donnions ci-dessus, car le terme  $K^2$  est négligeable devant  $2^N$ .

## Annexe 3 : Apprentissage statistique

### 1. Difficulté de l'estimation statistique d'une "boule"

Nous considérons un mécanisme d'apprentissage statistique d'une classification binaire dont la classe n°1 est une "boule" dans l'hypercube, c.à.d. qu'elle est constituée de tous les points à distance inférieure ou égale à  $R$  d'un point  $C$ . Nous cherchons à savoir combien d'exemples ce mécanisme statistique nécessite pour estimer  $C$  et  $R$ .

#### 1.1. Estimation du centre $C$ par l'estimation de la moyenne

Nous supposons que le mécanisme calcule le barycentre  $b$  réel des exemples  $x^i$  et choisit pour  $C$  le point du  $N$ -hypercube le plus proche de ce barycentre. Nous supposons que les  $x^i$  résultent d'un échantillonnage de la boule  $(C,R)$  (échantillons indépendants).

$$b = \frac{1}{K} \sum_{i=1}^K x^i$$

$b$  est un estimateur vectoriel sans biais de  $C$ . Estimons la distance euclidienne entre  $b$  et  $C$ . Nous chercherons ensuite la valeur de  $K$  qui rend cette distance inférieure à  $1/2$  avec une probabilité suffisante.

$$\begin{aligned} E(\|b - C\|^2) &= E\left(\sum_{n=1}^N \left(\frac{1}{K} \sum_{i=1}^K x_n^i - c_n\right)^2\right) \\ &= \frac{1}{K^2} \sum_{n=1}^N \sum_{i,j=1}^K E[(x_n^i - c_n)(x_n^j - c_n)] \\ &= \frac{1}{K^2} \sum_{n=1}^N \sum_{i,j=1}^K [E(x_n^i x_n^j) - c_n^2] \\ &= \frac{1}{K^2} \sum_{n=1}^N \sum_{i=1}^K [E(x_n^{i^2}) - c_n^2] \\ &= \frac{1}{K} \sum_{n=1}^N \sigma_x^2 \\ &= \frac{N\sigma_x^2}{K} \end{aligned}$$

Le centre  $C$  de la boule sera estimé correctement si le carré de la distance Euclidienne est en moyenne inférieur à  $1/4\lambda^2$ , où  $\lambda$  est un paramètre fixé dépendant de la loi de  $x$  (rappelons que le mécanisme choisit le point le plus proche de  $b$ , sachant que les points du  $N$ -hypercube sont à distance  $l$  de leurs voisins). On a donc ici une condition sur  $K$  :

$$K > 4N\lambda^2 \sigma_x^2$$

#### 1.2. Estimation du rayon $R$ par l'estimation de $\sigma_x$

Nous supposons la condition précédente garantie, autrement dit nous supposons que le centre  $C$  de la boule est estimé sans erreur. De plus, la connaissance de la loi de  $x$  permet de déterminer un paramètre  $\mu$  tel que  $R = \mu\sigma_x$ . Il s'agit donc d'estimer  $\sigma_x$  avec une erreur négligeable. Notons  $y = x - C$ .  $\sigma_x^2$  est estimé par :

$$\begin{aligned} v &= \frac{1}{K} \sum_{i=1}^K \|Y^i\|^2 \\ &= \frac{1}{K} \sum_{i=1}^K \sum_{n=1}^N Y_n^i{}^2 \end{aligned}$$

Noter que  $v$  est un estimateur sans biais. Le calcul de la variance de cet estimateur passe par le calcul de moments du quatrième ordre. Ceux-ci ne peuvent être estimés que lorsque l'on connaît la loi de  $x$ . Pour progresser, et par analogie avec le cas gaussien, nous supposons qu'ils sont de l'ordre de  $\sigma_x^4$ .

$$\begin{aligned} E(v^2) &= \frac{1}{K^2} \sum_{i,j=1}^K \sum_{n,m=1}^N E(Y_n^i{}^2 Y_m^j{}^2) \\ &= \frac{1}{K^2} \sum_{i=1}^K \sum_{n,m=1}^N E(Y_n^i{}^2 Y_m^i{}^2) \\ &\approx \sigma_x^4 + \frac{\alpha}{K} \sigma_x^4 \end{aligned}$$

Pour assurer une erreur négligeable sur le rayon, il faut assurer que la variance de l'estimateur précédent est inférieure à  $1/16\lambda^4$  :

$$K > 16\lambda^4 \alpha \sigma_x^4$$

Si l'on fait apparaître  $\mu$  dans la minoration précédente, on obtient la condition suivante pour le nombre minimal d'exemple permettant de déterminer une boule de rayon  $R$  :

$$K > 16\lambda^4 \alpha R^4 / \mu$$

Si l'on suppose que toute "boule" est susceptible d'être apprise, comme le rayon maximal d'une boule dans le  $N$ -hypercube est  $N$ , on constate que le nombre minimal d'exemples est de l'ordre de  $N^4$ .

## 2. Estimation de la corrélation entre deux variables

Nous supposons que l'on cherche à estimer la corrélation entre deux variables aléatoires gaussiennes centrées,  $X$  et  $Y$ , grâce à  $m$  tirages indépendants  $\omega_i$ . Nous notons  $\sigma_X$  et  $\sigma_Y$  les écarts-types respectifs de  $X$  et  $Y$ .

Le paramètre à estimer est la corrélation :

$$\rho = \frac{E(X(\omega) Y(\omega))}{\sigma_X \sigma_Y}$$

L'estimateur choisi est :

$$\hat{\rho} = \frac{\frac{1}{m} \sum_{i=1}^m X(\omega_i) Y(\omega_i)}{\sigma_X \sigma_Y}$$

On note que cet estimateur est sans biais :

$$E(\hat{\rho}) = \frac{\frac{1}{m} \sum_{i=1}^m E(X(\omega_i) Y(\omega_i))}{\sigma_X \sigma_Y} = \rho$$

Calculons sa variance  $E(\hat{\rho}^2) - \rho^2$  :

$$\begin{aligned}
E(\hat{\rho}^2) &= \frac{1}{m^2 \sigma_X^2 \sigma_Y^2} \sum_{i,j} E(X(\omega_i) Y(\omega_i) X(\omega_j) Y(\omega_j)) \\
&= \frac{1}{m^2 \sigma_X^2 \sigma_Y^2} \sum_{i,j} [ E(X(\omega_i) Y(\omega_i)) E(X(\omega_j) Y(\omega_j)) \\
&\quad + E(X(\omega_i) X(\omega_j)) E(Y(\omega_i) Y(\omega_j)) \\
&\quad + E(X(\omega_i) Y(\omega_j)) E(X(\omega_j) Y(\omega_i)) ]
\end{aligned}$$

car  $X$  et  $Y$  sont gaussiennes. Nous supposons les tirages distincts et indépendants :  $E(X(\omega_i) Y(\omega_j)) = 0$  si  $i \neq j$ .  
On a ainsi :

$$\begin{aligned}
E(\hat{\rho}^2) &= \frac{1}{m^2 \sigma_X^2 \sigma_Y^2} \left[ \sum_{i,j} \rho^2 \sigma_X^2 \sigma_Y^2 + \sum_i \sigma_X^2 \sigma_Y^2 + \sum_i \rho^2 \sigma_X^2 \sigma_Y^2 \right] \\
&= \frac{1}{m} [1 + \rho^2 + m\rho^2] \\
\sigma_{\hat{\rho}^2}^2 &= E(\hat{\rho}^2) - \rho^2 = \frac{1}{m} [1 + \rho^2]
\end{aligned}$$

L'erreur relative est :

$$e_r = \frac{\sigma_{\hat{\rho}^2}}{\rho} = \frac{\sqrt{1 + \rho^2}}{\rho \sqrt{m}}$$

Lorsque  $\rho$  est faible,  $e_r \cong \frac{1}{\rho \sqrt{m}}$ , et  $m$  doit être élevé pour garantir une erreur relative acceptable.



# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction : de la conversation à la cognition

L'étude des conversations spontanées nous a montré que les individus engagés dans une interaction verbale mettent en oeuvre des capacités cognitives qui semblent *a priori* complexes, mais qui néanmoins obéissent à un certain nombre de contraintes assez strictes. Mais comment caractériser exactement ces capacités cognitives ? Sont-elles spécifiques, ou générales ? Il est essentiel de mieux comprendre ces capacités cognitives liées à l'usage spontané du langage, parce qu'elles semblent centrales dans ce que la cognition humaine a d'original. De plus, nous comptons lier ces capacités à l'*apprentissage* des connaissances. Si l'on se rend compte du fait qu'une grande partie de ce que les adultes savent a été acquis à l'occasion de conversations, cette question devient cruciale, et débouche par exemple de nouveaux moyens d'enseigner à l'aide de systèmes artificiels.

Toutefois, dans ce chapitre, nous allons envisager l'apprentissage des concepts, non pas par le langage, mais grâce à l'expérience directe. Il est généralement admis que les individus acquièrent aussi une part importante de leurs connaissances grâce à leur expérience directe, c'est-à-dire grâce à la perception qu'ils ont des situations dans lesquelles ils sont plongés. Nous allons tenter de montrer les points suivants :

- Certaines connaissances acquises par l'expérience sont de même nature que celles qui sont échangées au cours des conversations
- L'apprentissage de ce type de connaissances suppose des capacités cognitives *spécifiques*, assurées par un mécanisme spécifique
- C'est ce mécanisme qui nous permet de converser
- Inversement; les mécanismes conversationnels jouent un rôle fondamental dans l'apprentissage des connaissances, même lorsqu'il n'y a pas de conversation et que le sujet apprend par lui-même !

Partant de la question de l'origine des connaissances qui nous permettent de converser, nous serons ainsi conduits à inverser le lien et à considérer que ce sont les processus conversationnels qui nous permettent d'acquérir des connaissances. Nous irons même plus loin, en suggérant que ces mêmes processus jouent un rôle dans la formation des nouveaux *concepts* qui entrent dans la composition des connaissances apprises. Mais commençons par tirer les conséquences d'une constatation toute simple : les humains sont capables de s'étonner.

## 2. La nature des connaissances conversationnelles

### 2.1. *L'étonnement et la réalité cognitive des connaissances*

L'observation des conversations quotidiennes montre que les interlocuteurs sont capables d'effectuer la discrimination des situations incohérentes. Cette capacité de juger de la cohérence des situations est fondamentale : elle est à la base de tout le comportement conversationnel et de la rationalité la plus élémentaire<sup>(1)</sup>. Ce que nous appelons ici *connaissances* est ce qui permet, entre autres, à l'individu de discriminer les situations normales des situations anormales (incohérentes) dans un contexte donné.

Nous laissons ainsi entendre que l'individu réalise le jugement de cohérence grâce à la manipulation de connaissances qu'il a acquises d'une manière que nous allons étudier. Mais après tout, pourquoi postuler l'existence de ces connaissances ? Ne pourrait-on pas supposer que l'individu a circonscrit l'ensemble des situations possibles directement grâce à son expérience ? Une situation incohérente serait simplement, dans ce cas, un "point aberrant" dans l'espace des situations, par rapport à l'ensemble des situations normales extrapolé à partir des situations vécues. Le savoir-faire de l'individu serait *implicite*, et résiderait dans cet ensemble de possibilités.

Or nous avons précédemment montré que ce genre d'apprentissage purement direct des situations possibles (chap. II.2) n'était pas imaginable. Nous rejetons donc catégoriquement l'hypothèse de l'apprentissage des possibles, ce qui nous contraint d'accorder une réalité cognitive *explicite* aux connaissances qui rendent possible le jugement de cohérence, puis de postuler l'existence de mécanismes spécifiques qui permettent à l'individu de fabriquer de nouvelles connaissances à partir de connaissances déjà possédées.

### 2.2. *Concepts et incohérence*

Lorsqu'une situation donnée (perçue ou évoquée) apparaît comme incohérente à un interlocuteur, il le fait généralement savoir aux personnes présentes, débutant ainsi une conversation sur le mode que nous avons appelé "paradoxal" (chap. I.1). Le caractère anormal de la situation est (ou peut être) explicité par ce locuteur comme la simultanéité de caractéristiques qu'il juge incompatibles.

On observe alors que les autres participants perçoivent correctement les éléments jugés incompatibles, ainsi que leur incompatibilité. Illustrons ceci avec l'extrait suivant :

[ex\_Lille]

En voiture à la sortie de Bruges, en Belgique, il s'agit de trouver la route qui mène à Lille afin de rentrer en France. Lille n'est indiquée nulle part, alors que d'autres villes (par ex. Rijsel), nécessairement petites puisqu'elles ne figurent pas sur notre carte, sont abondamment fléchées.

A : *Ils ont une ville d'un million d'habitants à 60 km et ils ne la flèchent même pas.*

Dans cet extrait, A fait part d'une incohérence qu'il perçoit. Pour lui, être une ville importante, être proche et ne pas être fléchée sont des caractéristiques incompatibles :

[ ville(V) & proche (V) & importante(V) & non fléchée(V) ] ⇒ F

Cette incompatibilité ne porte que sur certains des éléments de la situation. D'autres éléments, comme le fait qu'il fasse nuit, que cela se passe un samedi, que l'on soit à Bruges, sont bien présents à l'esprit des interlocuteurs, mais ils n'entrent pas dans l'incompatibilité et sont non pertinents en ce début de conversation.

Cette performance qui consiste à repérer le caractère incohérent d'une situation est à la base de nombreuses conversations. Il semble s'agir d'une performance spécifiquement humaine, car rien dans le comportement des animaux ne permet de supposer qu'ils sont sensibles comme nous à l'incohérence d'une situation. Certes, on pourrait admettre qu'ils sont capables de réagir à une situation inattendue (parce que

<sup>1</sup> Lorsqu'un individu n'est pas sensible aux incohérences élémentaires, il est considéré comme anormal. Le critère de cohérence est largement utilisé pour établir le diagnostic psychiatrique.

rare), mais cela est autre chose. Nous avons insisté ailleurs (voir chap. I.1) sur la différence qualitative entre incompatibilité et improbabilité, la première ne pouvant apparaître comme une limite (probabilité tendant vers 0) de la seconde. Il est par exemple possible de s'étonner, au sens de l'incompatibilité, d'une situation fréquente<sup>(2)</sup>. L'impossibilité logique est d'une autre nature que l'impossibilité statistique.

Lorsque, au cours d'une conversation, une situation est considérée comme anormale (bizarre, étonnante, incohérente ...), c'est parce que les interlocuteurs jugent *a priori* certaines de ses caractéristiques comme ne pouvant pas être simultanément présentes. On observe alors un accord entre les interlocuteurs sur les éléments de la situation qui sont incompatibles. Ce n'est pas la situation elle-même, avec tous les détails qu'elle comporte, qui est anormale. Le caractère incohérent porte sur certaines caractéristiques, toutes explicites, alors que d'autres caractéristiques de la situation n'entrent pas dans l'incompatibilité. Nous considérons donc cette incompatibilité comme *objective* dans le contexte de la conversation.

Le fait que l'incompatibilité, avec les éléments qu'elle regroupe, soit explicitée par les interlocuteurs, est *un fait expérimental* que l'on peut obtenir selon des méthodes opératoires (cf. expérience du "Et alors", chap. I.1). Nous défendons donc l'idée selon laquelle *l'incompatibilité, avec les éléments qu'elle regroupe, est effectivement représentée, de manière explicite, dans l'esprit du locuteur qui exprime son étonnement.*

**Concept logique** : nous appellerons concept logique toute représentation mentale entrant, ou pouvant entrer, dans une incompatibilité ressentie (i.e. capable de provoquer un étonnement).

**Connaissance conceptuelle** : une incompatibilité, qui peut provoquer un étonnement dans un contexte donné, constitue pour nous un élément de connaissance reliant des concepts logiques. Nous nous préoccupons ici de l'acquisition des connaissances conceptuelles ainsi définies.

Cette définition du concept (logique) a de quoi surprendre. La notion de concept recouvre généralement de nombreux aspects (perception, association, abstraction, etc.) qui sont absents ici. Mais notre définition utilise seulement une propriété caractéristique des concepts (logiques), celle de pouvoir entrer dans une relation d'incompatibilité. Elle est suffisante (et fournit un critère intuitif) pour distinguer les concepts (logiques) des autres représentations mentales. Par exemple une sensation de chaleur n'est pas un concept logique. En revanche, lorsque je m'étonne d'avoir chaud alors que le thermomètre indique 13°, alors nous devons considérer "avoir chaud" comme un concept logique, qui n'est pas la sensation elle-même.

Notre thèse est que la notion commune de *concept* (qui est généralement définie de manière très imprécise) coïncide avec notre notion de *concept logique*.

Cette idée peut choquer. Peut-on accepter que tout ce qui peut être "conçu" est susceptible d'entrer dans une relation d'incompatibilité ? Autrement dit tous les "concepts" sont-ils des concepts logiques ? Nous répondons par l'affirmative, et nous allons nous attacher à le montrer.

Pour cela, nous partirons de notre définition logique des concepts, qui présente l'avantage d'être opératoire : elle est ancrée sur une expérience cognitivement établie, celle de l'incompatibilité. Cette définition confère ainsi une réalité cognitive au concept logique, qui ne peut donc pas apparaître comme un artefact de modélisation. Nous verrons ensuite que les concepts logiques ont une *sémantique*, et que cette sémantique est liée à leur *ancrage* dans la perception, ce qui confèrera aux concepts toutes les propriétés généralement attribuées à la notion habituelle de concept. La nature de cet ancrage nous éclairera sur la nature cognitive (voire "physique") des concepts.

### 2.3. Concepts : c'est l'intension qui compte

On pourrait objecter que notre façon de caractériser les concepts ne correspond qu'à l'*intension* du concept (par opposition à son extension). Cela est tout à fait vrai, mais nous ne pouvions pas faire autrement

<sup>2</sup> Tous les trains de banlieue, même les express, qui circulent sur ma ligne, s'arrêtent dans cette gare où presque personne ne monte ni ne descend, alors que d'autres gares plus fréquentées sont sautées par les express. Le caractère incohérent de cette état de chose n'est pas atténué par l'habitude.

dans une définition. En effet, au niveau intensionnel, les relations entre concepts, comme la relation d'incompatibilité, apparaissent sans ambiguïté : elles fonctionnent en "tout ou rien". En revanche, il n'en est pas de même du niveau extensionnel. Les concepts logiques (par ex. *ville(.)* dans l'extrait précédent) n'ont pas de valeur définie (vraie ou fausse) sur tous les objets issus de la perception. Autrement dit, nous suggérons le fait qu'il n'est pas possible de définir les concepts logiques par leur *extension*.

En effet, les concepts, tels qu'ils interviennent dans les conversations, apparaissent plutôt comme des "outils" pour segmenter le monde perçu, outils bien imparfaits d'ailleurs (penser aux effets de typicalité étudiés par Rosch [1978]). L'opération inverse, qui conduirait d'une extension (réunion d'objets hétéroclites) à une intension, semble cognitivement peu vraisemblable telle qu'elle est décrite dans les modèles de catégorisation (cf. la critique de Medin [1989]). Il semble plutôt que les perceptions<sup>(3)</sup> servent à *évoquer* (*déclencher*) des concepts déjà forgés, ou *déclenchent* la formation de nouveaux concepts<sup>(4)</sup>. Mais comme notre définition l'indique, les concepts nouvellement forgés NE SONT PAS définis par la perception (toujours partielle, bruitée, etc.) qui a déclenché leur formation, mais par les liens logiques (que nous représentons par des incompatibilités) qui les relient à d'autres concepts. Ils héritent de leurs conditions d'évocation (imparfaite, floue, ...) à travers ces liens :

"nothing is a concept except by virtue of the way it is connected up with other things that are also concepts" [Hofstadter 1985:528]

Notre position concernant la représentation mentale des concepts est celle d'un "réalisme logique" :

**Réalisme logique** : nous supposons, dans notre modèle, que les concepts logiques et les relations d'incompatibilité qui les lient, dans la situation où l'incompatibilité est ressentie, sont matériellement représentés, de manière explicite, dans l'esprit des interlocuteurs<sup>(5)</sup>.

Certaines théories, comme la théorie cognitive connexionniste de P. Smolensky [1988], prennent soin de distinguer le modèle logique de la performance qu'il permet, et tentent de proposer des mécanismes non logiques capables de produire la même performance. La logique ne serait qu'un outil commode du modélisateur, un raccourci pour reconstituer la performance, mais elle n'aurait aucune plausibilité cognitive !

Malheureusement, ces théories a-logiques, qui sont bien entendues elles aussi des modèles, ne parviennent pas à reproduire une performance aussi naturelle pour les humains que celle qui consiste à détecter les situations incohérentes. La thèse selon laquelle les individus apprendraient l'ensemble des possibles, et qui ferait des incompatibilités un artifice de modélisation, est, selon nous, non fondée. Nous avons en effet montré que l'apprentissage direct des possibles n'est pas envisageable dans les situations concrètes (cf. chap. II.2).

Bien entendu, notre description des connaissances sous forme d'incompatibilités logiques est un modèle. Mais nous n'avons de toute façon accès qu'à des modèles. Notre "réalisme logique" vient de notre confiance dans la plausibilité cognitive de ce modèle, confiance que nous avons acquise en observant le comportement des interlocuteurs. Ceux-ci sont sensibles aux incohérences évoquées par leurs vis-à-vis, et ils identifient de manière très sélective les concepts impliqués dans l'incohérence<sup>(6)</sup>. La sensibilité à une incompatibilité se traduit par un étonnement durable, qui ne peut se résoudre que par une invalidation. Il nous semble donc parcimonieux de postuler une représentation cognitive de l'incompatibilité, d'autant que son implantation, même neuronale, ne pose pas de problème de principe (chap. II.4). De plus, rappelons-le, *aucun modèle non*

<sup>3</sup> Le mot *perception* est à prendre dans un sens plus large que le simple résultat d'une intégration sensorielle. Par ex., au théâtre, une situation impliquant trois personnages A, B, C peut déclencher chez le spectateur le concept de jalousie appliqué à A (*aime(A,B)*; *aime(B,C)*; *jalous(A,C)*).

<sup>4</sup> Notre thèse est bien sûr que nous forgeons la plupart de nos connaissances conceptuelles grâce au langage, à la faveur des conversations. Mais nous nous intéressons ici plus particulièrement à l'acquisition des concepts par l'expérience directe, et non par le langage.

<sup>5</sup> Ce réalisme logique ne suppose en aucune manière que les relations logiques sont présentes dans la nature, indépendamment de celui qui les perçoit. Bien au contraire !

<sup>6</sup> Le lecteur qui ne serait pas d'accord avec ces lignes trouvera qu'elles entrent en contradiction avec ce qu'il pense, et pourra certainement traduire son désaccord sous la forme d'une incompatibilité logique !

logique n'a été proposé qui rende compte avec succès de ce phénomène d'étonnement, tel qu'on peut l'observer par exemple dans les conversations sur le mode paradoxal.

**Résumé :** l'observation montre que différents individus sont capables du même étonnement durable face à des situations qu'ils rencontrent parfois pour la première fois. Au cours des conversations, ils explicitent spontanément leur étonnement sous la forme d'une incompatibilité. Nous considérons ces incompatibilités comme cognitivement représentées, et nous les regardons comme des unités de connaissance explicites reliant des concepts logiques. Le problème se pose alors de savoir comment ces relations conceptuelles sont forgées.

### 3. Les concepts sont-ils des perceptions intégrées ?

Notre définition du concept logique semble très éloignée de ce que l'on entend généralement par "concept". Mais est-ce vraiment le cas ? Nous examinons ici l'opinion, bien ancrée dans la tradition philosophique, selon laquelle les concepts ne seraient qu'une abstraction de représentations perceptives. Nous suggérons, comme certains auteurs, que les concepts sont qualitativement différents des représentations perceptives.

L'un des aspects les plus caractéristiques de la cognition humaine, celui dont beaucoup pensent qu'il nous distingue des autres êtres vivants, c'est notre *compréhension* des situations que nous percevons. Pour beaucoup d'auteurs, cette faculté est la quintessence d'une faculté perceptive qui n'aurait été à ce point développée vers l'abstraction que chez l'homme. Ces perceptions abstraites, très fortement intégrées, seraient ce que l'on appelle des *concepts*.

Les concepts, certains parlent d'idées, sont (pour l'instant) des objets purement cognitifs : seule l'introspection permet à chacun de se persuader qu'il manipule des représentations mentales du monde qui correspondent à ce que l'on appelle "concepts". Or ces représentations ont un nombre de propriétés impressionnant : elles peuvent se ressembler ; être apprises, mémorisées, généralisées, évoquées, spécialisées, combinées ; elles peuvent être véhiculées par le langage ; elles peuvent segmenter le monde perçu ; etc. Pour Douglas Hofstadter, les concepts sont intimement liés à une compétence humaine mystérieuse et néanmoins fondamentale, la créativité :

"The gist of my notion is that having creativity is an automatic consequence of having the proper representation of *concepts* in a mind. It is not something you add on afterward. It is built into the way concepts are. To spell this out more concretely : If you have succeeded in making an accurate model of *concepts*, you have thereby also succeeded in making a model of the creative process, and even of consciousness." [ . . . ] "the happy choice of the right concept at the right time is the essence of the creative." [Hofstadter 1985:528]

Dans quelle mesure les concepts sont-ils, ou ne sont-ils pas, des sortes de "super-perceptions", des perceptions fortement abstraites, débarrassées de leurs éléments contingents, comme les présente la tradition philosophique ? Les concepts semblent, il est vrai, posséder les propriétés des autres types de représentations, notamment les représentations perceptives. Les expériences des associationnistes, dans la première moitié de ce siècle, ont montré que des perceptions pouvaient évoquer d'autres perceptions (par exemple un air de musique en évoque un autre), ou évoquer (déclencher) des actions. De même des concepts peuvent évoquer d'autres concepts :

"[...] a concept consists of a central region surrounded by a dynamic, probabilistic 'halo' of potential associations. In its halo, 'driving' has such concepts as 'parking', 'getting stuck in traffic', 'having an accident', etc., each with a degree of association that changes in response to context." [Mitchell & Hofstadter 1990]

Les perceptions peuvent être plus ou moins abstraites (on peut avoir perçu le chat du voisin qui a une tache sur le cou, ou simplement, dans les mêmes circonstances, avoir perçu un chat) ; il en est de même pour une représentation d'action (on peut anticiper chaque détail d'un itinéraire ou bien simplement se représenter l'action de se diriger vers la destination). De même, les concepts peuvent être plus ou moins abstraits (c'est là

l'une de leurs caractéristiques les plus souvent relevées) : le concept d'être vivant est plus abstrait que celui de plante.

Le lien évident qui lie de nombreux concepts avec une réalité perçue, ainsi que la similitude entre les propriétés associatives des concepts et celles des perceptions, ont conduit beaucoup d'auteurs à considérer les concepts comme des *perceptions intégrées*. Pourtant les concepts semblent jouir de propriétés qui font défaut aux premiers niveaux de la perception, et ces propriétés semblent justement liées à la *signification* ainsi que le suggèrent J. Chalmers, R. French et D. Hofstadter :

"Perceptual processes form a spectrum, which for convenience we can divide into two components. Corresponding roughly to Kant's faculty of Sensibility, we have low-level perception, which involves the early processing of information from the various sensory modalities. High-level perception, on the other hand, involves taking a more global view of this information, extracting *meaning* from the raw material by accessing concepts, and making sense of situations at a conceptual level." [ . . . ] "High-level perception begins at that level of processing where *concepts* begin to play an important role." [Chalmers et al. 1991].

Doit-on considérer (comme nous le suggérerons dans ce chapitre) que les concepts sont des représentations distinctes des perceptions intégrées ? Nombreux sont les auteurs qui ressentent au moins la nécessité de considérer *deux* types de traitement cognitifs qui semblent qualitativement différents. C'était le cas dans la citation précédente, mais aussi chez Locke, qui distingue les idées issues de la sensation et celles qui proviennent de la réflexion<sup>(7)</sup> :

"First, our senses, conversant about particular sensible objects, do convey into the mind several distinct perceptions of things, according to those various ways wherein those objects do affect them. [...] This great source of most of the ideas we have, depending wholly upon our senses, and derived by them to the understanding, I call SENSATION.

Secondly, the other fountain from which experience furnisheth the understanding with ideas is the perception of the operations of our own mind within us, as it is employed about the ideas it has got; [...] I call this REFLECTION." [John Locke 1690]

On constate que même chez un empiriste, on trouve l'idée que certains concepts ne sont pas issus de l'expérience directe. Cette idée sera bien sûr acceptée par les rationalistes, pour qui les concepts sont avant tout les objets de la raison, avant d'avoir un lien avec les percepts. Une version récente du point de vue rationaliste est fournie par Jerry Fodor [1975] qui accorde aux concepts un rôle avant tout syntaxique, indépendant *a priori* des perceptions (ce qui rend d'ailleurs problématique l'origine de leur sémantique si celle-ci est, comme on s'accorde à le croire, liée à la perception [Harnad 1990]).

Le point de vue que nous voulons défendre ici consiste à accorder aux concepts un statut spécial, distinct de celui des perceptions. Même si les concepts ont des propriétés associatives et d'abstraction que l'on reconnaît aux perceptions, même si certains concepts sont intimement liés aux perceptions (évocation par et de certaines perceptions), l'existence de mécanismes spécifiques qui opèrent sur les concepts et non sur les perceptions va nous amener à considérer que les concepts sont qualitativement distincts des perceptions, et à postuler l'existence d'un appareil conceptuel relié d'une manière bien précise aux appareils perceptifs. C'est au niveau de cet appareil conceptuel que le sentiment d'une *compréhension*, d'une *signification* des situations, prend sa source. En d'autres termes, nous ne comprenons pas nos perceptions, nous ne pouvons comprendre que des relations conceptuelles<sup>(8)</sup>.

Pour justifier ce point de vue, nous partons du rôle joué par les concepts dans le langage spontané. Les contraintes qui limitent l'emploi des concepts dans ce type de communication va nous éclairer sur leur fonctionnement cognitif.

<sup>7</sup> Nous avons donné une citation plus complète de ce passage au début du chapitre II.1.

<sup>8</sup> Ce point de vue est assez proche de celui de Chalmers, French et Hofstadter qui donnent un critère définissant la perception supérieure par la sémantique qu'elle met en oeuvre : "The distinguishing mark of high-level perception is that it is semantic : it involves drawing *meaning* out of situations. The more semantic the processing involved, the greater the role played by *concepts* in this processing, and thus the greater the scope for top-down influences." [Chalmers et al. 1991].

## 4. Représentation cognitive des concepts logiques

### 4.1. Deux types de mécanismes

Comme nous l'avons indiqué, de nombreux auteurs définissent les concepts avant tout comme des représentations issues de la perception. Ils reconnaissent toutefois la possibilité que des concepts soient combinés entre eux, par exemple pour donner de nouveaux concepts, et acceptent même que ces combinaisons soient dans certains cas de nature logique. Mais un tel schéma nécessite *deux* types différents de mécanismes pour expliquer la formation des concepts. D'une part des mécanismes *conceptuels* qui opèrent sur les concepts, par exemple des mécanismes d'analogie ou d'abstraction ou encore de type logique, et d'autre part des mécanismes *externes*, faisant intervenir la perception.

L'existence de mécanismes externes, liés principalement à la perception, amène certains auteurs à décrire les concepts comme des *perceptions intégrées*. Ce sont des mécanismes externes, les mécanismes d'intégration perceptive, qui seraient supposés expliquer, par exemple, la formation des tout premiers concepts qui, forgés dès les premières semaines de la vie, vont pouvoir être ensuite combinés par les mécanismes conceptuels.

L'utilisation de ces perceptions intégrées pour engendrer une action et leur association à des symboles ont fait l'objet de modélisations dans le cadre des systèmes hybrides symboli-connexionnistes [Grumbach 93]. Ces perceptions intégrées peuvent être combinés (par exemple associés selon leur fréquence de cooccurrence) pour donner de nouvelles représentations du même type.

Toutefois, selon notre point de vue, ces représentations issues de la simple intégration des perceptions sont d'une nature qualitativement différente de celle des concepts logiques. En particulier, nous montrerons qu'aucune combinaison de perceptions ne peut constituer un concept logique. Ceux-ci ne sont donc pas issus de celles-là. En d'autres termes, nous réfutons une vision purement ascendante de la formation des connaissances conceptuelles.

Pour caractériser cette différence qualitative entre les perceptions intégrées et les concepts logiques, nous allons montrer que les premières NE PEUVENT PAS entrer dans des relations de type logique. Appelons  $\mathcal{M}_p$  le mécanisme externe qui forgerait un concept à partir d'un ensemble de perceptions, et  $\mathcal{M}_c$  un mécanisme logique qui opère sur des concepts (par ex. pour former un autre concept) :

$$\begin{array}{ccc} & \textit{perceptions} & \textit{concepts} \\ \mathcal{M}_p : & \{P_1, P_2, \dots, P_n\} & \rightarrow C \\ & & \\ & \textit{concepts} & \\ \mathcal{M}_c : & \{C_1, C_2, \dots, C_m\} & \rightarrow C' \end{array}$$

Les mécanismes  $\mathcal{M}_p$  qui sont couramment proposés dans un tel contexte sont des mécanismes d'extraction qui ont la propriété d'être "indifférents" (dans le sens que nous avons donné à ce terme dans le chapitre II.2), c.à.d. qu'ils admettent tout changement de repère dans le codage de leurs entrées. Ce seront par exemple des mécanismes statistiques, et le "concept"  $C$  construit apparaîtra comme un invariant perceptuel. Si tous les concepts provenaient ainsi directement ou indirectement de perceptions, alors nous devrions pouvoir écrire :

$$C' = \mathcal{M}_c( \{ \mathcal{M}_p( \{P_{\alpha\beta}\} ), \mathcal{M}_c( \{C_\gamma\} ) \} )$$

(où  $\{P_{\alpha\beta}\}$  est l'ensemble de perceptions qui formerait le "concept"  $C_\alpha$  par  $\mathcal{M}_p$ ). Ceci ne serait possible que si  $\mathcal{M}_c$  pouvait opérer sur les résultats de  $\mathcal{M}_p$ . Or nous verrons que cela n'a rien d'évident. En particulier, dans

le cas où il peut être considéré comme *systématiquement* applicable,  $\mathcal{M}_c$  devrait pouvoir opérer sur tous les résultats *potentiels* de  $\mathcal{M}_p$ . Nous allons essayer de montrer dans cette section (4.) que cela n'est pas possible, et que la formule ci-dessus ne peut pas avoir de sens si  $\mathcal{M}_p$  est indifférent.

Bien entendu, ce problème serait sans objet pour quelqu'un qui n'admettrait pas l'existence de mécanismes conceptuels de type logique  $\mathcal{M}_c$ , capables d'opérer sur les concepts. Nous avons rejeté cette position, car elle est manifestement contredite par le mécanisme de la détection d'incohérence dont nous devons accepter la réalité cognitive.

Non seulement de tels mécanismes logiques existent, mais certains d'entre eux permettent d'expliquer la formation de nouveaux concepts. Par exemple, je peux définir le concept de pamplemousse noir, à partir des concepts "pamplemousse" et "couleur noire". Cette chose, à propos de laquelle je peux certainement m'étonner (si on en vend depuis dix ans, pourquoi cette variété m'est-elle inconnue ?) est bien un concept, qui n'est sûrement pas extrait de mes perceptions. Ici, un nouveau concept est formé à partir de concepts existants.

Il existe d'autres mécanismes conceptuels. Considérons le mécanisme de la négation. Certains peuvent "nier" la pertinence cognitive de ce que l'on modélise par la négation logique, ce mécanisme qui permet de produire *non C* à partir d'un concept *C*. Nous ne le pouvons pas. En effet, les interlocuteurs ont besoin de concepts niés pour exprimer leur étonnement au cours des conversations. Par exemple, ayant compris que la couleur blanche confère un camouflage à un animal qui vit sur la neige, on pourra s'étonner que tel animal ne soit pas blanc, autrement dit qu'il soit *non blanc*, alors qu'il vit sur la neige et qu'il constitue une proie potentielle. Dans le contexte de l'animal sur la neige, (*non blanc*) est donc bien un concept, selon notre définition<sup>(9)</sup>.

Ce qui importe ici est que *non C* a une représentation mentale, et que mécanisme  $C \rightarrow \text{non } C$  doit donc nécessairement correspondre à ce que nous avons appelé un mécanisme conceptuel, opérant à partir des concepts de manière systématique. Nous allons examiner les conséquences de ce constat.

**Résumé :** nous avons défini les concepts logiques, d'autres définissent des concepts qui seraient de simples perceptions intégrées. Certains concepts logiques ne sont pas des perceptions intégrées. Nous comptons montrer qu'*aucun* concept logique n'est une perception intégrée ! Pour cela, nous allons montrer que les mécanismes conceptuels logiques (l'incompatibilité, la négation, etc.) peuvent opérer systématiquement sur *tous* les concepts, ce qui serait impossible si ces concepts *pouvaient* être issus de mécanismes perceptifs.

## 4.2. *Unicité des mécanismes conceptuels*

Les mécanismes que nous avons appelés conceptuels opèrent par définition sur des concepts logiques. Nous avons donné l'exemple de la négation. Notre objectif, ici, est de montrer que ces mécanismes sont matériellement uniques (au sens de la définition ci-dessous). Pour cela, nous allons utiliser le fait qu'il existe des mécanismes conceptuels qui ont la propriété d'être systématiques.

Considérons le mécanisme de la recevabilité d'une explication. L'observation des conversations montre que les individus considèrent comme recevable (ce qui ne signifie pas qu'ils l'acceptent) toute tentative d'explication d'un étonnement qui consiste à énoncer la négation de l'un des termes de l'incompatibilité qui provoque l'étonnement. Par exemple, si l'étonnement peut se traduire par l'incompatibilité suivante :

$$[p_1 \ \& \ p_2 \ \& \ \dots \ p_n] \Rightarrow \mathbf{F}$$

alors *non p<sub>i</sub>* est une tentative d'explication recevable. Dans l'exemple de la page 138, chaque prémisse peut être niée pour tenter une explication ("Lille" ne serait pas une ville, il s'agirait d'un groupe de villes<sup>(10)</sup> ; Lille ne serait pas une ville très importante ; Lille serait beaucoup plus éloignée que 60 km ; Lille serait

<sup>9</sup> Nous verrons que le statut des concepts niés est un peu différent. Pour la clarté de l'exposé, nous préférons nous en tenir à notre définition du concept qui inclut tout terme pouvant entrer dans une incompatibilité. Les concepts niés apparaissent donc comme des concepts. Mais ceci est sans incidence sur la description du mécanisme  $C \rightarrow \text{non } C$ .

<sup>10</sup> Comme St Quentin en Yvelynes.

fléchée<sup>(11)</sup>). Ce phénomène est systématique : *non p<sub>i</sub>* sera éventuellement "combattu", mais son énoncé ne provoquera jamais de réaction d'incompréhension (*i.e.* une réaction du type "pourquoi dis-tu cela?"). Il faut se rappeler qu'une conversation sur le mode paradoxal (le mode de l'étonnement) est très fortement contrainte (chap. I.1). Seuls un co-étonnement ou une invalidation directe ou indirecte sont admissibles (recevables). Nous parlons ici de la recevabilité systématique de l'explication par invalidation directe, dans le sens suivant :

**Mécanisme à valeur systématique** : un mécanisme  $\mathcal{M}$  est à valeur systématique par rapport à un sous-ensemble  $\mathcal{V}$  de son ensemble de départ s'il est constant sur  $\mathcal{V}$ .

Nous pouvons modéliser le mécanisme de recevabilité précédent par une fonction booléenne  $\mathcal{R}$  opérant sur une liste de concepts :  $\mathcal{R}([q, p_1, p_2 \dots p_n]) = 1$  ssi *non q* est une invalidation directe de  $[p_1 \& p_2 \& \dots p_n]$   $\Rightarrow$  **F**.  $\mathcal{R}$  vaut systématiquement 1 sur  $\mathcal{V} = \{[q|L] \text{ t.q. } q \in L\}$  <sup>(12)</sup>.

Dès que le caractère systématique d'un processus cognitif a été reconnu, il y a de très fortes contraintes sur la façon dont on peut matériellement le modéliser. Il semble en effet inévitable de conclure qu'un mécanisme systématique comme la recevabilité d'une explication est un mécanisme matériellement unique. La systématisme rend totalement improbable la multiplicité du mécanisme, car un mécanisme multiple n'est pas assez fiable :

**Unicité matérielle** : un processus peut être réalisé par un ensemble de mécanismes équivalents (mécanisme matériellement multiple), ou par un mécanisme unique (mécanisme matériellement unique). Nous considérons que la systématisme d'un processus est incompatible avec sa multiplicité, en raison de l'imperfection et du manque de fiabilité de tout mécanisme multiple.

Certains processus sont réalisés par des mécanismes matériellement multiples. Par exemple, les premiers "étages" de la perception visuelle ou auditive sont le fait de nombreux mécanismes identiques (*e.g.* les cellules ganglionnaires de la rétine pour le traitement des zones on/off, les cellules ciliées de la cochlée, les colonnes corticales de l'aire visuelle primaire, etc.). La fiabilité de ce genre d'organisation est très loin d'être totale (penser à la tache aveugle correspondant à la traversée du nerf optique, ou aux surdités sélectives dans le traitement des fréquences par la cochlée). Nous ne connaissons aucune organisation biologique multiple dont la fiabilité puisse approcher la perfection<sup>(13)</sup>. Noter que la multiplicité matérielle ne doit pas être confondue avec une redondance. Au contraire. Des mécanismes équivalents qui, ensemble, réalisent un processus assurent des traitements distincts (penser à la vision). La défection d'un seul d'entre eux affecte inévitablement le processus.

Le mécanisme de reconnaissance d'une explication doit donc, du fait de sa systématisme, être considéré comme matériellement unique. Mais cela va poser, par voie de conséquence, le problème de la "connexion" (au sens matériel) entre de tels mécanismes conceptuels matériellement uniques et les représentations matérielles des concepts.

### 4.3. Nécessité d'un appareil conceptuel

L'exigence d'unicité matérielle concerne non seulement les mécanismes conceptuels systématiques eux-mêmes, mais aussi leur *prise en compte* des concepts, comme indiqué schématiquement sur la figure 1.

<sup>11</sup> C'est cette dernière explication qui s'est révélée être la bonne : Lille était fléchée en Flamand, sous le nom de Rijsel !

<sup>12</sup> Nous utilisons la notation Prolog :  $[q|L]$  est la liste résultant de l'adjonction de l'élément  $q$  en tête de la liste  $L$ .

<sup>13</sup> Les imperfections des mécanismes perceptifs peuvent être masquées, et n'apparaître que dans des cas un peu particuliers (*e.g.* sifflement d'un poste de télévision, non perçu par certains). Cet argument du masquage des imperfections ne tient pas pour les mécanismes conceptuels : on n'a jamais décrit d'individu qui, étant par ailleurs normal, présente un défaut de systématisme dans l'application d'un mécanisme logique comme la recevabilité d'une explication (cf. le "bon sens" de Descartes).

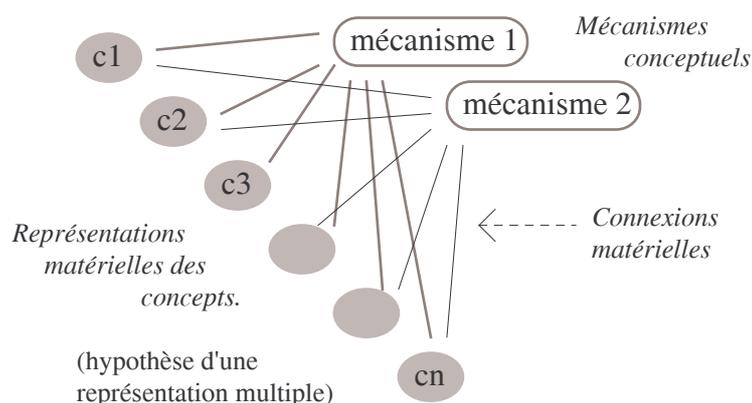


figure 1 : Représentation schématique de la problématique de l'unicité matérielle. Certains mécanismes conceptuels ne peuvent être que matériellement uniques. Mais si les représentations matérielles des concepts étaient, elles, physiquement distinctes, comme indiqué sur la figure, alors les liens qui relient ces représentations aux mécanismes seraient multiples. Or ceci est absurde si la *prise en compte* des concepts par l'un des mécanismes est systématique.

Pour fixer les idées, considérons l'expérience suivante. On demande à des sujets d'appuyer sur un bouton chaque fois que la carte qu'on leur présente, prise dans un paquet de cartes comportant des dessins (ou des mots, ou des rébus, etc.), représente (ou évoque), disons, un cheval. Les cartes peuvent être choisies pour que la tâche ne présente aucune ambiguïté. Tout sujet normal résout parfaitement la tâche, du premier coup.

Personne ne mettra en doute le fait que la performance du sujet a un aspect systématique : tous les concepts connus du sujet (ou du moins, dans ce protocole, ceux que l'on parvient à évoquer facilement à l'aide d'une carte) seront *considérés*, qu'ils conduisent ou non à l'appui du bouton (si l'on préfère une action systématique, il est possible de prévoir trois boutons : "oui", "non", "sans opinion"). On ne peut pas imaginer un sujet qui formerait correctement le concept de "chat" en voyant une carte, mais qui resterait dans l'incapacité totale de prendre une quelconque décision.

Personne n'a jamais décrit d'individu (normal) capable de *considérer* un concept pour un traitement donné (par ex. pour le recevoir comme une explication), et qui serait incapable de le *considérer* pour un autre traitement (par ex. l'expérience du paquet de cartes ci-dessus). Ce phénomène banal est pourtant fondamental. Il met en évidence le fait que des processus conceptuels, comme celui du test ci-dessus, ont une *prise en compte systématique des concepts évoqués par la situation*. Il en est de même de la négation qui, elle aussi, est *systématiquement applicable* : pouvons-nous concevoir un concept logique  $C$  tel que *non C* ne puisse pas avoir de représentation cognitive (par ex. en apparaissant au sein d'une incompatibilité) ?

**Mécanisme systématiquement applicable** : un mécanisme conceptuel  $\mathcal{M}$  est systématiquement applicable s'il est une application (au sens mathématique) ; toute configuration des concepts sur lesquels  $\mathcal{M}$  opère conduit à un résultat.

Un mécanisme comme la recevabilité d'une explication (voir p. 144), que nous avons noté  $\mathcal{R}()$ , a été décrit comme systématique et donc comme matériellement unique. Nous constatons maintenant qu'il est systématiquement applicable. Dans un contexte où  $[p_1 \& p_2 \& \dots p_n] \Rightarrow F$  est une incompatibilité ressentie,  $\mathcal{R}([q, p_1, p_2 \dots p_n])$  a une valeur booléenne (recevable / non recevable) *pour tout concept q* connu du sujet. La figure 2 résume la situation.

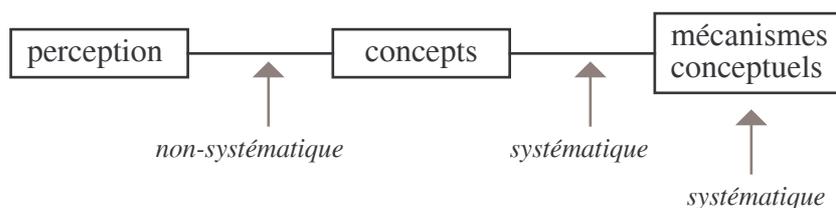


figure 2: représentation schématique des aspects systématiques dans le traitement cognitif des concepts. La systématique se situe au niveau de certains mécanismes conceptuels, ainsi qu'au niveau de la prise en compte des concepts par ces mécanismes. En revanche, l'évocation des concepts par les perceptions n'a aucun caractère systématique.

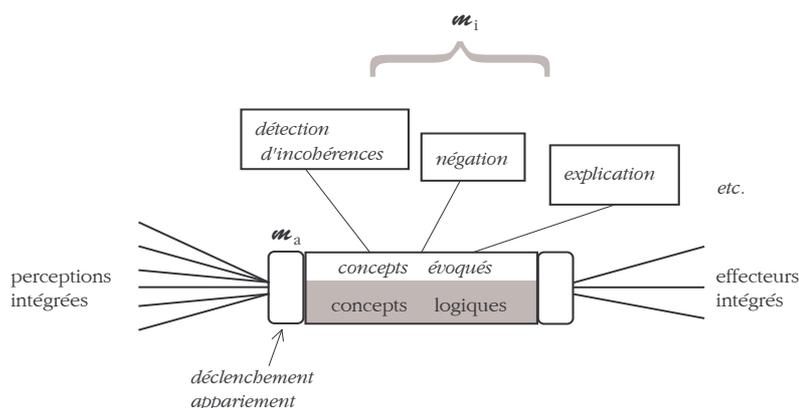
Il existe donc des mécanismes matériellement uniques systématiquement applicables, par exemple le mécanisme de recevabilité d'une explication. Ce constat a des conséquences fondamentales. Notre point de vue sera encore une fois le même : un processus systématique est forcément unique matériellement. Cette règle s'applique donc non seulement aux mécanismes conceptuels systématiques, mais aussi aux liens qui les "relient" aux concepts. En d'autres termes, le schéma de la figure 1 n'est pas acceptable. La défaillance d'un seul des liens représentés sur cette figure aurait des conséquences qui n'ont jamais été observées chez aucun individu.

Dans ces conditions, comment imaginer l'implantation matérielle de la prise en compte des concepts par un mécanisme conceptuel systématique ? Nous ne voyons qu'une seule solution :

*Tous les concepts logiques sont hébergés par un seul système et partagent la même représentation matérielle.*

Certes, il pourrait sembler plus "naturel" d'accorder aux concepts des représentations matérielles distinctes, notamment si l'on croit que certains concepts sont construits par la perception sensorielle qui, elle, est assurée par des mécanismes multiples. Malheureusement, une telle hypothèse ne résiste pas à l'exigence de systématique. On ne peut imaginer que la *prise en compte* d'un concept, dans un test comme celui que nous avons décrit ci-dessus, soit laissée aux aléas d'une connexion matérielle. Pour un traitement donné, les sujets seraient totalement et inexorablement "aveugles" à certains concepts, alors qu'ils pourraient utiliser ces concepts dans d'autres traitements !

Nous voilà donc contraints de postuler l'existence d'un "appareil conceptuel" matériellement unique, alimenté par la perception (voir notes 3 et 4 p. 140), mais distinct des appareils perceptifs. Cet appareil spécialisé hébergerait tous les concepts logiques connus ou potentiels en une même représentation neuronale (figure 3).



**figure 3:** schéma de l'appareil conceptuel censé abriter les représentations matérielles de tous les concepts. Il est relié par des connexions multiples (non systématiques) aux appareils perceptifs (perceptions intégrées) et aux appareils effecteurs intégrés (pour le déclenchement des actions). Les mécanismes conceptuels systématiques  $\mathcal{M}_c$  opèrent dans cet appareil conceptuel.

Cette hypothèse concernant l'existence d'un système conceptuel matériellement unique permet de résoudre le problème de la systématisme des mécanismes opérant sur les concepts. De tels mécanismes, nécessairement uniques (au sens de l'unicité matérielle), peuvent avoir une prise en compte systématique des concepts évoqués par la situation, car ceux-ci partagent une même représentation matérielle.

Cet appareil conceptuel serait relié aux appareils perceptifs par des voies multiples. Cette fois, l'argument de la systématisme ne joue plus car la prise en compte des perceptions n'a aucune raison d'être systématique. De même, dans le test imaginé plus haut, l'action prescrite (appuyer sur le bouton) aurait une représentation dans cet appareil conceptuel, et l'appareil serait relié de manière non systématique aux effecteurs intégrés<sup>(14)</sup>.

Selon ce modèle, les concepts logiques sont des représentations internes à l'appareil conceptuel. C'est ce que nous voulons dire en notant qu'ils partagent la même représentation matérielle. Ceci indique qu'ils ne peuvent pas être des perceptions intégrées. Nous avons annoncé ce résultat au début de cette section 4. en disant que les mécanismes conceptuels logiques  $\mathcal{M}_c$ , capables de combiner des concepts, ne pouvaient pas opérer sur des "concepts perceptuels" produits par un mécanisme d'intégration indifférent  $\mathcal{M}_p$ . Dans la figure précédente, nous avons donc remplacé  $\mathcal{M}_p$  par un mécanisme d'appariement  $\mathcal{M}_a$  qui permet aux représentations internes de l'appareil conceptuel d'être déclenchées par des perceptions. Nous allons tenter de justifier ces différents points en définissant l'ancrage des concepts logiques.

## 5. La nécessité d'ancrer les concepts logiques

### 5.1. Le problème de l'ancrage

Dans notre modèle, nous avons défini les concepts logiques sans utiliser de référence à la perception, et nous leur avons assigné une représentation matérielle différente de celle des perceptions. Tout cela pose le problème de la *sémantique* des concepts logiques, autrement dit de leur ancrage dans la réalité perçue, qui est un cas particulier du problème de l'ancrage des symboles soulevé par Harnad [1990].

En effet, des processus systématiques, comme celui de la recevabilité d'une explication rappelé ci-dessus, sont de type syntaxique, c'est-à-dire qu'ils agissent sur des représentations d'après leur forme et non d'après

<sup>14</sup> Ces connexions nombreuses, complexes et bidirectionnelles avec les appareils perceptifs et les effecteurs empêchent de considérer que cet appareil conceptuel, bien qu'unique et chargé d'une fonction particulière, puisse être un module au sens de J. Fodor [1983]. Il est même parfaitement "pénétrable", contrairement aux modules.

leur contenu. Par exemple une explication *non q* dans le contexte de l'incompatibilité  $[p_1 \& p_2 \& \dots p_n] \Rightarrow F$  sera jugée recevable dès que  $q = p_i$ , quel que soit ce que représentent  $q$  et les  $p_i$ . Or si les processus ne tiennent pas compte des concepts eux-mêmes, mais seulement de leurs relations formelles, l'appareil conceptuel risque de donner l'impression d'héberger des symboles déconnectés de la réalité, et donc d'être cognitivement invraisemblable.

Cette question de l'ancrage des symboles nous semble fondamentale, et elle doit être traitée dès que l'on a reconnu la non-identité entre les symboles et ce qu'ils sont censés "signifier". Il serait en effet artificiel de s'intéresser aux processus formels eux-mêmes sans avoir compris l'origine cognitive des symboles sur lesquels ils opèrent, et c'est sans doute ce genre de démarche qui a donné l'impression à certains que l'ancrage de tout symbole engendré par une manipulation formelle était impossible.

Or notre définition du concept logique (voir p. 139) est interne à l'appareil conceptuel. Elle ne fait aucune référence à un quelconque ancrage, et nous risquons de tomber dans le travers que nous venons de dénoncer. Ce serait oublier que cette définition n'indique pas comment les concepts sont formés ni quelle est leur fonction, mais donne simplement un moyen de les reconnaître parmi les représentations utilisées par l'individu. Une description plus complète de ce que sont les concepts va nous permettre de comprendre comment ils sont ancrés.

## 5.2. Les concepts logiques vus comme des détecteurs

Nous avons défini les concepts logiques d'une manière "syntaxique", par l'une des propriétés qui les distingue des autres représentations mentales : leur capacité à entrer dans des relations d'incompatibilité. Nous avons cependant clairement indiqué que ces concepts avaient une *sémantique*, liée à leurs conditions d'évocation. Le moment est venu d'étudier ces conditions d'évocation, car cela va nous permettre de progresser dans la compréhension du mode de formation des concepts logiques.

Nous avons indiqué que les concepts (logiques) n'étaient pas des perceptions intégrées. Quel est donc leur lien avec les perceptions ? Il existe, bien sûr. Ce lien est tellement manifeste que beaucoup croient pouvoir définir les concepts par leur prétendue "origine" perceptive. Le dilemme auquel nous sommes confrontés est le suivant :

- soit on considère que les concepts sont des perceptions intégrées, ce qui entraîne que les mécanismes logiques, du fait de leur systématisme, manipulent autre chose que des concepts !
- soit on considère que les concepts sont ce que manipulent les mécanismes logiques, mais alors quel sont les liens entre les concepts et les perceptions ?

Notre point de vue est sans ambiguïté : la première de ces deux façons de voir ne nous semble pas tenable. En revanche, il est facile de voir le lien *percept-concept* comme une opération de *détection*.

**Détection** : un détecteur est un mécanisme capable d'entrer en activité pour certaines configurations de ses entrées. Nous présentons ici les concepts comme des détecteurs. Un concept présent dans une incompatibilité ressentie correspond à un détecteur activé. ||

Le phénomène de détection est fondamental, car il semble intervenir à tous les niveaux de l'organisation biologique. Dans [Dessalles 1992c], nous avons décrit la perception comme structurée en niveaux de détection, ce qui nous a permis d'expliquer le fait que les objets perçus sont eux-mêmes structurés en hiérarchies d'appartenance, conformément à ce qu'exige l'interprétation de phénomènes comme la sensation d'émergence ou les illusions. Dans [Dessalles 1993b] nous avons recherché un mécanisme neuronal plausible pour la détection (voir chap. II.4).

Cette manière de voir les concepts va nous permettre de les "ancrer", c'est-à-dire de définir leur *sémantique*. Dans un modèle logique, par exemple en logique des propositions, la *sémantique* des symboles est fournie par une *interprétation*, c'est-à-dire une fonction qui associe une valeur de vérité (vrai/faux) à chaque symbole.

L'état d'activité d'un concept vu comme un détecteur, *dans une situation donnée*, ne doit pas être confondu avec la valeur de vérité qu'une modélisation logique lui accorde. Le lien peut être toutefois établi si l'on adopte la sémantique d'un monde ouvert<sup>(15)</sup> :

détecteur $D_C$	concept $C$
actif	vrai
inactif	indéterminé

Comme nous l'avons indiqué p. 144, le statut des concepts niés est un peu différent. (*non C*) est défini à partir de  $C$  et de la relation problématique (au sens donné à ce terme dans les chapitres précédents), par exemple de l'incohérence, à laquelle  $C$  participe dans le contexte :

détecteur $D_C$ participant à un contexte problématique	concept ( <i>non C</i> )
actif	faux
désactivé	vrai

Nous aurons l'occasion, dans le chapitre II.4, de revenir sur le statut cognitif de *non C*. Pour l'instant, nous nous contentons d'affirmer le fait suivant qui, si nous parvenons à l'étayer par des expériences psychologiques, pourra avoir un caractère spectaculaire : la négation logique d'un concept n'intervient qu'en tant qu'invalidation d'une situation problématique. Les concepts niés sont, de ce point de vue, des concepts "temporaires".

En décrivant les concepts comme des détecteurs, nous pouvons définir l'ancrage d'un concept :

**Sémantique d'un concept logique** : les concepts sont ancrés dans ce qu'ils détectent dans le contexte. Autrement dit, la "sémantique" d'un concept  $C$  *dans une situation donnée* correspond à l'ensemble  $\{P_\alpha\}$  de perceptions<sup>(16)</sup> qui l'a activé :

$$C = \mathcal{M}_a(\{P_\alpha\})$$

Comme nous l'avons indiqué, cette activation n'a aucune raison d'avoir un caractère systématique. Ceci signifie que les concepts logiques sont des détecteurs approximatifs, et qu'il n'est pas possible de définir un concept à partir de l'ensemble  $\{\{P_\alpha\}\}$  de toutes les configurations  $\{P_\alpha\}$  susceptibles de l'activer. Cet ensemble correspondrait à l'extension de  $C$ , mais hors des situations mathématiquement axiomatisées, il est inaccessible. Cette impossibilité a été maintes fois constatée, par exemple à travers l'étude de la typicalité [Barthélémy 1990], mais il serait tout de même hâtif de renoncer au caractère logique des concepts simplement parce que leur extension ne peut être définie.

Les phénomènes comme la typicalité<sup>(17)</sup> tendent à montrer que les concepts ne sont pas des détecteurs fiables. Leurs frontières de détection sont tellement floues que certains chercheurs en sont venus à donner des définitions extensionnelles probabilistes des concepts [Rosch 1978]. Cependant, il est à noter que ces phénomènes ont été tous observés dans des situations expérimentales artificielles, en l'absence de contexte. La présence d'un contexte restaure l'efficacité détectionnelle des concepts [Medin 1989, Houdé 1990].

Les concepts logiques, considérés comme des détecteurs, peuvent avoir la fonction de segmentation du monde perçu qu'on leur reconnaît généralement, sans pour autant que cette fonction les définisse. En revanche, un concept peut être défini par les relations logiques (que nous modélisons par des incompatibilités) qu'il entretient avec d'autres concepts, et hériter sa fonction de détection de ces relations inter-conceptuelles.

<sup>15</sup> Dans un monde "fermé", tout ce qui n'est pas vrai est faux, et réciproquement. Dans un monde ouvert, une proposition peut être vraie, fausse, ou non encore déterminée. Ici, un concept peut être vrai ou indéterminé.

<sup>16</sup> (voir note 3 p. 140)

<sup>17</sup> Par exemple, un canari est plus typique (et plus vite reconnu) comme un oiseau que ne l'est un pinguin. Une voiture est "plus" un véhicule que ne le sont des patins à roulettes.

Il nous reste donc un certain nombre de points essentiels à éclaircir : comment les concepts logiques sont-ils formés ? Lorsqu'une perception est à l'origine de la formation d'un concept, comment les relations inter-conceptuelles sont-elles établies avec le nouveau concept ? Ensuite, lorsqu'un concept est formé et relié aux autres concepts, comment sont définies ses capacités de détection ? Plus précisément, nous voudrions pouvoir définir, pour un mécanisme  $\mathcal{M}_C$  donné et pour tout concept  $C$  formé avec  $\mathcal{M}_C$  à partir d'autres concepts, un ensemble  $\{P_\beta\}$  de perceptions susceptible de déclencher  $C$  :

$$C = \mathcal{M}_C(\{C_\alpha\}) = \mathcal{M}_a(\{P_\beta\})$$

Ces questions ne sont pas évidentes, mais nous croyons que le fait de les poser dans ces termes devrait mener à une meilleure compréhension des processus d'apprentissage proprement humains.

## 6. La formation de concepts ancrés

### 6.1. Le problème de la bidétermination

Nous avons donné la primauté à la définition intensionnelle des concepts, pour ensuite les présenter comme des *détecteurs*. Nous y avons été contraints par l'existence d'un certain nombre de phénomènes, à commencer par la capacité des individus à repérer le caractère incohérent d'une situation. Si l'on accepte cette description, alors on se trouve confronté à une situation redoutable. Certain de ces détecteurs sont formés à partir de perceptions (voir note 3 p. 140). Comment des perceptions vont-elles pouvoir provoquer la formation d'un détecteur qui sera justement capable de les détecter ?

Ce problème a déjà été rencontré en biologie, notamment en immunologie. Un antigène (*i.e.* un corps étranger) provoque l'apparition d'un clone lymphocytaire capable de le détecter et de le détruire. Nous donnons le nom de *bidétermination* à ce problème :

**bidétermination** : une situation bidéterminée est une situation dans laquelle un objet  $A$  provoque la création matérielle d'un détecteur  $D$  tel que  $D(A) = 1$ . L'objet  $A$  *détermine* le détecteur  $D$ , qui à son tour *détermine* un ensemble détecté qui contient  $A$ .

Le problème de la bidétermination est manifestement un problème crucial lorsqu'il s'agit d'expliquer le rôle des perceptions dans la formation des concepts. Il existerait évidemment une solution simple à ce problème : il suffirait de rechercher une *ressemblance* entre le détecteur et ce qu'il détecte. Effectivement, si une perception crée un concept qui lui "ressemble", et qu'un concept détecte les configurations qui lui ressemblent, alors le problème de la bidétermination semble résolu. Voire !

En fait, la notion même de détecteur est *a priori* étrangère à toute idée de ressemblance. Le détecteur, qui peut être dans deux états (actif/inactif) n'a pas de raison de ressembler à ce qui le déclenche. Pour prendre une image, un interrupteur ne ressemble pas au doigt qui l'actionne. Un anticorps, en tant que molécule formée d'acides aminés organisés sur deux chaînes lourdes et deux chaînes légères, ne ressemble pas à l'antigène avec lequel il entre en affinité. Le rapport entre le déclencheur et le déclenché est un rapport d'affinité, d'adéquation, et non un rapport de ressemblance.

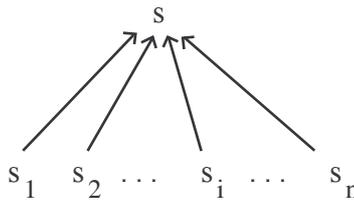
Pour nous, la représentation matérielle d'un concept ne ressemble pas à la représentation matérielle des perceptions, et ce n'est donc pas ainsi que l'on peut expliquer la bidétermination dans le cas des concepts. En effet, pour rendre compte de la systématisme de certains processus, nous avons postulé l'existence d'un appareil conceptuel matériellement unique, et les concepts logiques sont alors apparus comme des représentations partageant le même substrat matériel. Lorsqu'une perception est à l'origine de la formation d'un nouveau concept logique, la représentation matérielle ce dernier est distincte de celle de la perception, et n'a aucune raison d'entretenir une quelconque ressemblance avec celle-ci. Ces deux représentations ont, au contraire, des raisons de ne pas se ressembler, comme nous allons le constater.

Il semble ainsi que dans notre modèle, la création d'un nouveau concept nécessite un *recodage* de l'information. C'est ainsi que nous allons pouvoir expliquer comment le problème de la bidétermination peut être résolu (voir section 7.3.).

Nous proposons de distinguer deux modes différents d'extraction d'une information pertinente, dont un seul peut, selon nous, décrire la formation des concepts.

## 6.2. Formation de concepts par extraction analogique

Certaines situations d'apprentissage sont décrites comme résultant d'un mécanisme d'extraction :  $\{s_1, s_2, \dots, s_n\} \rightarrow s$ .



**figure 4** : illustration schématique d'une extraction de type analogique. La représentation  $s$  est extraite à partir d'un ensemble de représentations  $s_i$  auxquelles elle ressemble.

La représentation apprise  $s$  est extraite d'un ensemble  $\{s_i\}$  de représentations de situations vécues :  $s$  est ce qu'il y a de commun aux  $s_i$ . Ce type d'apprentissage impose donc un lien particulier, un lien de ressemblance, entre ce qui est apprenable et un ensemble de situations effectivement rencontrées.

Un tel mécanisme est souvent invoqué pour expliquer l'apprentissage de formes (au sens large) stables en tant qu'invariants perceptifs. Si nous voyons en  $s$  la représentation matérielle de la forme apprise (par ex. le pattern d'activité des neurones d'un réseau de Hopfield utilisé comme classifieur et entraîné sur des exemples tous différents), et si les  $s_i$  désignent les représentations matérielles directement provoquées par les perceptions, alors  $s$  *ressemble* (au sens matériel, par ex. pour la distance de Hamming dans l'exemple de Hopfield) aux  $s_i$ .

Nous qualifierons ce mécanisme d'*extraction analogique*<sup>(18)</sup>, puisque  $s$  ressemble à chacune des représentations perceptives  $s_i$ .

**Mécanisme d'extraction analogique** : Une application  $g : E^K \rightarrow E$  est un mécanisme d'extraction analogique si

$$\forall \{s_1, \dots, s_K\} \in E^K \quad \frac{1}{K} \sum_{i=1}^K d_o(s, s_i) \ll N \quad \text{avec } s = g(s_1, \dots, s_K)$$

$d_o(\cdot)$  est la distance permettant d'estimer la ressemblance, par exemple la distance de Hamming dans l'hypercube (le critère peut être remplacé par un critère équivalent, faisant par exemple intervenir le carré des distances).

Précisons cette définition dans le cas de la formation des concepts. Il s'agit ici d'expliquer comment  $s$ , extraite analogiquement d'un ensemble de situations  $\{s_i\}$ , peut servir de détecteur. La solution généralement proposée consiste à considérer que  $s$  est utilisée dans un détecteur de proximité :

<sup>18</sup> Ne pas confondre avec l'"apprentissage par analogie", qui intervient dans un tout autre contexte. Noter que le terme "ancrage" a deux significations : il peut désigner ce sur quoi le concept est ancré, mais aussi l'action d'ancrer. C'est cette action qui est produite par un mécanisme analogique, ici.

**Détecteur de proximité** : Un détecteur  $D_s$  détecte la proximité par rapport à une représentation  $s$  s'il est de la forme :

$$D_s(s') = \begin{cases} 1 & \text{si } d(s, s') < \xi \\ 0 & \text{sinon} \end{cases}$$

où  $d()$  est la distance utilisée par le détecteur pour estimer la ressemblance et  $\xi$  un seuil caractéristique du détecteur.  $d()$  et  $\xi$  peuvent dépendre des  $\{s_j\}$ .

Nous pouvons ainsi donner la définition d'un mécanisme analogique de création de concept :

**Formation analogique de concept** : Un concept, considéré comme un détecteur, est formé analogiquement s'il s'écrit comme un détecteur de proximité  $D_s$  où  $s = g(s_1, \dots, s_K)$  est le produit d'un mécanisme d'extraction analogique  $g()$ .

Ce mécanisme permet de résoudre le problème de la bidétermination (cf. p. 151). Si  $d_o()$  et  $d()$  sont des distances topologiquement équivalentes,  $D_s$  va être tel que  $D_s(s_i) = 1$  pour tous les  $s_i$  ayant participé à la formation de  $D_s$  (ou au moins la plupart d'entre eux). Nous allons voir, cependant, que cette formation analogique des concepts soulève des problèmes d'ordre théorique.

Auparavant, nous allons considérer les mécanismes de formation analogique des concepts qui ont de plus la qualité d'être indifférents. En effet, les mécanismes d'extraction analogique généralement proposés sont de type indifférent (*i.e.* ils tolèrent tout changement de repère en entrée, cf. chap. II.2). Donnons la définition pour le cas qui nous occupe ici.

**Mécanisme d'extraction indifférent** : Une application  $g : E^K \rightarrow E$  est un mécanisme d'extraction indifférent si :

$$\forall \{s_1, \dots, s_K\} \in E^K \quad \forall \rho \text{ isométrie dans } E \quad g(\rho(s_1), \dots, \rho(s_K)) = \rho(g(s_1, \dots, s_K))$$

Autrement dit le mécanisme est insensible aux changements de repère. Par exemple, les mécanismes d'extraction statistique sont indifférents. Rappelons maintenant la définition d'un mécanisme d'apprentissage indifférent.

**Mécanisme indifférent de formation de concept** : le mécanisme de formation analogique de concept  $\{s_1, \dots, s_K\} D_{g(s_1 \dots s_K)}$  est un mécanisme d'apprentissage indifférent si :

$$\text{pour toute isométrie } \rho, \quad D_{g(\rho(s_1), \dots, \rho(s_K))} \circ \rho = D_{g(s_1, \dots, s_K)}$$

Cette définition est conforme à celle que nous avons donné au chapitre II.2. Elle signifie que le mécanisme de formation du concept est indifférent à un changement de repère.

Dans la mesure où  $g()$  peut être considérée comme surjective (ce qui est garanti si l'on accepte que  $g(s, \dots, s) = s$ , propriété vérifiée par tous les mécanismes analogiques courants), l'indifférence du mécanisme de formation du concept est équivalente, pour un détecteur de proximité formé par un mécanisme d'extraction lui-même indifférent, à la propriété suivante :

$$D_{\rho(s)} \circ \rho = D_s$$

Pour la discussion qui suit, nous allons envisager l'hypothèse selon laquelle les concepts seraient formés par un mécanisme d'extraction analogique indifférent.

### 6.3. Critique de l'extraction analogique

L'extraction analogique des concepts logiques poserait un certain nombre de problèmes qui semblent réhibitoires.

### 6.3.1. Indifférence des mécanismes analogiques

Le premier de ces problèmes vient du caractère indifférent de l'extraction analogique (isotropie, relativité concernant divers paramètres perceptifs) qui exigerait un grand nombre d'exemples pour la formation du moindre concept à partir de perceptions.

Nous avons vu en effet que les propriétés d'indifférence du mécanisme d'apprentissage entraînaient soit l'invariance de la forme apprise pour de nombreuses transformations, ce qui n'est certainement pas le cas ici pour les concepts<sup>(19)</sup>, soit une taille importante pour le jeu d'apprentissage (cf. chapitre II.2). Nous ne serons donc pas surpris de constater que les mécanismes d'extraction analogique généralement proposés rendent compte de l'extraction de la représentation  $s$  à partir d'un ensemble  $\{s_j\}$  de cardinal nécessairement significatif. Ce sont le plus souvent des mécanismes de type statistique. Or ce n'est pas ce que l'on observe chez les individus, qui sont capables de former un concept à partir d'un seul exemple (voir le cas de l'EBL en section 7.3.).

### 6.3.2. Manque de plausibilité neuronale d'une copie analogique

Une deuxième difficulté vient du manque de plausibilité neuronale d'un tel mécanisme d'extraction analogique. En raison de la systématisme de certains processus opérant sur les concepts, nous avons dû conclure que les structures impliquées dans la perception et celles impliquées dans le traitement des concepts logiques sont nécessairement distinctes. Or si elles sont distinctes, une extraction analogique nécessite une copie analogique des premières vers les secondes.

On s'aperçoit alors qu'une copie analogique entre deux structures au mode de fonctionnement très différent n'a rien d'évident. Alors que les copies analogiques sont facilement exécutées par des dispositifs techniques (transducteurs), les systèmes neuronaux que l'on a imaginés jusqu'à maintenant encodent les caractéristiques retenues *sur place*, dans les poids synaptiques et dans l'activité des neurones. Comment imaginer qu'au sein de l'appareil conceptuel, une représentation soit créée qui reproduise fidèlement (comme une copie analogique) tout ce qui est encodé dans les appareils perceptifs ?

### 6.3.3. Encodage implicite des paramètres analogiques

L'extraction analogique pose d'autres problèmes. Les mécanismes analogiques encodent les différents paramètres d'un objet perçu (taille, durée, luminosité, puissance sonore, etc.) de manière implicite dans la représentation. Or comme nous allons le rappeler, seules les caractéristiques *explicites* peuvent jouer un rôle logique et constituer des concepts.

Par exemple, imaginons que les  $s_j$  soient provoquées par la projection d'un même objet successivement sur des endroits différents de la rétine. La représentation  $s$  retiendra la forme de l'objet, mais pas sa position. Si maintenant on recommence l'expérience en présentant un objet homothétique aux mêmes positions et que cela provoque le changement d'une caractéristique de toutes les  $s_j$ , par ex. le nombre de neurones impliqués, alors cette caractéristique aura changé dans  $s$ . En d'autres termes, certaines caractéristiques matérielles de la représentation  $s$  sont couplées analogiquement à des caractéristiques des objets du jeu d'apprentissage.

Pour prendre un autre exemple, la durée effective<sup>(20)</sup> d'un morceau de musique est encodée implicitement par un système d'enregistrement audio, même si celui-ci est digital.

Nous nous intéressons à la formation des concepts, que nous présentons comme des détecteurs. Dans la théorie de la formation analogique des concepts, c'est le détecteur qui est supposé être issu d'une extraction analogique. Or cette théorie suppose que toutes les caractéristiques du concept sont incluses dans la représentation, c'est-à-dire ici dans le détecteur lui-même. Dans quelle mesure une caractéristique reconnue par un détecteur peut-elle être encodée de manière implicite dans ce détecteur ?

Intuitivement, un paramètre est codé implicitement dans une représentation s'il peut être extrait de cette représentation. Plus précisément :

<sup>19</sup> Les concepts ne consistent certainement pas des "bonnes formes", au sens de la psychologie gestaltiste (cf. discussion chap. II.1).

<sup>20</sup> La durée que le lecteur affiche ne résulte pas d'une mesure, et pourrait ne pas correspondre à la durée effective.

**Encodage implicite** : Le paramètre  $\alpha$  est encodé implicitement dans une représentation  $s$  si il existe une fonction  $h$  telle que  $\alpha = h(s)$ . Nous dirons aussi que  $\alpha$  est encodé implicitement dans le détecteur de proximité  $D_s$  s'il est encodé implicitement dans  $s$ .

**Paramètre strict** : Un paramètre  $\alpha = h(s)$ , encodé implicitement dans une représentation  $s$ , est un paramètre *strict* si il existe une projection  $\tau$  de  $E$  dans  $E$ , différente de l'identité, telle que  $h \circ \tau = h$ . Cela signifie que  $\alpha$  ne dépend que d'une partie des composantes de  $s$ .

Nous allons démontrer le résultat fondamental suivant :

*Un détecteur de proximité produit par un mécanisme analogique et indifférent ne peut pas détecter les paramètres stricts dont il dépend implicitement.*

Intuitivement, cela vient du fait que le détecteur de proximité  $D_s$  est "obligé" de prendre en compte *toute* l'information contenue dans  $s$ , c'est-à-dire ici toutes les composantes de  $s$ .

Lemme : Soit  $\alpha = h(s)$  un paramètre strict encodé implicitement dans  $s$ . Alors il existe  $s^o$  et une isométrie  $\rho_o$  de  $E$  dans  $E$  telle que

$$\rho_o(s^o) \neq s^o \quad \text{et} \quad \forall s \quad h(\rho_o(s)) = h(s).$$

En effet, il existe par hypothèse,  $\alpha$  étant un paramètre strict, une projection  $\tau$  telle que :

$$h \circ \tau = h \quad \text{et} \quad \exists x_o \quad \tau(x_o) \neq x_o$$

On peut prendre alors  $s^o = \tau(x_o)$  et  $\rho_o(x) = x \ (x_o \ \tau(x_o))$ , où désigne le "ou" exclusif. On a bien :

$$\rho_o(s^o) = \tau(x_o) \ \ x_o \ \ \tau(x_o) = x_o \ \ s^o$$

car  $\tau(x_o) \oplus \tau(x_o) = 0$ , et

$$h(\rho_o(s)) = h(\tau(\rho_o(s))) = h(\tau(s \ \ x_o \ \ \tau(x_o))) = h(\tau(s) \ \ \tau(x_o) \ \ \tau(x_o))$$

car  $\tau$  est linéaire et  $\tau^2 = \tau$ . Il reste donc :

$$h(\rho_o(s)) = h(\tau(s)) = h(s)$$

démonstration :

Soit  $D_s$  un détecteur de proximité. Soit  $\alpha = h(s)$  un paramètre strict encodé implicitement dans  $s$ . D'après le lemme, il existe  $s^o$  et une isométrie  $\rho_o$  tels que

$$\rho_o(s^o) \neq s^o \quad \text{et} \quad \forall x \quad h(\rho_o(x)) = h(x)$$

Nous allons montrer par l'absurde que  $D_s$  ne peut pas détecter  $\alpha$ . Supposons donc le contraire, c'est-à-dire que  $D_s$  peut s'écrire :

$$\forall s' \quad D_s(x) = \Delta_s(h(x))$$

Ceci signifie bien que  $D_s(x)$  est une fonction booléenne  $\Delta_s$  du paramètre  $\alpha$  codé implicitement dans  $x$ .

Comme  $D_{s^o}(x)$  est issu d'un mécanisme analogique indifférent, nous avons la propriété suivante :

$$\forall \rho \text{ isométrie} \quad \forall x \quad D_{s^o}(x) = D_{\rho(s^o)}(\rho(x))$$

En particulier pour  $\rho_o$  :

$$\begin{aligned} \forall x \quad D_{s^o}(x) &= D_{\rho_o(s^o)}(\rho_o(x)) = \Delta_{\rho_o(s^o)}(h(\rho_o(x))) \\ &= \Delta_{\rho_o(s^o)}(h(x)) \end{aligned}$$

$$= D_{\rho_0(s^0)}(x)$$

Or cette conclusion  $\forall x \ D_{s^0}(x) = D_{\rho_0(s^0)}(x)$  est absurde dans la mesure où  $D_s$  est un détecteur de proximité et que  $s^0 \neq \rho_0(s^0)$ . CQFD.

En particulier, un détecteur de proximité  $D_s$  ne peut pas détecter les composantes de sa référence  $s$ .

Par exemple, dans un extrait de conversation que nous avons recueilli, un interlocuteur attire l'attention sur le dard exceptionnellement long d'un insecte qu'il découvre sur une pierre. Son étonnement est relatif à un concept, *dard\_long*. Est-il possible que la longueur du dard soit un paramètre implicite du concept *insecte*, ce concept ayant été extrait de manière analogique des perceptions antérieures ? Non, car cette longueur joue un rôle logique, elle constitue donc un concept logique à part entière, elle a une représentation propre au même titre que les autres concepts logiques, et elle peut être manipulée par les mêmes mécanismes qu'eux. C'est dans ce sens que nous qualifions cette représentation d'*explicite*. Si la longueur du dard était au départ un paramètre implicite de la représentation perceptive de l'insecte, alors elle a dû, pour devenir explicite, subir un *décodage*, autrement dit elle doit résulter d'une détection. Or comme nous venons de le voir, le concept *insecte* ne peut pas réaliser lui-même cette détection.

Ceci veut dire que toutes les caractéristiques de l'objet susceptibles de jouer un rôle logique (ici les ailes, les antennes, la taille, etc.) doivent être détectés pour elles-mêmes, par des détecteurs distincts du détecteur *insecte*. Elles ne peuvent pas jouer ce rôle logique en restant implicites dans la représentation *insecte*.

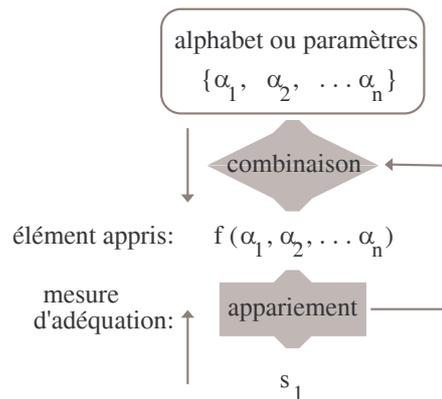
#### 6.3.4. Gradualité des systèmes analogiques

Notons enfin que les processus qui opèrent sur les concepts logiques fonctionnent de manière *non graduelle*, contrairement à ce que l'on attendrait d'une extraction analogique. Les fameuses expériences de Shepard [Cooper & Shepard 1985] ont montré que certains traitements sur des perceptions dépendaient de manière graduelle de certains paramètres des formes perçues. Par exemple, le temps mis par un sujet (humain ou singe) pour juger si deux formes gauches sont superposables est proportionnel à l'angle entre les deux formes. On n'observe rien de tel pour les traitements conceptuels. Un mécanisme comme la recevabilité d'une tentative d'explication (voir p. 144) ne tient compte que du rapport logique formel entre les termes explicites du contexte, et non des termes eux-mêmes. L'acceptation d'une explication tient bien compte des valeurs de vérité localement accordées aux termes du contexte, mais elle ne varie pas graduellement en fonction d'une quelconque caractéristique des objets détectés par ces termes<sup>(21)</sup>.

#### 6.4. Formation de concept par appariement

Pour toutes ces raisons, il nous faut considérer un autre type de mécanisme pour former des concepts logiques ancrés. Or le mode d'extraction analogique de l'information n'est pas le seul qui a été proposé. Il y en a au moins un autre, l'appariement, que nous allons décrire ici car il supprime la contrainte de ressemblance entre la représentation matérielle du concept logique et les perceptions que celui-ci détecte :

<sup>21</sup> Nous devrions argumenter plus longuement pour montrer que des processus conversationnels comme la banalisation ou la réaction antagoniste (cf. chap. I.1) ne sont pas en dépendance graduelle par rapport à un paramètre implicite d'un terme de la situation. Si l'auteur d'une telle réaction est amené à justifier son intervention, il *explicite* des valeurs caractéristiques des paramètres ou de leur variation, comme dans les modélisations qualitatives du raisonnement [Stepankova 1992]. Nous ne pouvons discuter ce point en détail ici.



**figure 5** : illustration schématique d'un apprentissage par appariement. Les objets  $\alpha_i$  sont donnés au départ. Le système crée des combinaisons  $f(\alpha_1, \alpha_2, \dots, \alpha_n)$  de ces objets pour qu'elles soient en *adéquation* avec la situation  $s_1$  à REprésenter. Le résultat de l'apprentissage est une combinaison adéquate.

Le mécanisme d'appariement :  $s_1 \rightarrow f(\alpha_1, \alpha_2, \dots, \alpha_n)$  obéit à un principe fondamentalement différent de celui de l'extraction analogique. L'élément appris provient d'un "alphabet" limité d'éléments préexistants  $\alpha_i$ , généralement à la suite d'une combinaison  $f()$ . Cet élément appris  $f(\alpha_1, \alpha_2, \dots, \alpha_n)$  ne ressemble pas à la situation  $s_1$  qui provoque l'apprentissage, dans la mesure où il est d'une autre nature et n'a aucune raison d'avoir des propriétés en commun avec  $s_1$ . En revanche il doit être en *adéquation* avec la situation déclenchante  $s_1$ .

On pourrait imaginer qualifier un tel mécanisme de *numérique*, pour reproduire la distinction analogique / numérique telle qu'on la fait par exemple en théorie des communications. En effet, la présence d'un alphabet d'éléments préexistants  $\{\alpha_i\}$  de cardinalité limitée et la mesure d'adéquation sont essentielles dans le mécanisme d'appariement comme pour les communications de type numérique<sup>(22)</sup>. Cependant, dans le cas de l'apprentissage, la forme appariée à la situation déclenchante est le résultat d'une combinaison  $f()$ , rendue nécessaire par la limitation de l'alphabet.

*Un système d'appariement est par essence non indifférent (i.e. il est sensible aux changements de repère dans le codage de ses entrées).* Cette non-indifférence est liée à l'existence de l'alphabet  $\{\alpha_i\}$ . En effet, un tel modèle suppose l'existence d'éléments  $\alpha_i$  indépendants de toute expérience. Ces  $\alpha_i$  jouent le rôle de primitives conceptuelles. Nous sommes contraints de postuler leur existence, mais il n'est pas évident qu'il soit facile de les déterminer. Les concepts de symétrie [Leyton 1993], de permanence dans le temps ou de proximité sont peut-être des exemples de telles primitives (cf. aussi la "whole object assumption" [Markman 1990], ou les primitives conceptuelles de la sémantique spatiale inspirée par Lakoff : co-localisation, inclusion, orientation [Veale & Keane 1992]).

La non-indifférence inhérente au mécanisme d'appariement le libère de la contrainte qui pèse sur les systèmes analogiques. Ceux-ci, lorsqu'ils ne conduisent pas obligatoirement à des formes fortement symétriques, nécessitent de nombreux exemples pour fonctionner (cf. chapitre II.2).

Avec un mécanisme d'appariement combinatoire, un nombre très limité de situations (voire une situation unique) peut provoquer l'apprentissage d'une combinaison adéquate. Contrairement aux exigences de l'extraction analogique, nous n'attendons pas d'avoir vu, par exemple, 136 chausse-pieds pour en former le concept. Un seul de ces objets, dont nous comprenons la fonction, peut nous suffire pour que nous lui fassions correspondre une représentation conceptuelle que nous relions très vite à d'autres concepts. Avec un seul exemple, nous comprenons par exemple que la couleur de l'objet n'est pas pertinente pour sa fonction. Le détecteur conceptuel est alors en place. Si le deuxième chausse-pied que nous voyons est rouge alors que

<sup>22</sup> En communications numériques, les éléments  $\alpha_i$  sont par exemple des impulsions électriques idéales, et la mesure d'adéquation avec le signal reçu (qui correspond à une impulsion distordue et bruitée) se fait généralement par une opération de corrélation (par "filtrage adapté") suivie d'un échantillonnage.

le premier était brun, nous ne serons certainement pas surpris, alors que nous serons étonnés s'il est fait d'une matière molle.

Ce mécanisme d'appariement combinatoire peut être invoqué pour expliquer plusieurs aspects de la cognition (par ex. l'apprentissage du langage) ou de l'organisation du vivant (par ex. apprentissage d'un antigène par le système immunitaire). Dans chaque cas il faut déterminer quatre choses :

- ◆ l'alphabet,
- ◆ les modes de combinaison,
- ◆ le mécanisme qui pousse à essayer certaines combinaisons,
- ◆ la mesure d'adéquation.

Dans le cas du système immunitaire, le problème de la bidétermination (voir p. 151) est résolu : il n'y a pas de copie analogique de l'antigène vers l'anticorps. Le choix des combinaisons "essayées" est même totalement indépendant de l'antigène. C'est un mécanisme de sélection qui réalise la mesure d'adéquation. L'anticorps n'est pas déterminé par l'antigène, il est simplement sélectionné.

#### *immunologie*

---

alphabet :	sites actifs
modes de combinaison :	concaténation
choix des combinaisons :	hasard
mesure d'adéquation :	sélection selon l'affinité

---

Pour le système immunitaire, l'alphabet est constitué par les sites actifs qui peuvent apparaître dans les chaînes légères des immunoglobulines. Ils sont codés par les gènes recombinés des clones lymphocytaires. Chaque clone réalise une combinaison en réunissant des sites actifs le long des chaînes légères. Le hasard seul semble présider au choix d'une nouvelle combinaison. Il est corrigé par la détection d'activité auto-immune. L'adéquation, enfin, est évaluée par le lymphocyte qui détecte que les immunoglobulines qu'il porte sont entrées en réaction d'affinité avec une molécule externe. L'apprentissage est obtenu, entre autres, par la multiplication du lymphocyte adéquat.

Dans le cas de l'apprentissage conceptuel, la réponse à ces quatre questions est loin d'être simple compte tenu des connaissances actuelles. C'est pourtant là que réside le problème de la bidétermination dans le cas des concepts. Certaines théories cognitives [Changeux & Dehaene 1989] suggèrent un mécanisme sélectif semblable à celui que nous venons d'évoquer pour le système immunitaire. Les concepts (idées) préexisteraient tous, et seraient sélectionnés selon leur efficacité dans le traitement cognitif. Cette théorie a le mérite de résoudre le problème de la bidétermination.

Notre point de vue, en ce qui concerne les concepts logiques, n'est toutefois pas celui-là. Nous pensons que, contrairement au scénario sélectif, la situation (contexte conceptuel et perceptif) dirige la synthèse de nouveaux concepts adéquats (appariement guidé). Nous allons donc essayer de suggérer des mécanismes plausibles de combinaison et de mesure d'adéquation qui puissent résoudre le problème de la bidétermination dans le cas des concepts.

## 7. Exemples d'appariement guidé pour la formation des concepts

Dans cette section, nous proposons divers mécanismes par lesquels de nouveaux concepts peuvent être formés. Dans le cadre dans lequel nous nous plaçons, cette création de concepts ne va pas de soi, car elle pose le problème de la bidétermination (voir p. 151) : la présence d'un ou plusieurs objets doit provoquer la création, par appariement, d'un nouveau concept qui est justement capable de reconnaître ces objets. Les mécanismes que nous évoquons dans cette section sont des techniques développées dans le domaine de l'apprentissage symbolique automatique (ASA). Nous les citons ici parce qu'elles ont cette impressionnante capacité de résoudre le problème de la bidétermination.

Toutefois, en examinant la vraisemblance cognitive de ces techniques, nous laisserons entendre qu'aucune d'elles n'est suffisante pour rendre compte de la formation de nouveaux concepts par les humains : toutes ces techniques négligent en effet de préciser les *conditions* dans lesquelles l'apprentissage doit avoir lieu. Ce dernier point, qui fera l'objet de la section 8., nous permettra de montrer l'importance d'un contexte de type conversationnel pour la création d'un concept.

Nous allons examiner successivement trois techniques utilisées en ASA : l'apprentissage par similarités (SBL), l'appariement structurel, et l'apprentissage basé sur une explication (EBL). Notre objectif sera, dans chaque cas, de montrer comment le problème de la bidétermination est résolu lors de la création de nouveaux concepts.

### 7.1. Apprentissage de concepts basé sur les similarités

Dans l'approche basée sur les similarités (SBL : Similarity Based Learning), on recherche une *caractérisation*, généralement exprimée en logique, d'une classification d'exemples. Chaque exemple est décrit par les valeurs d'un certain nombre d'attributs. Dans notre contexte, les attributs apparaissent comme des concepts élémentaires (primitives conceptuelles) activés ou non par les différents exemples. L'objectif de la caractérisation est de fournir une définition intensionnelle concise de chaque classe à partir des exemples fournis. Dans le cas le plus simple, il n'y a que deux classes, et on parle alors d'exemples et de contre-exemples.

On peut ainsi voir dans le SBL un mécanisme d'appariement symbolique résolvant le problème de la bidétermination (v.p. 151) : les algorithmes de SBL permettent de forger de nouveaux concepts (un par classe caractérisée) à partir d'exemples qui feront partie de leur "extension". Le tableau suivant explicite cette relation entre SBL et appariement :

<i>SBL</i>	
alphabet :	valeurs d'attributs
modes de combinaison :	combinaison logique (et / non)
choix des combinaisons :	dépend de l'algorithme
mesure d'adéquation :	simplicité et précision

Dans leur version de base, les algorithmes SBL opèrent sur des objets définis par les valeurs de leurs attributs. Ces attributs, et les valeurs possibles, sont donnés au départ et ne peuvent plus être modifiés. Dans la version non supervisée, l'algorithme doit classer les objets suivant leur ressemblance (estimée d'après les valeurs d'attributs). Dans la version supervisée, la classification est fournie. Mais dans tous les cas, supervisés ou non, l'important est de parvenir à une caractérisation (sous forme logique ou équivalente) de la classification <sup>(23)</sup>.

Par exemple CLUSTER/2 [Michalski 1983] commence par classer les objets selon leur ressemblance avec  $k$  objets pris au hasard (les semences). Il retient pour chaque classe obtenue une conjonction d'attributs communs aux éléments de la classe et permettant de distinguer cette classe des autres. Il évalue la qualité de la caractérisation obtenue en fonction de sa simplicité et de sa précision (couverture et séparation des classes), puis tente de modifier les  $k$  classes par un choix de nouvelles semences. Cet algorithme mène donc simultanément les tâches de classification et de caractérisation [Fisher & Pazzani 1991].

Les algorithmes SBL sont utiles dans la mesure où ils fournissent une description intensionnelle des classes qui a la qualité d'être concise et parfois hiérarchisée (par ex. sous la forme d'un arbre de décision comme dans ID3 [Quinlan 1983]). Mais cette description a de plus l'ambition d'être *prédictive*, dans la mesure où elle donne avec succès une classe adéquate à des objets qui n'ont pas encore été rencontrés.

<sup>23</sup> Bien qu'il y ait utilisation de la ressemblance, les algorithmes SBL ne sont pas analogiques. Les concepts caractérisés sont des formules logiques, ce ne sont pas des éléments typiques. Ces algorithmes sont toutefois indifférents (par rapport aux attributs primitifs), ce qui, d'ailleurs, explique le fait qu'ils nécessitent un nombre significatif d'exemples.

Ces algorithmes, outre leur intérêt pratique et technique, présentent pour nous, comme nous l'avons dit, l'avantage de montrer la faisabilité d'un appariement conceptuel qui résout le problème de la bidétermination (voir p. 151) : des exemples (en nombre certes significatif) vont diriger la création de nouveaux concepts (caractérisations intentionnelles des classes) qui vont détecter ces mêmes exemples.

Toutefois, la pertinence psychologique de tels algorithmes est discutable. Il n'est pas évident que les humains classifient spontanément les objets d'après une ressemblance perceptive (les "attributs" déclenchés par la perception) et cherchent à caractériser les classes obtenues. Ils en sont certes plus ou moins capables dans des situations expérimentales [Dubois 1991]. Mais ils semblent préférer des modes de classification thématiques et fonctionnels [Houdé 1990], ils prennent en compte des descriptions structurelles des objets et utilisent leur connaissance du domaine.

La formation de concepts à partir de la simple constatation des ressemblances existe peut-être, mais il nous semble qu'elle ne peut jouer qu'un rôle marginal. Elle exige en effet que les objets soient comparés selon une liste immuable d'attributs. Les situations concrètes spontanées qui donnent lieu aux échanges conversationnels n'offrent pas de telles collections d'objets comparables selon une liste de critères, et pourtant elles sont le théâtre de nombreuses créations conceptuelles.

L'utilisation des *connaissances du domaine* permet de forger des exemples à partir d'un nombre d'exemples beaucoup plus restreint que ce qu'exige l'apprentissage par similarité. Nous allons mentionner deux exemples : l'appariement structurel et l'apprentissage basé sur l'explication.

## 7.2. *Apprentissage de concepts par appariement structurel*

Les concepts formés par l'apprentissage par similarité sont généralement de simples conjonctions d'attributs. Ces techniques sont incapables de former des concepts structurés, comme le montrent Thompson et Langley [1991]. Reprenons l'exemple donné par ces auteurs :



Ces objets sont structurés, c.à.d. qu'il ont des composants qui sont eux-mêmes des objets. L'apprentissage des concepts correspondant à de tels objets suppose la prise en compte des relations entre les composants d'un même objet. Par exemple la description d'un concept peut être : "il y a trois objets,  $X$ ,  $Y$  et  $Z$ ;  $X$  est posé sur  $Y$ ;  $X$  et  $Y$  sont à gauche de  $Z$ ". Il est difficile d'imaginer qu'une technique de SBL puisse y parvenir : même sur cet exemple simple, il faudrait que *toutes* ces relations (*à-gauche-de*, *à-droite-de*, *au-dessus-de*, *etc.*) aient été prévues et précisées dans la description de chaque objet exemple. En revanche, avec une représentation structurelle des objets, il est possible d'apprendre des concepts comportant des relations non explicites dans les attributs (e.g. un objet LeftStack avec trois composants empilés, si seules les relations à deux objets ont été prévues comme attributs dans la description).

Un programme comme LABYRINTH [Thompson & Langley 1991] repose sur des représentations structurelles. Il est capable de fabriquer un concept qui décrit chaque nouvel objet structuré présenté, résolvant en cela le problème de la bidétermination (voir p. 151).

appariement structurel (LABYRINTH)

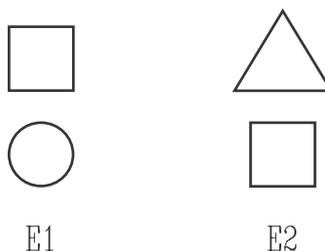

---

alphabet :	composants et relations élémentaires
modes de combinaison :	ensembles de composants et de relations qui les lient.
choix des combinaisons :	description de l'exemple; essai d'appariement avec les concepts existants; seuil d'utilité; éclatement, promotion, fusion des noeuds de l'arbre conceptuel.
mesure d'adéquation :	succès du processus d'appariement ; utilité (prob. <i>a posteriori</i> des éléments structurels).

---

LABYRINTH utilise des représentations structurées et des hiérarchies conceptuelles pour former de nouveaux concepts à partir de quelques exemples. Il tente d'apparier un objet avec les concepts existants de la hiérarchie. L'objet est décrit par ses composants particuliers (e.g. *cercle-1*, *carré-3*) et quelques relations (e.g. *cercle-1* à droite-de *carré-3*). Les concepts sont décrits par des composants généraux (e.g. *bloc* est plus général que *cercle* qui est plus général que *cercle-1*) et certaines de leurs relations (dans LABYRINTH, ces relations sont probabilistes; e.g., pour deux composants *Comp1* et *Comp2* identifiés par le programme, "*Comp1* à-gauche-de *Comp2* avec prob. 0,67"). Pour chaque concept candidat, le mécanisme d'appariement essaie les différentes combinaisons *composant-composant*, *relation-relation* possibles, puis calcule la qualité des appariements obtenus. LABYRINTH favorise les appariements qui rendent les concepts plus "utiles", c.à.d. qu'il retient le concept qui donne la plus forte probabilité aux éléments structurels des objets qu'il recouvre (y compris le nouvel objet). Au passage, le concept est modifié (mise à jour des probabilités des éléments structurels, généralisation des composants) et de nouveaux concepts sont créés au sein d'une hiérarchie selon la procédure définie dans le système COBWEB de D. Fisher [Fisher & Pazzani 1991]: un noeud de l'arbre des concepts peut être éclaté ou promu en fonction de sa qualité recalculée au vu du nouvel exemple. Par exemple la branche *root-LeftStack1* pourra devenir *root-LeftStack-[Leftstack1; LeftStack2]*, puis *root-Stack-[LeftStack-[Leftstack1; LeftStack2]; RightStack1]* (les noms *root*, *Stack*, *LeftStack* n'existent pas dans le programme, mais ils sont utilisés ici pour désigner les concepts qu'il crée).

Nous mentionnerons aussi l'appariement structurel réalisé par le système AGAPE de Y. Kodratoff et J.G. Ganascia [Ganascia 1987]. Il s'agit là encore, grâce à l'utilisation d'une connaissance du domaine constituée d'une hiérarchie de concepts primitifs, d'apparier des objets structurés dans le but de donner une description qui les englobe. Considérons l'exemple suivant [Ganascia 1987] :



Ces deux exemples peuvent être décrits par l'une ou l'autre des descriptions suivantes : "il y a deux objets l'un sur l'autre, l'objet du dessus étant un polygone", ou "il y a deux objets l'un sur l'autre, dont un carré". L'appariement structurel effectué par AGAPE consiste à mettre en correspondance les deux exemples grâce à des réécritures successives de leur description.

Nous avons là encore une solution au problème de la bidétermination, avec l'utilisation d'objets pour diriger la formation d'un concept qui va recouvrir ces objets :

appariement structurel (AGAPE)


---

alphabet :	concepts et relations élémentaires
modes de combinaison :	conjonction
choix des combinaisons :	descriptions d'objets; généralisation des descriptions après appariement des exemples entre eux
mesure d'adéquation :	succès du processus d'appariement ; exclusion des contre-exemples après appariement

---

AGAPE part des descriptions initiales de E1 et E2 :

E1 = (CARRE A) & (CERCLE B) & (AU-DESSUS A B)

E2 = (TRIANGLE C) & (CARRE D) & (AU-DESSUS C D)

AGAPE est obligé de remonter la hiérarchie conceptuelle pour mettre E1 et E2 en correspondance :

E'1 = (POLYGONE-CONVEXE Y1 X1) & (FORME X5 X4 Y2 X2) & (AU-DESSUS X1 X2)

avec les liens : X1=A; X2=B; X1≠X2; X3=CARRE; X3=Y1; Y1=CARRE; Y2=CERCLE; X3≠Y2; Y1≠Y2; X4=ELLIPSOIDE; X5=PATATOIDE

E'2 = (POLYGONE-CONVEXE Y1 X1) & (FORME X5 X4 Y2 X2) & (AU-DESSUS X1 X2)

avec les liens : X1=C; X2=D; X1≠X2; X3=CARRE; Y1=TRIANGLE; X3≠Y1; X3=Y2; Y2=CARRE; Y1≠Y2; X4=POLYGONE-CONVEXE; X5=POLYGONE

L'algorithme a conservé jusque-là toute l'information contenue dans les exemples. L'obtention d'une description commune aux exemples passe maintenant par une généralisation, obtenue par l'élimination de certains détails. AGAPE va ici conserver tout ce qui est commun aux deux descriptions :

Eg = (POLYGONE-CONVEXE Y1 X1) & (FORME X5 X4 Y2 X2) & (AU-DESSUS X1 X2)

avec les liens : X1≠X2; X3=CARRE; Y1≠Y2;

pour donner ce que l'on peut traduire par "il y a deux objets X1 et X2 l'un sur l'autre; X1 est un polygone convexe. Ces objets ont deux formes Y1 et Y2 différentes et il y a un carré X3".

Une telle généralisation est généralement trop complexe. Elle contient en fait plusieurs généralisations partielles ("il y a deux objets", ou "il y a un carré", "il y a un polygone convexe sur un autre objet", etc.). AGAPE utilise des contre-exemples pour affiner le généralisé  $E_g$ . Un bon contre-exemple doit être suffisamment proche des exemples pour être appariaable, et contenir des éléments (qu'AGAPE stockera au niveau des liens) qui le distingue des exemples (*near miss*). L'utilisation des contre-exemples permet de désigner les liens qui doivent être abandonnés.

Ces mécanismes d'appariement structurel donnent des solutions intéressantes au problème de la bidétermination pour l'apprentissage de concepts structurés. Ils présentent aussi l'intérêt de fonctionner avec très peu d'exemples. Ils comportent toutefois des aspects peu plausibles d'un point de vue cognitif, comme la comparaison systématique aux concepts connus dans LABYRINTH, ou la mémorisation de tous les liens dans le cas d'AGAPE.

### 7.3. Apprentissage de concepts basé sur l'explication

L'apprentissage à base d'explication (Explanation Based Learning, EBL) est parfois présenté comme une simple "opérationnalisation" des connaissances, c.à.d. comme une transformation de connaissances générales en connaissances spécialisées, plus facilement utilisables, plus opérationnelles. Suivant [Kodratoff 1987] et [Mooney 1991], nous préférons voir dans l'EBL une technique conduisant à la formation de *nouveaux* concepts à partir d'un exemple et d'une théorie du domaine.

La théorie du domaine  $TD$  comporte des relations logiques qui vont permettre d'*expliquer* l'exemple  $E$ , c'est-à-dire de trouver un concept connu  $C$  qui détecte  $E$ . Cette explication peut être l'occasion de définir un concept  $C'$  plus restreint que  $C$  (*i.e.* tel que  $C' \Rightarrow C$ ) qui détecte  $E$ .

L'EBL utilise l'exemple pour diriger la formation d'un concept qui détecte l'exemple, et résout en cela le problème de la bidétermination (voir p. 151) :

<i>EBL</i>	
alphabet :	concepts (prédicats) de la théorie
modes de combinaison :	conjonction, instanciation de variables
choix des combinaisons :	généralisation d'une trace de preuve, sélection d'une portion d'explication liant une variable
mesure d'adéquation :	succès de l'unification et du processus de preuve

Prenons un exemple, adapté de [Mooney 1991] :

*Théorie du domaine :*

$\text{repose\_sur}(X,Y) \ \& \ \text{plat}(Y) \Rightarrow \text{stable}(X)$

$\text{peut\_être\_saisi}(X) \ \& \ \text{léger}(X) \Rightarrow \text{portable}(X)$

$\text{concave}(Y) \ \& \ \text{partie\_de}(Y,X) \ \& \ \text{vers\_le\_haut}(Y) \Rightarrow \text{récipient}(X)$

$\text{partie\_de}(Y,X) \ \& \ \text{forme}(Y, \text{demi-tore}) \ \& \ \text{adapté\_aux\_doigts}(Y) \ \& \ \text{rigide}(Y) \Rightarrow \text{peut\_être\_saisi}(X)$

$\text{isolant}(Y) \ \& \ \text{fait\_de}(X,Y) \ \& \ \text{petit}(X) \Rightarrow \text{peut\_être\_saisi}(X)$

$\text{isolant}(\text{céramique})$

$\text{isolant}(\text{porcelaine})$

$\text{petit}(X) \Rightarrow \text{léger}(X)$

$\text{diamètre}(X) > 2\text{cm} \ \& \ \text{diamètre}(X) < 12\text{cm} \Rightarrow \text{adapté\_aux\_doigts}(X)$

$\text{stable}(X) \ \& \ \text{portable}(X) \ \& \ \text{récipient}(X) \Rightarrow \text{tasse}(X)$

*description de l'exemple :*

$\text{appartient\_à}(\text{obj1}, \text{Fred}) \ \& \ \text{léger}(\text{obj1}) \ \& \ \text{couleur}(\text{obj1}, \text{rouge}) \ \& \ \text{partie\_de}(\text{h1}, \text{obj1}) \ \& \ \text{rigide}(\text{h1}) \ \& \ \text{diamètre}(\text{h1}) = 3\text{cm} \ \& \ \text{forme}(\text{h1}, \text{demi\_tore}) \ \& \ \text{repose\_sur}(\text{obj1}, \text{b1}) \ \& \ \text{plat}(\text{b1}) \ \& \ \text{concave}(\text{c1}) \ \& \ \text{partie\_de}(\text{c1}, \text{obj1}) \ \& \ \text{vers\_le\_haut}(\text{c1})$

A partir de cette description, un démonstrateur (par ex. le moteur Prolog) peut établir *tasse(obj1)*. L'analyse du raisonnement du démonstrateur (on parle d'explication) et sa généralisation (transformation de obj1, c1, b1 et h1 en variables) conduit à une condition suffisante pour qu'un objet soit une tasse :

$\text{léger}(X) \ \& \ \text{partie\_de}(H, X) \ \& \ \text{rigide}(H) \ \& \ \text{diamètre}(H) > 2\text{cm} \ \& \ \text{diamètre}(H) < 12\text{cm} \ \& \ \text{forme}(H, \text{demi\_tore}) \ \& \ \text{repose\_sur}(X, B) \ \& \ \text{plat}(B) \ \& \ \text{concave}(C) \ \& \ \text{partie\_de}(C, X) \ \& \ \text{vers\_le\_haut}(C) \Rightarrow \text{tasse}(X)$

Un autre exemple aurait pu conduire à :

$\text{petit}(X) \ \& \ \text{fait\_de}(X,Y) \ \& \ \text{isolant}(Y) \ \& \ \text{repose\_sur}(X, B) \ \& \ \text{plat}(B) \ \& \ \text{concave}(C) \ \& \ \text{partie\_de}(C, X) \ \& \ \text{vers\_le\_haut}(C) \Rightarrow \text{tasse}(X)$

Ces descriptions ne sont pas équivalentes au concept *tasse* de la théorie. Ce sont des spécialisations du concept *tasse* (appelons-les G001 et G002). Nous pourrions les appeler "tasse à anse" et "petite tasse" respectivement. La technique de l'EBL permet donc de *créer* des concepts en tant que spécialisations de concept connus. Elle permet aussi de créer de nouveaux concepts à partir d'une portion d'explication. Toutes les contraintes qui portent sur une même variable, par ex. la variable H, sont regroupées pour définir un concept inédit (que nous appellerions "anse") :

$\text{G003}(H, X) \Leftrightarrow \text{partie\_de}(H, X) \ \& \ \text{rigide}(H) \ \& \ \text{diamètre}(H) > 2\text{cm} \ \& \ \text{diamètre}(H) < 12\text{cm} \ \& \ \text{forme}(H, \text{demi\_tore})$

Nous avons maintenant :

$G001(X) \Leftrightarrow \text{léger}(X) \ \& \ G003(H, X) \ \& \ \text{repose\_sur}(X, B) \ \& \ \text{plat}(B) \ \& \ \text{concave}(C) \ \& \ \text{partie\_de}(C, X) \ \& \ \text{vers\_le\_haut}(C)$

Ces nouveaux concepts G001, G002, G003 peuvent être utilisés directement pour analyser des objets et les appairer à des concepts. Par exemple, G003 peut servir à reconnaître l'anse d'une valise.

Cette technique a le mérite de fabriquer des concepts et de les intégrer en même temps dans la théorie du domaine. On lui reproche parfois de ne pas créer de concepts réellement nouveaux. En accord avec [Mooney 1991], nous pensons que cela est inexact. S'il s'agit de créer des concepts à partir d'objets qui ne sont pas des concepts, alors aucun algorithme d'apprentissage symbolique ne crée de nouveaux concepts ! Toutes les techniques forment des concepts en combinant d'autres concepts. Pour SBL, ce sont les valeurs des attributs. Pour l'appariement structurel, ce sont les concepts de base et ceux de la hiérarchie. L'EBL forme des concepts nouveaux par spécialisation de concepts connus et par isolement de portions d'explications. En décrivant la formation des concepts comme un appariement symbolique, nous suggérons ici le fait que les humains forment de même de nouveaux concepts par combinaison de concepts déjà connus.

*A priori*, tout prédicat de la théorie du domaine peut être spécialisé par l'algorithme EBL. Dans l'exemple précédent, le premier exemple a pu créer une spécialisation de *portable* :

$G004(H, X) \Leftrightarrow \text{léger}(X) \ \& \ \text{partie\_de}(H, X) \ \& \ \text{rigide}(H) \ \& \ \text{diamètre}(H) > 2\text{cm} \ \& \ \text{diamètre}(H) < 12\text{cm} \ \& \ \text{forme}(H, \text{demi\_tore})$

Dans certaines applications, on désigne à l'avance les concepts qui seront spécialisés. Il arrive aussi que l'on mette des contraintes sur les prédicats utilisés dans l'expression de la spécialisation. En effet, la trace de la preuve qui permet de reconnaître l'exemple comme un concept connu peut être plus ou moins détaillée. On donne parfois la préférence à des prédicats qui ne sont pas nécessairement les plus détaillés, que l'on désigne comme "opérationnels". Cela permet de définir un langage "cible" dans lequel le nouveau concept sera exprimé. Par ex. si *peut\_être\_saisi* est un prédicat "opérationnel", alors G001 s'exprime de la manière suivante :

$G001(X) \Leftrightarrow \text{léger}(X) \ \& \ \text{peut\_être\_saisi}(X) \ \& \ \text{repose\_sur}(X, B) \ \& \ \text{plat}(B) \ \& \ \text{concave}(C) \ \& \ \text{partie\_de}(C, X) \ \& \ \text{vers\_le\_haut}(C)$

Les considérations sur le diamètre de h1 ont contribué à la preuve, mais elles ne sont pas conservées dans l'expression du nouveau concept.

Noter que les concepts créés font eux-mêmes partie du langage cible. L'EBL définit ainsi deux niveaux de connaissance : les connaissances exprimées dans le langage cible, qui sont supposées d'accès facile car elles peuvent s'apparier directement avec des exemples, et les connaissances "profondes" à accès supposé plus lent, car leur utilisation suppose la mise en oeuvre du démonstrateur.

## 8. Apprentissage de concepts pertinents

Les différentes techniques que nous venons d'évoquer sont autant de modèles qui montrent la faisabilité d'une formation des concepts dirigée par des exemples. Leur plausibilité cognitive peut être discutée, mais elles démystifient le problème de l'extraction des concepts logiques.

Le problème de la bidétermination (voir p. 151), qui est inhérent à tout apprentissage par l'expérience, est résolu ici par le *guidage* de l'appariement. Les mécanismes que nous avons mentionnés (SBL, appariement structurel, EBL) utilisent les exemples pour guider la formation de nouveaux concepts. Ces concepts, formés à partir de concepts connus, vont couvrir (*i.e.* détecter) les exemples qui ont aidé à leur formation.

On comprend grâce à ces exemples comment la formation d'un concept peut être dirigée par un ou quelques exemples. Mais on comprend aussi comment l'ancrage du nouveau concept, *i.e.* ses capacités de détection, sont définies. Le nouveau concept est formé par combinaison de concepts déjà connus, et ses capacités de détection découlent de ces combinaisons. Par exemple, si le concept logique *C* est défini comme

la conjonction de concepts  $C_j$ , un objet sera détecté par  $C$  dans un contexte donné dès qu'il est détecté par chaque  $C_j$  dans ce contexte.

Toutefois, ainsi que nous l'avions annoncé, les mécanismes que nous avons mentionnés sont tous insuffisants, et pour une même raison. Aucun de ces mécanismes, qu'il s'agisse du SBL, de l'EBL ou des appariements structurels, ne donne les *conditions* dans lesquelles un nouveau concept doit être formé. Par exemple, un algorithme d'EBL va former un concept dès qu'il le pourra, c.à.d. dès que l'exemple pourra servir à prouver un prédicat de la théorie.

Pour expliquer de manière plausible la formation des concepts logiques par les humains, il est nécessaire de résoudre les deux problèmes suivants :

- les individus sont exposés à un nombre de situations qui est très supérieur au nombre de concepts logiques qu'ils formeront dans leur vie. Etant donnée l'efficacité de certains modèles d'appariement qui autorisent la formation de concepts à partir d'une seule situation (performance courante chez les individus), il faut expliciter les conditions dans lesquelles la formation de concept aura lieu.
- nous nous sommes intéressés ici à la formation des concepts dirigée par l'expérience, c.à.d. par la perception directe de situations (voir note 3 p. 140). Mais nous n'oublions pas, bien entendu, que les concepts s'acquièrent aussi par le langage, notamment à la faveur des conversations. Serons-nous obligés de postuler deux mécanismes différents, l'un pour la formation directe des concepts, l'autre pour leur formation à travers le langage ?

Nous suggérons le fait que ces deux problèmes ont une solution commune. L'observation des conversations et des situations d'apprentissage nous a amenés à considérer que les concepts sont forgés à l'occasion de situations "problématiques". Des suggestions approchantes ont été faites dans la littérature. Par exemple Bersini [1990] suggère que nous n'utilisons nos capacités symboliques dans l'action qu'à la suite d'une "rupture" (*breakdown*), caractérisée par une surprise, une incertitude ou une appréhension. Nous retrouvons une idée analogue par exemple chez R. Schank :

"In any situation, a person has expectations about that situation. When all his expectations are confirmed, no learning takes place because none is necessary. Learning only takes place when an expectation is violated. After an expectation-failure, an individual must re-examine his expectations in order to revise his understanding of the situation and to prepare himself for it and for other similar situations in the future." [Schank & Edelson 1989]

Schank suggère même que la mémoire elle-même est organisée grâce à des ruptures (*failure*) qui servent d'indices pour rappeler des situations analogues :

"When our view of what should happen next is found to contradict the facts of the matter, we must attempt to *explain* it. That is, failures lead to attempts to *explain failures*". "[...] memory is organized, at least in part, by a classification of explanations of other people's behavior." [Schank 1980]

Nous suggérons de même le fait que la formation des concepts se produit dans des situations non routinières. Elle semble donc résulter d'un processus de traitement d'exceptions. Notre proposition consiste à définir ces conditions propices à la création conceptuelle de la manière suivante :

**Contexte "problématique"** : nous qualifions de "problématiques" les situations perçues comme *incohérentes*, hautement *improbables*, hautement *indésirables* ou *désirables*<sup>(24)</sup>. ||

*La formation et la mémorisation de nouveaux concepts exigent un contexte problématique.*

<sup>24</sup> Le terme "problématique" est peu approprié pour désigner les situations hautement désirables, mais celles-ci semblent jouer un rôle relativement mineur dans la création conceptuelle.

Cette contrainte sur les conditions de formation des concepts nous est dictée par l'observation des conversations spontanées. Notre propos ici est de l'étendre à la formation de concepts dirigée par l'expérience.

Nous ne disposons pas des justifications expérimentales qui nous permettraient d'ancrer cette extrapolation. Nous nous contentons donc de l'émettre en tant que suggestion, pour deux raisons principales : (1)-elle rappelle une intuition qui, comme nous l'avons signalé, est commune à de nombreux auteurs ; (2)-elle permet une unification séduisante avec les processus conversationnels. Elle renforce par là l'intérêt de l'étude des conversations pour la compréhension des processus d'apprentissage conceptuel.

## 9. Conclusion

Nous avons cherché ici à cerner les contraintes qui résultent de la systématisme de certains processus conceptuels sur l'apprentissage des concepts logiques par l'expérience. Ces contraintes nous ont amenés à postuler l'existence d'un *appareil conceptuel* hébergeant les représentations matérielles des concepts logiques, ceux-ci apparaissant alors comme distincts des perceptions intégrées. Le rapport entre les perceptions et les concepts apparaît alors comme un rapport de *déclenchement* (pour l'évocation de concepts connus) et d'*appariement* (pour la création de nouveaux concepts), et non comme un rapport de *ressemblance*. Nous avons présenté les concepts logiques comme des *détecteurs*, puis évoqué plusieurs mécanismes susceptibles de réaliser la formation de nouveaux concepts selon un appariement guidé par des exemples. Enfin, nous avons suggéré le fait que la formation de nouveaux concepts n'intervenait que dans des contextes problématiques.

Dans le chapitre II.2, nous avons montré que les connaissances conversationnelles ne pouvaient pas résulter d'une *extraction* à partir des situations vécues. Ici, nous avons montré que les concepts eux-mêmes, qui sont les termes de ces connaissances, ne pouvaient pas être issus d'une *extraction* à partir de perceptions. Les connaissances, comme les concepts, sont déclenchées par la situation, grâce à un mécanisme d'appariement. En présence d'une situation *problématique*, un mécanisme d'appariement logique conduit à une révision des connaissances (par exemple selon le processus de l'explication, cf. chap. III.4) alors qu'un mécanisme d'appariement conceptuel (du genre de ceux que nous avons évoqués dans ce chapitre) conduit éventuellement à la formation d'un nouveau concept.

Des contextes problématiques naissent chaque fois que nous percevons une incohérence, une rareté, un événement très positif ou très négatif dans notre environnement. Mais de très nombreux contextes problématiques nous sont présentés par d'autres personnes, lors des conversations. Les interlocuteurs semblent même contraints d'introduire tout nouveau sujet de conversation en le présentant comme problématique. Les conversations apparaissent ainsi comme le lieu privilégié du traitement des concepts : nous avons en effet défini les concepts comme des représentations mentales pouvant entrer dans une relation problématique (p. 139) et indiqué que ces dernières étaient propices à la création de nouveaux concepts. L'analyse des processus conversationnels qui permettent la création de nouveaux concepts et l'ajustement des concepts déjà connus devrait donc nous éclairer, en retour, sur la façon dont nous formons ou corrigeons nos concepts grâce à notre expérience directe.

Les réflexions menées dans ce chapitre ont des conséquences très contraignantes sur l'architecture de la cognition. Nous avons prédit l'existence *matérielle* d'un appareil conceptuel autonome, et nous avons décrit les concepts comme des détecteurs, de même nature que les détecteurs perceptifs du système visuel. Nous sommes conscients qu'une telle façon de voir a de quoi choquer tous ceux qui considèrent qu'il est impossible d'imaginer une réalisation neuronale biologiquement plausible des systèmes symboliques (comme le système conceptuel tel que nous le décrivons). Voici pourquoi nous consacrons le chapitre II.4 à la description d'une telle implantation.

# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

Nous proposons ici un modèle neuronal simplifié de la détection d'incohérences. Ce modèle n'a pas encore donné lieu à des implantations, mais notre objectif est de montrer que les traitements symboliques ne sont pas nécessairement étrangers à toute modélisation biologiquement plausible.

Le modèle fait intervenir un ensemble particulier de neurones, la *couche "C"*, spécialisé dans le traitement des concepts logiques. Cette couche implante les concepts logiques suivant un schéma ensembliste, et permet la détection des situations incohérentes, caractérisées par l'activation simultanée de représentations conceptuelles incompatibles.

Ce schéma permet de donner un début de réponse satisfaisant à plusieurs problèmes soulevés indépendamment : problème de la systématique et de l'unicité matérielle des processus conceptuels, problème de la compositionnalité et de la constituence, problème de la dualité extension / intension.

## 2. Traitement symbolique vs. traitement connexionniste

Les mécanismes logiques sont peut-être, parmi les mécanismes symboliques, ceux qui sont considérés comme les plus éloignés de ce que les modèles connexionnistes habituels sont capables de reproduire. L'une des rares manières d'approcher la performance des traitements logiques avec une architecture connexionniste consiste à remplacer les implications logiques  $[p \Rightarrow q]$  par des inférences statistiques [Smolensky 1988]. Nous avons montré ailleurs (voir chapitre II.2) à quel point une telle approximation de la performance logique est non satisfaisante. Rappelons simplement que l'inférence statistique est symétrique (corrélation), alors que l'implication ne l'est pas ; qu'elle exige un échantillon important, alors que l'implication peut être formée sans expérience du domaine (cf. chap. II.2) ; de plus la reconnaissance des incompatibilités logiques (par ex.  $[p \& \text{non } q]$ ), performance humaine quotidienne, est distincte d'un traitement statistique, puisqu'elle peut s'exercer sur des situations indépendamment de leur fréquence.

Or la systématique de certains processus logiques nous a amenés à postuler l'existence d'un appareil conceptuel autonome (voir chapitre II.3). Cet appareil est supposé héberger les représentations matérielles de tous les concepts. Comment imaginer un fonctionnement neuronal pour un tel appareil, qui serait capable d'effectuer des opérations logiques comme la reconnaissance d'incohérences ?

L'exigence de systématique a parfois été perçue, à tort, comme incompatible avec une implantation neuronale. Par exemple l'argument donné par J. Fodor et Z. Pylyshyn [1988], qui repose en grande partie sur la systématique des traitements linguistiques et logiques, est souvent cité pour opposer traitement symbolique et implantation connexionniste, alors qu'il consiste simplement à imposer des contraintes à une

telle implantation, contraintes auxquelles certaines architectures connexionnistes semblent pouvoir satisfaire [Chalmers 1990].

Les modèles qui cherchent à "réconcilier" les contraintes liées à l'approche connexionniste et celles liées à l'approche symbolique sont de deux types : d'une part les modèles ascendants ou homogènes, qui cherchent à faire émerger les symboles de fonctionnements sub-symboliques [Grumbach 1990 ; Hofstadter & McGraw 1993], et d'autre part les approches descendantes ou hétérogènes, qui garantissent une autonomie au niveau symbolique et accordent simplement un ancrage aux symboles [Harnad 1990].

Les approches hétérogènes (descendantes) sont souvent critiquées pour deux raisons. Premièrement parce que leurs implantations sont elles-mêmes hétérogènes : comme nous venons de le voir, le fonctionnement symbolique n'a pas de correspondant plausible dans ce que les architectures connexionnistes proposent couramment. Or celles-ci sont généralement reconnues comme plus proches du fonctionnement neuronal naturel.

La deuxième raison vient du problème de l'ancrage. Comment obtenir que les symboles du niveau supérieur aient une signification intrinsèque fournie par le niveau inférieur ?

Nous avons traité cette question de l'ancrage dans le cas des concepts (chap. II.3). Nous avons décrit les concepts comme des détecteurs, nous avons défini l'ancrage d'un concept par les objets que ce concept peut détecter dans une situation donnée, et nous avons décrit plusieurs mécanismes par lesquels de nouveaux concepts pouvaient être formés pour détecter de nouveaux objets.

Pour justifier la plausibilité de notre modèle, qui est résolument hétérogène, il nous faut donc maintenant suggérer une architecture vraisemblable pour le traitement symbolique des concepts. C'est ce que nous nous proposons de faire pour un cas particulier, qui jouit pour nous d'une grande importance, celui de la détection des incohérences.

### 3. Implémentation neuronale de la détection

Dans le chapitre II.3, nous avons présenté les concepts comme des détecteurs, au sens fonctionnel, mais en laissant entendre qu'ils étaient réalisés matériellement par des détecteurs neuronaux. Notre premier objectif ici va donc être de proposer une implantation neuronale pour les détecteurs (qu'ils interviennent dans le traitement conceptuel ou perceptif).

Dans [Dessalles 1993b], nous avons proposé une modélisation neuronale du phénomène de détection. Ce phénomène, que nous invoquons ici pour caractériser les concepts, intervient aussi dès les premiers niveaux de la perception et opère selon nous à chaque changement de niveau d'intégration. Nous avons donné des arguments pour suggérer le fait que le phénomène de détection ne peut pas être le fait de neurones uniques. Nous avons donc proposé un modèle qui décrit les détecteurs comme des ensembles localisés de neurones susceptibles d'entrer en activité synchrone, et nous avons qualifié ces détecteurs de *collectifs*, dans la mesure où aucune information n'est détenue par un seul neurone<sup>(1)</sup>. Nous reproduisons ici une partie de la description de ces détecteurs collectifs.

#### 3.1. Détection collective et cohérence temporelle

Nous partons de l'axiome suivant, dicté par un souci de plausibilité biologique : aucune fonction cognitive ne peut être supportée par un seul neurone. Ce sera en particulier le cas de la fonction de détection.

Tout modèle plausible du fonctionnement d'un détecteur se doit par conséquent de respecter cette contrainte de non-centralisation de l'information détectée. En particulier, l'activité du détecteur doit être portée par un phénomène physique défini au niveau *collectif* et dépourvu de signification au niveau d'un

---

<sup>1</sup> Le mot *collectif* est plus fort que le mot *distribué*. Dans un système collectif, les neurones individuels ne sont pas porteurs d'information, contrairement aux systèmes distribués, comme les réseaux connexionnistes habituels, dans lesquels chaque neurone détient une partie de l'information (fournie ou apprise).

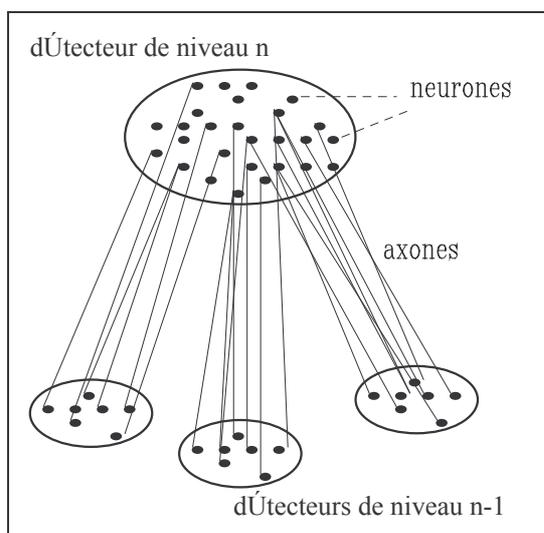
composant du détecteur. Il en est de même pour le système physique (connexion du détecteur) qui provoque l'activation : il doit lui aussi rester collectif.

Si un détecteur collectif est réalisé par un ensemble de neurones, par quoi se traduit son activation ? Pour quelques raisons que nous allons expliciter, nous suggérons le fait que cette activation peut être caractérisée par une activité cohérente des neurones qui composent le détecteur.

Dans le cas des neurones naturels, la cohérence se traduit par une émission synchronisée de potentiels d'action. Les neurones émettent spontanément des potentiels d'action, mais généralement sans coordination avec d'autres neurones. Nous définissons donc un détecteur comme un ensemble de neurones susceptibles de se synchroniser entre eux : le détecteur est actif si les neurones sont synchrones, et inactif si l'activité des neurones est désordonnée.

La cohérence temporelle est un bon candidat pour caractériser l'activation d'un détecteur collectif neuronal pour au moins trois raisons.

Premièrement, il s'agit d'une relation simple entre les neurones, et nous sommes "en droit" d'attendre une relation simple. En effet, le problème étant symétrique par nature, nous pouvons exiger du motif d'activation qu'il soit lui-même symétrique, ou stochastiquement symétrique (*i.e.* rôle statistiquement équivalent pour chaque neurone). La synchronisation est l'une des relations symétriques les plus simples que l'on puisse imaginer. L'avantage d'une relation simple est que son déclenchement pourra être simple lui aussi. On comprend aisément comment, dans une hiérarchie d'intégration perceptive, la synchronisation des neurones d'un détecteur de niveau donné  $n$  peut être obtenue par synchronisation avec les neurones afférents (de niveau  $n-1$ ), autrement dit par l'action des détecteurs afférents actifs au niveau  $n-1$ . Nous détaillerons cet aspect plus loin.



**figure 6** : implémentation neuronale de deux niveaux de détection collective. Les neurones d'un détecteur reçoivent des synapses excitatrices de leurs voisins (connexions latérales), ainsi que des synapses excitatrices en provenance de neurones appartenant à des détecteurs de niveau inférieur (ou à des capteurs). Si peu de détecteurs de niveau  $n-1$  sont actifs de manière synchrone, peu de neurones du détecteur de niveau  $n$  peuvent sommer une activité suffisante, et cette activité "se perd" dans les connexions latérales. En revanche, l'activité d'un nombre suffisant de détecteurs de niveau  $n-1$  permet au détecteur de niveau  $n$  de "s'embraser" : ses neurones deviennent synchrones.

Le deuxième argument vient de ce que la synchronisation neuronale est biologiquement vraisemblable. Elle est observée par exemple dans le cortex visuel, où l'on vérifie que les neurones répondant à une même zone du champ visuel, mais situés dans des aires différentes (comme l'aire V4 qui traite la couleur et l'aire

V5 (aire MT) où les mouvements sont détectés), deviennent synchrones lorsqu'ils répondent au même objet du champ visuel [Zeki 1992]<sup>(2)</sup>. Le même phénomène a été observé dans des zones sous-corticales comme le bulbe olfactif [Freeman 1991].

La troisième raison qui nous pousse à considérer la synchronisation comme un phénomène plausible pour marquer le résultat d'une détection collective vient de notre connaissance de mécanismes inter-neuronaux qui la rendent possible. Par exemple, une stabilisation des synapses qui suit la loi de Hebb peut rendre un réseau de neurones artificiels *cohérent* : deux neurones d'un réseau cohérent seront soit toujours simultanément actifs, soit jamais actifs en même temps.

Il reste à expliquer comment l'activation est transmise d'un niveau à l'autre, et grâce à quel câblage. Il s'agit ensuite de se demander comment un tel câblage peut s'établir.

### 3.2. Déclenchement d'une détection collective neuronale

Nous sommes confrontés à la question de savoir comment l'activité d'un détecteur collectif peut être déclenchée par l'activité de détecteurs collectifs du niveau inférieur, autrement dit nous devons préciser le type de câblage, au niveau neuronal, qui permet "d'alimenter" les détecteurs collectifs. Plus précisément, dans notre optique, il s'agit de comprendre comment l'activité synchrone de neurones de niveau  $n-1$  peut déclencher l'activité d'un groupe de neurones donné au niveau  $n$ . Nous devons en quelque sorte définir ce qui caractérise les niveaux d'un point de vue structurel.

#### 3.2.1. Des détecteurs collectifs cohérents

L'hypothèse la plus simple que nous pouvons faire semble être qu'un détecteur collectif est formé d'un ensemble "architecturalement cohérent" de neurones : les connexions internes au détecteur sont excitatrices, et les connexions avec des neurones externes (*i.e.* les neurones de la même couche, mais externes au détecteur) sont inexistantes<sup>(3)</sup>. Il peut s'ensuivre un phénomène de seuil, caractéristique de tout détecteur : si très peu de neurones du détecteur sont actifs de manière synchrone, alors les excitations qu'ils propagent ont peu de chance de se rencontrer en entrée d'autres neurones du détecteur pour y être sommées. Les neurones restent donc insensibles à l'activité de leurs voisins, et le détecteur garde une activité désordonnée. En revanche, sous l'effet d'une excitation synchrone extérieure, un nombre suffisant de neurones du détecteur peuvent devenir actifs et ainsi "embraser" le détecteur (cf. figure 6).

Un tel schéma suppose l'existence d'un câblage excitateur interne au détecteur, et d'un câblage excitateur en provenance des détecteurs afférents. La question se pose alors de savoir comment un tel câblage peut s'établir. Cette question est essentielle dans les systèmes naturels, pour lesquels il faut résoudre le contraste entre l'apparente complexité du câblage ( $10^{15}$  synapses) et la relative concision de son codage génétique (il y a moins de  $10^5$  gènes au total, et seule une petite partie de ces gènes a une influence sur la spécificité du câblage neuronal). Nous abordons maintenant cette question du câblage.

#### 3.2.2. Des détecteurs collectifs pseudo-locaux

Sans prétendre fournir une réponse précise, nous pouvons rappeler ici quelques éléments de réponse souvent mentionnés dans des contextes similaires. Le câblage excitateur interne peut être associé à la topologie des détecteurs. Si nous supposons que les détecteurs sont *pseudo-locaux*, c'est-à-dire qu'ils sont constitués non pas d'un seul neurone (hypothèse localiste extrême), mais de neurones proches, alors un système de connexions latérales rappelant le "chapeau mexicain" de T. Kohonen [1984] peut définir un ensemble de détecteurs collectifs avec un coût de codage architectural très faible. De plus, Lumer et Huberman [1992] ont montré qu'un tel ensemble de neurones (chez ces auteurs, il s'agit d'oscillateurs connectés entre eux avec des valeurs de couplage qui décroissent exponentiellement avec la distance de connexion) pouvait se synchroniser.

<sup>2</sup> Imaginer un petit objet rouge en mouvement horizontal. Les neurones détectant, dans cette zone du champ visuel, la couleur rouge (dans l'aire V4) et le mouvement horizontal (dans l'aire V5) seront synchrones entre eux.

<sup>3</sup> Des connexions externes inhibitrices permettraient d'augmenter le contraste, comme le note T. Kohonen [1984].

De plus, le câblage excitateur afférent peut être lui aussi réalisé avec un faible coût grâce aux propriétés topologiques de tels détecteurs collectifs pseudo-locaux. Il suffit que des connexions en provenance d'un détecteur afférent se projettent à l'intérieur du détecteur cible. La mise au point fine peut être assurée par une sélection synaptique. Seules sont conservées les synapses qui sont fonctionnelles au sens de Hebb : lorsque tous les détecteurs afférents, ou une grande partie d'entre eux, sont actifs, le détecteur cible devient actif aussi, même si le câblage est grossier. Les synapses qui favorisent cette synchronisation sont alors stabilisées. En revanche les synapses qui se perdent dans des détecteurs voisins, non actifs, dégénèrent. Une telle stabilisation peut s'opérer sans apport d'information extérieure, par exemple grâce à des générateurs aléatoires qui respectent la topologie [Shatz 1992]<sup>(4)</sup>. Le coût de codage d'une telle architecture reste faible.

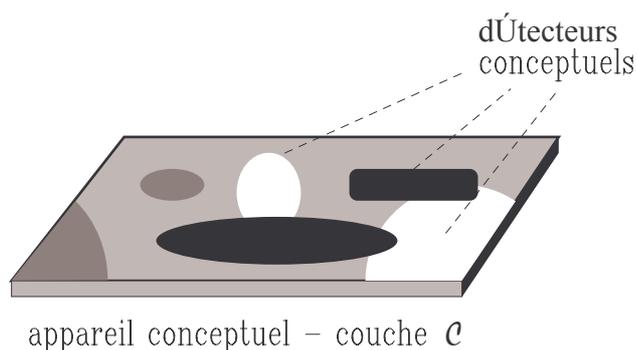
## 4. Modélisation neuronale de la détection d'incohérence

Nous proposons de considérer que la détection des incohérences, qui est une compétence humaine facile à mettre en évidence par l'observation des conversations, peut, malgré son caractère éminemment symbolique, être réalisée par un mécanisme neuronal plausible. Selon notre définition, une incohérence relie des concepts logiques. Nous suggérons le fait que tous les concepts logiques connus ou potentiellement connus d'un individu sont hébergés dans une structure unique que nous dénommons "appareil conceptuel". C'est cet appareil qui, dans notre modélisation, réalise la détection d'incohérences.

### 4.1. Architecture de l'appareil conceptuel

Nous allons tenter ici de suggérer un fonctionnement plausible pour un "appareil conceptuel" autonome. Nous avons décrit les concepts comme des détecteurs. Nous considérons maintenant qu'ils sont *réalisés* par des détecteurs neuronaux collectifs pseudo-locaux. Ces détecteurs sont reliés physiquement aux étages intégrés des appareils perceptifs par des faisceaux d'axones comme schématisé sur la figure 6. Leur entrée en activité dépend donc des perceptions.

Nous supposons que l'appareil conceptuel consiste en un ensemble de neurones impliqués dans des détecteurs collectifs dans lesquels les concepts sont implantés. Pour fixer les idées, imaginons cet ensemble bidimensionnel et appelons-le *couche*  $\mathcal{C}$  (figure 7). Les connexions internes à la couche  $\mathcal{C}$  sont excitatrices au sein des détecteurs (que nous supposons pseudo-locaux, au sens défini plus haut).



**figure 7** : représentation imagée de la couche " $\mathcal{C}$ ". Les représentations conceptuelles sont constituées d'ensembles de neurones connectés par des synapses excitatrices, qui constituent des détecteurs. Tous les neurones de la couche  $\mathcal{C}$  sont impliqués dans de tels détecteurs. Cette couche est reliée aux représentations intégrées de la perception et du langage. La représentation d'un concept actif *dans un contexte donné* est traduite par l'activité synchrone d'un ou plusieurs de ces détecteurs.

<sup>4</sup> Shatz a observé de tels générateurs aléatoires dans des rétines d'embryons.

L'appareil conceptuel tel que nous le décrivons héberge donc les représentations conceptuelles (détecteurs pseudo-locaux). Il reçoit des faisceaux de connexion en provenance des appareils perceptifs. Mais nous devons prendre en compte aussi le fait que les concepts sont intimement liés aux mécanismes linguistiques, et qu'ils sont capables d'évoquer des perceptions. On est donc en droit d'imaginer les connexions correspondantes. Toutefois, nous n'en parlerons pas ici, car ces aspects n'entrent pas dans le cadre de la modélisation partielle que nous décrivons.

#### 4.2. Principe de fonctionnement de la couche "C"

Chaque situation perçue va activer un certain nombre de détecteurs de la couche  $\mathcal{C}$ . Les neurones de ces détecteurs deviennent actifs de manière synchrone. Cette synchronisation est liée à la perception de la situation : il n'y a qu'une seule synchronisation possible à un moment donné, autrement dit la couche  $\mathcal{C}$  ne traite qu'une seule situation à la fois. Les neurones n'appartenant pas aux détecteurs activés par la situation ont une activité désordonnée.

*L'agencement des détecteurs est tel qu'une incohérence est détectée lorsque tous les neurones de la couche  $\mathcal{C}$  sont synchronisés. La couche  $\mathcal{C}$  elle-même se comporte alors comme un détecteur activé.*

La figure 7 montre de manière très schématique ce que peut être la couche  $\mathcal{C}$ . Les détecteurs conceptuels peuvent se recouvrir, et tout neurone de la couche appartient au moins à un détecteur. Noter qu'un neurone donné appartenant à un détecteur non activé peut néanmoins être en phase avec la synchronisation de référence : il suffit qu'il appartienne aussi à un détecteur activé. Ce raisonnement est valable pour tous les neurones du détecteur. Ce dernier peut donc se retrouver actif en quelque sorte "malgré lui", c.à.d. sans que son activité ait été provoquée par ses propres connexions afférentes (en provenance des systèmes perceptifs).

#### 4.3. Principe du câblage de la couche "C"

L'agencement des détecteurs permet une représentation directe des relations logiques. Si plusieurs concepts sont simultanément détectés dans une situation, les neurones de tous ces détecteurs dans la couche  $\mathcal{C}$  seront synchrones. Nous constatons donc que ce qui correspond à la conjonction logique (présence simultanée) est représenté par la *réunion* des détecteurs correspondant aux concepts conjoints.

Nous devons en fait attribuer à la couche  $\mathcal{C}$  des propriétés duales de celles accordées généralement à l'extension des concepts : le "et" logique correspond à l'*union* des détecteurs correspondants ; l'implication  $p \Rightarrow q$  est représentée par l'*inclusion inverse* des détecteurs  $D_q \subset D_p$ , et l'incohérence, par la couverture de toute la couche. De cette manière, et cela constitue l'élément central de notre modèle :

*L'ensemble des neurones de la couche  $\mathcal{C}$  encode les relations logiques grâce à un schéma ensembliste.*

<i>relation logique entre concepts</i>		<i>relation ensembliste entre détecteurs</i>	
conjonction	$r = p \ \& \ q$	union ensembliste	$D_r = D_p \cup D_q$
implication	$p \Rightarrow q$	inclusion ensembliste	$D_q \subset D_p$
incompatibilité	$[p \ \& \ q] \Rightarrow F$	totalité	$D_p \cup D_q = \text{couche } \mathcal{C}$

#### 4.4. Formation des concepts dans la couche "C"

L'agencement des détecteurs conceptuels dans la couche  $\mathcal{C}$  doit être issu de l'encodage des relations logiques. Comment cela se peut-il ? Nous ne prétendons pas être en mesure de résoudre ce problème, mais simplement suggérer le fait qu'il n'est pas insoluble.

Nous nous plaçons dans le cadre d'un mécanisme d'appariement pour expliquer la formation des concepts (chap. II.3). Nous postulons donc l'existence, préalable à toute expérience, d'un certain nombre de détecteurs conceptuels, que nous appelons détecteurs primitifs<sup>5</sup>. Leur agencement permet la perception de certaines incohérences, c'est-à-dire que la synchronisation de plusieurs d'entre eux entraîne la synchronisation de toute la couche  $\mathcal{C}$ .

Dans notre modèle, le câblage interne qui caractérise les détecteurs primitifs de la couche  $\mathcal{C}$  ne se modifie pas après sa mise en place. Il est caractérisé par des synapses fortement excitatrices, il définit la topologie des détecteurs primitifs, et donc leurs relations d'incompatibilité. Nous devons donc nous demander comment sont codés les concepts non primitifs et leurs relations d'incompatibilité.

Les concepts formés par les humains semblent exprimables par des conjonctions de concepts connus [Johnson-Laird 1983:424; Medin et al. 1987; Richard 1990:68]. Si l'on remonte jusqu'aux concepts primitifs postulés dans notre modèle (ce qu'un individu est peut-être incapable de faire de manière explicite), tout concept  $C$  peut être défini en théorie comme une conjonction de concepts primitifs :  $C = C_1 \& C_2 \& \dots \& C_n$ . Les neurones de la couche  $\mathcal{C}$  vont traduire cette relation : le détecteur conceptuel  $D_C$  qui représente le concept  $C$  regroupe, au sens de la réunion ensembliste, les neurones impliqués dans les détecteurs primitifs  $D_{c_i}$ . Il nous faut donc caractériser matériellement les capacités de détection de  $D_C$ , ce qui revient à justifier l'existence d'un détecteur non primitif comme  $D_C$  en tant que groupe de neurones de la couche  $\mathcal{C}$  capables de se synchroniser.

Nous avons vu qu'un détecteur pseudo-local devait son existence à un câblage excitateur interne et à un câblage afférent en provenance d'autres détecteurs. L'ensemble de neurones  $D_C$  possède en partie ce câblage excitateur interne. Toutefois  $D_C$  est hétérogène, formé de régions (les détecteurs primitifs  $D_{c_i}$  qu'il réunit) fortement connectées en interne, mais éventuellement non connectées entre elles. Une solution possible pour caractériser  $D_C$  consiste à postuler l'existence d'un *second câblage* de la couche  $\mathcal{C}$ . À côté du câblage de base aux synapses fixées sur des valeurs fortement excitatrices, nous supposons qu'il existe un deuxième câblage dont les synapses sont, elles, susceptibles de changer de valeur.  $D_C$  sera ainsi caractérisé par des synapses excitatrices de ce deuxième câblage, dont les valeurs sont issues de l'expérience selon un mécanisme à préciser, et qui relie des neurones de  $D_C$  (sans tenir compte des limites des détecteurs primitifs qui le composent)<sup>6</sup>.

Nous appellerons ces détecteurs non primitifs, capables eux aussi de se synchroniser pour certaines configurations des perceptions intégrées, *détecteurs mémorisés*.

<sup>5</sup> Cette hypothèse concernant les primitives conceptuelles est parfois dénoncée comme non parcimonieuse. Il faut cependant bien comprendre qu'elle est indissociable de celle du mécanisme d'appariement. Rappelons que l'alternative, c'est-à-dire le refus de tout appariement, obligerait à considérer que les connaissances conceptuelles résultent d'un apprentissage de type "indifférent" (i.e. qui tolère tout changement de repère dans le codage de ses entrées). Nous avons montré qu'un mécanisme indifférent n'est pas assez efficace pour expliquer l'apprentissage des connaissances manipulées, par exemple, dans les conversations (chap. II.2). De plus, un mécanisme indifférent ne permet qu'un encodage *implicite* des connaissances, ce qui exclut toute manipulation logique (chap. II.3). Or les humains sont capables de manipulations logiques (e.g. le mécanisme de recevabilité d'une explication). L'abandon des systèmes indifférents nous conduit forcément à postuler l'existence de "connaissances" *a priori*. Dans le mécanisme d'appariement que nous proposons, ces connaissances prennent la forme de primitives conceptuelles.

<sup>6</sup> Noter que c'est ce deuxième câblage qui rend nécessaire, au niveau matériel, la centralisation physique de l'appareil conceptuel que nous avons postulée dans le chapitre II.3.

**Détecteur primitif** : la couche  $\mathcal{C}$  comporte, du fait de son câblage interne de base, un certain nombre de détecteurs, appelés détecteurs primitifs, capables de se synchroniser pour certaines configurations des perceptions intégrées. Tout neurone de la couche appartient à un tel détecteur. La structure de ces détecteurs primitifs, ainsi que leurs capacités de détection, sont supposées indépendantes de l'expérience de l'individu.

**Détecteur mémorisé** : les perceptions intégrées relatives à certaines situations peuvent activer simultanément plusieurs détecteurs primitifs. Une telle configuration (réunion de détecteurs primitifs) peut être mémorisée, dans certaines conditions. Nous suggérons pour cela l'existence d'un câblage secondaire dans la couche  $\mathcal{C}$  dont certaines synapses reliant des neurones de  $D_{\mathcal{C}}$  seraient positionnées, dans les conditions de mémorisation et selon un mécanisme à définir, sur des valeurs excitatrices.

Les détecteurs mémorisés sont davantage que des réunions ensemblistes de détecteurs primitifs. Ce sont des représentations "structurées" (voir définition p.175), qui réunissent d'autres détecteurs mémorisés moins étendus.

Ce qui distingue l'établissement de  $D_{\mathcal{C}}$ , en tant qu'ensemble de neurones susceptibles de se synchroniser, du type de mémorisation que l'on observe dans un réseau associatif habituel, c'est qu'il peut être réalisé en une fois. Nous ne sommes pas en mesure de suggérer un mécanisme détaillé de ce processus de mémorisation. Ce que nous suggérons, et que nous avons suggéré dans d'autres contextes (chap. II.3 et III.2), c'est que l'établissement d'un nouveau détecteur comme  $D_{\mathcal{C}}$ , autrement dit la formation d'un nouveau concept  $c$ , se produit préférentiellement lorsqu'une incohérence a été détectée. Voyons comment.

Le fait que la couche  $\mathcal{C}$  entre complètement en synchronisation (c.à.d. qu'elle devienne active exactement comme les détecteurs qui la composent) est utilisé comme signal pour une révision des concepts activés par la situation. Cette révision peut entraîner une création de concept ou une modification des concepts existants, à travers le mécanisme de l'*explication* que nous avons décrit ailleurs (voir chapitre III.4). Ce genre de phénomène, dans le contexte de cette modélisation neuronale, peut être décrit en 4 phases :

- 1- mise en synchronisation de nombreux détecteurs primitifs de la couche  $\mathcal{C}$ , due souvent à une perception sommaire (moins la perception est détaillée, plus les suppositions "par défaut" sont actives ; en termes connexionnistes, la généralisation au niveau des mécanismes perceptifs est plus importante dans le cas d'observations partielles) ;
- 2- détection d'une incohérence par entrée en synchronisation de tous les neurones de la couche  $\mathcal{C}$  ;
- 3- révision de la perception (comme dans l'étape 1), l'attention plus soutenue conduisant à des informations plus complètes, et donc éventuellement à moins de détecteurs activés ;
- 4- mémorisation éventuelle de l'exception : un aspect de la perception avait été négligé dans la phase 1 ; du fait de sa prise en compte, la couche  $\mathcal{C}$  n'est plus entièrement synchrone. Par différence avec la situation en -2-, un concept (groupe de détecteurs) peut être mémorisé ou redélimité, et ses liens (relation d'incompatibilité) avec les autres concepts actifs établis ou corrigés.

## 5. Points forts du modèle

### 5.1. *Compatibilité avec les contraintes de la systématité*

L'une des principales caractéristiques de ce modèle est d'héberger en une même structure les représentations de tous les concepts. Ceci constitue, on s'en souvient (cf. chap. II.3), une exigence liée à la systématité de certains processus conceptuels : systématité des résultats, systématité de prise en compte. Nous avons pris l'exemple de la recevabilité d'une explication, dont la systématité exigeait qu'elle fût un processus matériellement unique (systématité des résultats) et que sa réalisation matérielle opérât à partir

d'un appareil conceptuel lui-même unique hébergeant toutes les représentations conceptuelles (systématicité de prise en compte).

Pour une incohérence que nous représentons de la façon suivante :

$$[p_1 \& p_2 \& \dots p_n] \Rightarrow \mathbf{F}$$

$q$  est une tentative d'explication recevable dès que  $q = \text{non } p_i$ . Dans le modèle neuronal que nous venons d'esquisser, la couche  $\mathcal{C}$  constitue ce lieu où toutes les représentations conceptuelles sont hébergées, et le mécanisme de recevabilité peut directement opérer sur la couche  $\mathcal{C}$ . Lorsque  $p_i$  cesse d'être perçu, la situation est révisée (au sens de l'étape -3- dans le scénario esquissé ci-dessus), et le détecteur correspondant au terme  $p_i$  a toutes les chances de n'être plus synchrone<sup>(7)</sup>. Cela a pour effet de supprimer la perception de l'incohérence, car par définition  $p_i$  était nécessaire à l'incompatibilité. L'inactivation de  $p_i$ , dans le contexte de l'incohérence ci-dessus, définit le concept ( $\text{non } p_i$ ) qui est ainsi reçu comme une tentative d'explication, puisque sa présence supprime l'incohérence initiale.

En revanche, il faut noter que l'évocation d'un concept  $r$  qui serait sans rapport avec l'incompatibilité n'a aucun effet sur l'état de la couche  $\mathcal{C}$  qui reste entièrement synchrone :  $r$  ne sera pas reçu comme une tentative d'explication<sup>(8)</sup>.

## 5.2. Constituence des représentations

Dans leur célèbre article, J. Fodor et Z. Pylyshyn [1988] insistent sur le fait que les représentations mentales, en particulier les concepts, doivent avoir un certain nombre de propriétés, dont nous allons mentionner certaines. Tout modèle de la cognition doit, selon ces auteurs, respecter ces propriétés. Nous voulons indiquer ici que notre modèle respecte les propriétés indiquées par Fodor et Pylyshyn. La principale qui nous intéresse ici est la propriété de *constituence*, qui concerne les représentations mentales structurées :

**Représentations structurées** : Les éléments d'un ensemble de représentations mentales sont structurés si certaines de ces représentations sont atomiques, et que les autres sont moléculaires, *i.e.* résultent de la combinaison de composants qui sont eux-mêmes des représentations atomiques ou moléculaires.

Les concepts logiques, dans notre modèle, ont des représentations structurées. Les représentations atomiques sont les détecteurs primitifs, et les représentations moléculaires sont les détecteurs mémorisés, qui résultent de la réunion ensembliste d'autres détecteurs, primitifs ou mémorisés.

**Constituence** : un ensemble de représentations structurées a la propriété de constituence si les composants des représentations sémantiquement interprétables de cet ensemble sont eux-mêmes sémantiquement interprétables.

Dans notre modèle, l'interprétation sémantique correspond aux capacités de détection. Un détecteur est mémorisé par la fixation de la co-activité de plusieurs détecteurs, dans des conditions particulières (consécutives à la détection d'une situation problématique, par ex. incohérente). Tout nouveau détecteur ainsi formé possède des capacités de détection, il est donc sémantiquement interprétable. Les détecteurs qui le composent, c.à.d. les détecteurs qui étaient activés au moment de sa mémorisation, sont aussi sémantiquement interprétables pour la même raison. La propriété de constituence des représentations est donc vérifiée dans notre modèle.

<sup>7</sup> Si ce n'est pas le cas, autrement dit si la perception continue à évoquer  $p_i$  alors que  $\text{non } p_i$  est évoqué par ailleurs (par ex. par le langage), il y a détection d'une autre incohérence par la couche  $\mathcal{C}$  qui masque l'incohérence initiale.

<sup>8</sup> Ceci est vrai même si  $r$  est relatif à un élément de la situation :  $r = \text{non } s$  où  $s$  a été déclenché dans un premier temps. La désactivation de  $s$  n'affectera pas la détection de l'incohérence, si  $s$  n'était pas un élément de l'incompatibilité. Certes le détecteur  $D_s$  associé à  $s$  ne sera peut-être plus excité par les perceptions qui avaient provoqué son activation dans la première phase. Cependant, les neurones de  $D_s$  resteront synchrones avec tous les autres neurones de la couche  $\mathcal{C}$ , car ils appartiennent à d'autres détecteurs que  $D_s$ . Ainsi,  $s$  ne sera pas détecté comme une tentative d'explication, car la couche  $\mathcal{C}$  restera entièrement synchrone.

Noter que la constituence est obtenue, dans notre modèle, par les propriétés d'inclusion matérielle des constituants dans les représentations, comme l'avaient plus ou moins prévu Fodor et Pylyshyn :

[Properties like constituency] "constrain the physical realizations of symbol structures. In particular, the symbol structures in a Classical model are assumed to correspond to real physical structures in the brain and the combinatorial structure of a representation is supposed to have a counterpart in structural relations among physical properties of the brain. For example, the relation 'part of', which holds between a relatively simple symbol and a more complex one, is assumed to correspond to some physical relation among brain states." [Fodor & Pylyshyn 1988]

Nous pouvons extrapoler en disant que des représentations structurées sont nécessairement matériellement structurées :

**représentations matériellement structurées** : des représentations structurées sont matériellement structurées si elles contiennent (au sens de l'inclusion ensembliste) leurs composants.

Nous pouvons écrire symboliquement : constituence  $\Rightarrow$  structuration  $\Rightarrow$  structuration matérielle.

Le fait que notre modèle vérifie les contraintes identifiées par Fodor et Pylyshyn est tout sauf anodin. L'argument de Fodor et Pylyshyn vise principalement les architectures connexionnistes classiques qui, selon eux, ne respectent aucune des trois propriétés ci-dessus (ainsi que d'autres qui en découlent). Ils décrivent les architectures connexionnistes "orthodoxes" comme utilisant la transmission d'activité entre les noeuds d'un graphe. Ces noeuds, qui peuvent être des neurones ou des groupes de neurones, sont les unités sémantiques du système. Ces noeuds sont tous considérés comme sémantiquement atomiques, et l'architecture connexionniste ainsi décrite ne contient pas de représentation matériellement structurée. Elle ne peut donc pas héberger des représentations structurées, et vérifie encore moins la propriété de constituence <sup>(9)</sup>.

Le modèle neuronal que nous proposons dans ce chapitre pour la détection de l'incohérence semble respecter la propriété de constituence, principalement parce que ses représentations sont matériellement structurées. Par exemple, le détecteur correspondant à  $[p \& q]$  correspond à l'union des neurones impliqués dans les détecteurs associés à  $p$  et à  $q$ . Il *contient* donc les représentations de  $p$  et de  $q$ .

Qu'en est-il de la représentation de  $[p \text{ ou } q]$  ? La contrainte de constituence exige que si  $[p \text{ ou } q]$  a une représentation, alors elle doit inclure la représentation de  $p$  et celle de  $q$ . Cela semble poser un problème majeur à notre modèle. En effet, dans notre représentation neuronale des relations logiques, on devrait s'attendre à ce que  $[p \text{ ou } q]$  ait comme représentation *l'intersection* des représentations de  $p$  et de  $q$ , ce qui serait tout à fait contraire à l'exigence de structuration matérielle qui voudrait que la représentation de  $[p \text{ ou } q]$  contînt celle de  $p$  et celle de  $q$ . De plus, les représentations des concepts dans la couche  $\mathcal{C}$  sont pour nous des détecteurs primitifs, ou la réunion de plusieurs détecteurs primitifs. Rien ne garantit que l'intersection de deux détecteurs sera une telle représentation : elle risque d'inclure des "morceaux" de détecteurs primitifs <sup>(10)</sup>, et donc de ne pas avoir de capacité de détection propre.

En fait, c'est dans ce dernier point que réside la solution. Pour nous,  $[p \text{ ou } q]$  N'A PAS de représentation directe dans la couche  $\mathcal{C}$ , dans la mesure où généralement  $D_p \cap D_q$  n'est pas un détecteur. En termes cognitifs,  $p \text{ ou } q$  n'est pas un objet élémentaire de la pensée <sup>(11)</sup>. Ce qui peut être représenté, c'est l'incohérence  $[non p \& non q] \Rightarrow F$ . La sortie de cette incohérence exige "que  $p$  ou que  $q$ " soient vrais. D'après notre modèle, un individu ne pense  $[p \text{ ou } q]$  qu'en ayant à l'esprit l'impossibilité que  $non p$  et  $non q$  soient tous deux vrais <sup>(12)</sup>. Le problème de la structuration matérielle ne se pose donc pas pour des expressions comme  $[p \text{ ou } q]$  qui, dans notre modèle, n'ont pas de représentation directe.

<sup>9</sup> Certaines architectures connexionnistes semblent échapper à cette description [Chalmers 1990].

<sup>10</sup> Dans sa version actuelle, le modèle ne suppose aucunement que les détecteurs primitifs soient disjoints.

<sup>11</sup> Cela rejoint la préférence manifeste que montrent les individus pour les définitions conjonctives, cf. p. 173.

<sup>12</sup> On peut dire que  $[p \text{ ou } q]$  a une représentation mentale dynamique ( $p$  et  $q$  sont "pensés" en séquence), contrairement à la représentation statique de  $[p \& q]$ .

Il en est exactement de même pour la négation logique *non p* d'un concept  $p^{(13)}$ . (*non p*) n'existe pas en tant que représentation mentale directe (sauf dans les cas où *non p* coïncide avec un détecteur). Comme nous l'avons annoncé dans le chapitre II.3, l'individu pense *non p* seulement après avoir pensé que *p* participe à un contexte problématique. (*non p*) n'intervient donc que comme concept temporaire, en tant qu'invalidation d'une première relation problématique. Noter que dans le contexte de cette relation problématique, (*non p*) correspond souvent à un détecteur. Par exemple, nous avons vu dans le chapitre I.1 un extrait où, dans le contexte de la conversation, *remorque* et *voiture break* apparaissaient comme l'exacte négation l'une de l'autre.

Citons encore une propriété exigée par Fodor & Pylyshyn pour un système cognitif, la propriété de compositionnalité, qu'ils déniaient aux architectures connexionnistes "orthodoxes", et qui semble vérifiée par notre modèle. Nous exprimons cette propriété ainsi :

**Compositionnalité sémantique** : un ensemble de représentations structurées vérifie la propriété de compositionnalité si l'interprétation sémantique d'une représentation *R* peut être calculée à partir de l'interprétation sémantique des constituants de *R* et de leurs relations dans *R*.

Les représentations de notre modèle sont clairement compositionnelles : la sémantique d'un détecteur *D* réside dans ses capacités de détection, et celles-ci peuvent être calculées par conjonction des capacités de détection des composants de *D*.

### 5.3. *Compatibilité avec les modèles associatifs*

Nous avons défini les concepts par leurs propriétés logiques (capacité à entrer dans une relation d'incohérence, cf. chap. II.3. Mais nous n'avons jamais nié l'importance des propriétés associatives des concepts (propriétés qu'ils partagent avec d'autres représentations mentales). Il semble que ces propriétés soient d'ailleurs considérées comme fondamentales par certains auteurs.

Parmi les aspects de notre modèle que nous pouvons considérer comme des points positifs, nous devons précisément noter le fait qu'il permet la cohabitation des processus logiques et des processus associatifs qui opèrent sur les concepts. En effet les concepts, même s'ils entrent dans des relations logiques comme la relation d'incompatibilité, ont néanmoins de nombreuses propriétés de type associatif, qu'ils partagent d'ailleurs avec d'autres représentations mentales (voir chap. II.3). Les associations et analogies comme celles qui sont étudiées dans le programme COPYCAT [Hofstadter & Mitchell 1988] seraient réalisées, dans notre modèle, par des transferts d'activation au sein du deuxième câblage de la couche *C* que nous avons dû postuler pour justifier de l'existence de concepts mémorisés. Toutefois, nous sommes loin d'avoir élucidé le fonctionnement de ce câblage, en particulier les aspects *dynamiques* de ce fonctionnement qui semblent essentiels pour le déclenchement des analogies.

### 5.4. *Simplicité et plausibilité neuronale*

Le modèle du traitement des incohérences que nous venons de proposer repose sur peu de principes. Tout d'abord le mécanisme de la détection, réalisé par des détecteurs pseudo-locaux qui sont caractérisés par des synapses internes excitatrices, et dont l'activation se traduit par une synchronisation. Ce type de détection ne suppose aucune propriété des neurones autres que celles qui sont bien établies sur un plan biologique.

Nous avons postulé ensuite l'existence de détecteurs "précâblés" dans la couche *C*. Les organismes vivants disposent de nombreux dispositifs neuronaux précâblés, comme les détecteurs de segments orientés des aires visuelles primaires, mis en évidence dans les expériences de D. Hubel et T. Wiesel [1979], ou comme les mécanismes innés de déclenchement (les *Innate Release Mechanisms* postulés par les éthologistes [Lorenz 1978]).

Nous avons essayé de montrer que le coût architectural, et donc génétique, de détecteurs comme ceux de la couche *C* est relativement faible. Le principe de fonctionnement de la couche *C* elle-même présente

<sup>13</sup> Je remercie Bernard Victorri d'avoir attiré mon attention sur ce point concernant la négation.

l'intérêt d'être le même que celui des détecteurs qu'elle contient : c'est sa synchronisation qui traduit la détection d'une incohérence. Nous pouvons imaginer, en termes évolutifs, qu'elle soit issue d'un détecteur pseudo-local standard. Son câblage interne comporte deux types de câblage, l'un fixe, l'autre de type associatif. Elle a donc une structure interne, ce que n'a pas le détecteur de base. Nous pouvons supposer que la couche  $\mathcal{C}$  n'est pas une création unique, et émettre l'hypothèse que des structures complexes comme la couche  $\mathcal{C}$  sont impliquées dans d'autres opérations cognitives, y compris la perception.

### 5.5. *Parallélisme et temps de réponse*

Le sentiment qu'une situation est incohérente est souvent un sentiment quasi instantané. Contrairement aux systèmes symboliques couramment implantés dans les ordinateurs, les individus ne prennent pas le temps d'enchaîner des inférences pour juger de l'incompatibilité de faits perçus. Le modèle que nous proposons permet de rendre compte de ce temps de réponse étonnant.

Le problème de la détection d'une incohérence ne doit pas être confondu avec le problème, de complexité exponentielle, de savoir si  $n$  termes donnés sont toujours incompatibles (problème de la satisfiabilité). Le problème résolu par notre modèle neuronal est celui de la *satisfaction* d'une relation d'incompatibilité dans une certaine interprétation. L'interprétation (au sens logique, *i.e.* la valuation qui assigne une valeur de vérité aux propositions) est fournie par la situation perçue. Le fonctionnement de la couche  $\mathcal{C}$  permet de savoir si, compte tenu de ces valeurs de vérité, une contradiction sémantique est obtenue. Ce problème de satisfaction d'une relation d'incompatibilité est ainsi résolu de manière parallèle par notre modèle.

### 5.6. *Lien logique - ensembles*

Ce modèle de la détection des incohérences permet, comme souvent en pareil cas, de résoudre un problème qui était passé inaperçu. En effet, la logique correspond souvent à des raisonnements abstraits totalement déconnectés de toute interprétation sur des collections d'objets (penser à un raisonnement du genre "S'ils se sont disputés, alors il refusera de lui écrire !"). Par ailleurs, tout humain semble capable de raisonner sur des ensembles (possession et partage d'ensembles d'objets, inclusion de territoires, etc.). Par exemple nous pouvons colorier des intersections ou des réunions dans un diagramme de Venn sans jamais réfléchir aux liens logiques entre les appartenances des éléments. Appelons cette capacité "compétence spatio-ensembliste", dans la mesure où elle semble mettre en jeu nos capacités de représentation spatiale.

Or la logique cadre parfaitement avec une interprétation spatio-ensembliste : l'intersection et la conjonction, l'implication et l'inclusion, la disjonction et la réunion, etc. sont autant de paires de notions dont les termes peuvent être définis indépendamment, et que l'esprit humain ne peut pourtant pas dissocier dès qu'il prend en compte la relation d'appartenance. Par ex.  $\forall x ; x \in A \Rightarrow x \in B$  si et seulement si  $A \subset B$ . Au point que dans la tradition philosophique, les concepts ont toujours eu deux définitions simultanées ! L'une logique (intensionnelle) et l'autre ensembliste (extensionnelle). Par exemple J-B. Grize [1967] présente la "logique des classes" comme faisant presque double emploi avec la logique des propositions !

Mathématiquement, il y a homomorphisme entre les deux. Si  $\mathcal{P}(E)$  est l'ensemble des parties d'un ensemble  $E$ , et  $\Phi$  la bijection de  $\mathcal{P}(E)$  dans l'ensemble des prédicats logiques sur  $E$  (*i.e.* les fonctions de  $E$  dans  $\{V;F\}$ ) :  $\Phi(A)(x) = V$  ssi  $x \in A$ , alors on a par exemple  $\Phi(A \cap B) = \Phi(A) \& \Phi(B)$ .

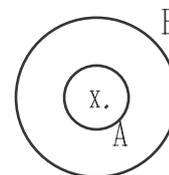
Certes nous avons clairement opté pour une définition intensionnelle des concepts logiques (chap. II.3), et nous avons indiqué que l'extension des concepts ne pouvait pas être définie parmi les objets du monde réel. Mais il reste que dans les situations axiomatisées (ou si l'on préfère, dans les contextes très clairement circonscrits), les concepts logiques peuvent jouer leur rôle de détecteur de manière parfaite : on peut décider sans ambiguïté, dans ces cas-là, quel est l'ensemble des objets qui correspondent au concept. On retombe alors sur cette adéquation obligée logique / ensembles.

Or cette adéquation est, *a posteriori*, pour le moins troublante. Elle nous est tellement naturelle que nous omettons de nous en étonner, mais elle est étonnante. Pourquoi notre compétence logique et notre compétence spatio-ensembliste montrent-elles une telle convergence parfaite ?

Jean Piaget est l'un des rares auteurs à avoir soulevé cette question, et à avoir suggéré une réponse. Pour Piaget, notre compétence logique est construite, elle *dérive* de l'évolution des opérations que nous maîtrisons au cours de notre enfance, entre autres de notre expérience spatio-ensembliste [Piaget & Inhelder 1947:530].

G. Lakoff aussi a repéré la convergence entre notre compétence spatio-ensembliste et notre compétence logique. Lui aussi fait dériver la deuxième de la première :

"One of the most interesting properties of image-schemas is that they have built-in logics. For example, BOUNDED REGION schemas, also called CONTAINER-schemas, have essentially a Boolean logic. Consider two CONTAINER-schemas, A and B, such that A is in B, and an object X is in A. We 'see' instantly, without doing any logical deduction, that X is in B. All we need to do is to shift our focus to the relationship between X and B." [Lakoff 1988]



Nous considérons pour notre part que la compétence spatio-ensembliste et la compétence logique *ne dérivent pas* l'une de l'autre. Ce genre d'hypothèse nous semble invraisemblable, compte tenu des observations que nous avons pu faire. Piaget observe la compétence logique en recherchant la trace d'un raisonnement hypothético-déductif. Or l'étude des conversations nous conduit à d'autres conclusions : les capacités conversationnelles des jeunes enfants, dès 3ans, mettent en jeu des compétences logiques qui dépassent largement certaines de leurs capacités spatio-ensemblistes telles qu'elles sont apparentes dans les tests psychologiques [Garitte & Dessalles 1994].

Nous pensons donc que ces deux compétences apparaissent indépendamment chez l'enfant. Mais dans ce cas, comment expliquer qu'il existe entre elles une telle adéquation, que l'on peut caractériser par un homomorphisme mathématique ?

Le modèle neuronal que nous avons décrit ici apporte un élément important qui rend beaucoup moins surprenante cette adéquation mathématique entre compétence logique et compétence spatio-ensembliste, sans poser l'exigence, selon nous peu vraisemblable, d'une origine spatiale ontogénétique de la logique ni, inversement, d'une origine logique de la perception spatiale. Selon notre modèle, les relations logiques (relations d'incompatibilité, d'implication, etc.) sont représentées de manière *ensembliste* dans le cerveau. Dans ces conditions, la coïncidence semble beaucoup moins "miraculeuse", car il est beaucoup plus facile d'accepter qu'il existe une adéquation mathématique entre notre compétence spatio-ensembliste et l'implantation neuronale, ensembliste, des relations logiques. Noter toutefois que la relation d'appartenance impose, dans la perception spatio-ensembliste, des relations duales de celles qui existent entre les neurones de la couche  $\mathcal{C}$ .

## 6. Limites du modèle

Le modèle neuronal de la détection d'incohérences esquissé ici est loin d'être complet. Un certain nombre de points restent à préciser, avant que l'on puisse songer à une implémentation :

- la nature des détecteurs primitifs. Ceux-ci sont souvent postulés, mais rares sont les auteurs qui se risquent à donner des exemples de concepts innés. A titre de simple illustration et sans aucune justification, on peut suggérer par exemple le concept d'entité spatio-temporelle (cf. la "whole object assumption" [Markman 1990]).
- l'agencement des détecteurs primitifs dans la couche  $\mathcal{C}$ . Cet agencement prévoit des incompatibilités indépendamment de toute expérience. Peut-on suggérer, à titre d'exemple, l'impossibilité d'avoir deux localisations distinctes et simultanées pour une entité spatio-temporelle ?
- la mémorisation des concepts. Dans notre modèle, nous n'avons pas décrit ce mécanisme de mémorisation. Nous avons simplement indiqué qu'il n'était pas du type associatif habituel, mais qu'il utilisait l'information issue de la détection d'une incohérence donnée par la couche  $\mathcal{C}$ .
- le fonctionnement dynamique du câblage associatif de la couche  $\mathcal{C}$ , qui permet de déclencher des associations et des analogies.

D'un autre côté, le modèle est aussi incomplet parce qu'il est très loin de rendre compte de l'ensemble des processus conceptuels. Nous n'avons fait que mentionner le fait que les concepts pouvaient évoquer des perceptions, provoquer l'activation de schémas d'action intégrés, et qu'ils avaient des liens très étroits avec le système linguistique. De même, les traitements conceptuels que nous avons observés dans les conversations sur les modes improbable et (in)désirable (cf. chap. I.1) ne sont pas couverts par ce modèle neuronal.

Ces points ne sont pas rédhibitoires compte tenu de notre objectif qui est de suggérer un mécanisme plausible pour la détection d'incohérences. Toutefois, nous pensons qu'il serait particulièrement intéressant de proposer des implémentations de ce modèle, qui permettrait de prouver sa fécondité, son intérêt pour la modélisation cognitive et enfin, peut-être, son intérêt technique pour l'apprentissage automatique.

## 7. Conclusion : "ramener la logique au neurone"

Dans le contexte scientifique actuel, la séparation symbolique/sub-symbolique est tellement affirmée que la réalité cognitive des processus symboliques est parfois niée [Smolensky 1988], principalement parce que l'on n'est pas en mesure de lui trouver une implantation neuronale plausible. L'objectif de ce chapitre était justement de suggérer une implantation plausible d'un mécanisme symbolique, le mécanisme logique de détection d'incohérence.

Notre modèle fait intervenir un appareil neuronal spécifique, la couche "C", pour le traitement de l'incohérence et des relations logiques. Il est en accord avec les exigences de la systématique, de la constituence et de la compositionnalité. Il est de plus plausible sur le plan neuronal, car les structures qu'il fait intervenir (détecteur collectif) résultent des propriétés de base des neurones et présentent un coût architectural faible.

Notre modèle est bien sûr incomplet. Mais, rappelons-le, notre objectif était avant tout de démystifier le caractère symbolique de certains traitements réalisés par les individus, en montrant qu'une architecture neuronale relativement simple pouvait expliquer des performances comme la détection d'incohérences pour lesquelles on est obligé de postuler un traitement de type logique. D'une certaine façon, nous avons voulu montrer qu'il n'était pas déraisonnable de vouloir "ramener la logique au neurone", de relier les propriétés d'un traitement symbolique particulier, la détection d'incohérences, aux propriétés que l'on reconnaît généralement aux cellules du cerveau.

**remerciements** : je remercie Olivier Hudry et Jean-Bernard Auriol pour les discussions qui m'ont permis de progresser dans la définition de ce modèle neuronal de la détection d'incohérences. Je remercie Cédric Chappelier et Bernard Victorri pour leurs critiques détaillées de ce texte.

## **III- Enseignement des concepts par ordinateur**

<b>III.1</b>	<b>Les approches de l'EIAO</b>	<b>185</b>
<b>III.2</b>	<b>SAVANT3</b>	<b>199</b>
<b>III.3</b>	<b>Un système auteur conversationnel</b>	<b>217</b>
<b>III.4</b>	<b>Explication et critique argumentatives dans les systèmes à base de connaissances</b>	<b>225</b>



## Introduction à la partie III

### Expliquer par l'incohérence

#### *L'utilisation d'un mode de la conversation spontanée en E.A.O. et dans les systèmes à base de connaissances*

Notre sensibilité à l'*incohérence*, si manifeste dans les conversations, se retrouve dans l'architecture de notre cognition, voire de notre cerveau. Cette constatation a des conséquences fondamentales.

Dans cette partie, nous explorons ses conséquences dans le domaine de l'apprentissage des connaissances et de la compréhension des explications. Ce domaine est essentiel, car il présage d'une utilisation différente des systèmes à base de connaissances : ces systèmes pourront enseigner leurs connaissances, les expliquer en situation, et les utiliser pour critiquer les actions de leur utilisateur.

Notre apport à l'amélioration de tels systèmes réside entièrement dans cette idée concernant la sensibilité humaine à l'incohérence. La thèse qui est défendue ici est qu'une connaissance est comprise, puis mémorisée, dès qu'elle est reliée logiquement à un contexte *problématique*, en particulier un contexte perçu comme *incohérent*.

- *Le chapitre III.1 est consacré à la distinction, fondamentale pour nous, entre concepts et savoir-faire. Avec des exemples de systèmes conçus pour enseigner, nous montrons que les savoir-faire s'acquièrent par la pratique, alors que les connaissances conceptuelles s'acquièrent par la logique, grâce à des contextes problématiques.*
- *Dans le chapitre III.2, nous décrivons notre système d'enseignement assisté SAVANT3. Ce système fonctionne avec comme seul principe celui d'essayer d'amener l'étudiant à s'auto-contredire. C'est dans cette situation d'incohérence momentanée que l'étudiant est supposé comprendre et apprendre le plus efficacement.*
- *SAVANT3 présente l'avantage d'être simple, et d'exiger peu de connaissances pour fonctionner. Dans le chapitre III.3, nous décrivons le système auteur que nous avons développé pour créer ces connaissances, notamment le programme ATOL qui repose sur les mêmes principes conversationnels que SAVANT3.*

- *L'enseignement n'est bien sûr pas la seule application des principes conversationnels que nous avons découverts. L'ordinateur du XXI<sup>ème</sup> siècle sera utilisé pour argumenter à propos des actions qui lui sont demandées ou pour critiquer les actions de son utilisateur. Dans le chapitre III.4, nous montrons comment ces interactions peuvent bénéficier de la reconnaissance préalable, par les deux partenaires, d'une problématique identique.*

# **Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage**

## **1. Introduction**

Notre objectif est d'appliquer à l'enseignement des connaissances les principes que nous avons dégagés à partir de l'observation des conversations et d'une réflexion sur l'apprentissage des connaissances conceptuelles. Nous allons essayer de montrer comment ces principes s'intègrent à la problématique de l'Enseignement Assisté par Ordinateur (EAO). Pour cela, nous n'allons pas dresser une liste des nombreux systèmes qui ont été décrits, mais plutôt évoquer les principaux courants qui vont nous permettre de positionner notre propre approche. Notre propos étant l'enseignement des concepts, nous allons subordonner notre analyse des travaux du domaine de l'EAO à la distinction, essentielle pour nous, entre apprentissage de concepts et apprentissage procédural.

Nous essayons de distinguer les différents systèmes selon le paradigme cognitif qu'ils mettent en oeuvre, indépendamment de leurs aspects novateurs sur le plan technique. En effet, nous restons convaincus que le progrès dans ce domaine ne viendra pas d'une simple augmentation de la complexité technique, mais de la définition d'une théorie cognitive convenable de l'apprentissage des connaissances.

## **2. La problématique de l'EAO**

L'enseignement est l'un des domaines essentiels de l'organisation d'une société. C'est pourtant l'un des domaines les moins touchés par les avancées de la technique. On apprend encore comme l'on apprenait il y a un siècle. Face à ce constat, nous avons vu deux attitudes. Celle qui consiste à sous-estimer le problème et inversement celle qui voit là une tâche insoluble par principe.

Après les premiers essais de l'Enseignement Programmé, nous avons vu à la fin des années 70 de très grandes entreprises informatiques investir dans des systèmes et des productions d'E.A.O. (Enseignement Assisté par Ordinateur). Ces investissements sont retombés au milieu des années 80, principalement parce que l'on s'est très vite heurté aux limites de l'approche classique : les produits coûtaient très cher, et étaient mal tolérés tant par les apprenants que par les formateurs, notamment en formation initiale.

Au même moment, la recherche a produit de nouveaux systèmes, comme les tuteurs intelligents, dont les objectifs sont beaucoup plus ambitieux que ceux de l'approche classique, puisqu'ils visent à modéliser l'élève pour adapter l'interaction. Mais la plupart de ces produits restent au stade de prototype, et leur transposition d'un domaine enseigné à un autre reste presque aussi coûteuse que le développement initial.

Face à ces grands espoirs plus ou moins déçus, beaucoup en sont venus à considérer que les ambitions devaient être diminuées. Certains considèrent même que le formateur humain est indispensable, et que la machine ne peut le remplacer pour des raisons fondamentales liées à la relation psychologique. Nous ne

connaissions toutefois pas de travaux qui aient pu établir ce point sans ambiguïté <sup>(1)</sup>. D'autres en sont venus à privilégier l'approche "ordinateur boîte à outils", à l'exclusion de toute tentative d'apporter des connaissances à l'apprenant.

Notre travail en EAO depuis 1980 nous a donné la conviction que la capacité d'enseigner des connaissances pouvait être rendue accessible à des machines. Cependant, le développement de produits efficaces et bien acceptés par les apprenants (idéalement des produits qui seront préférés aux formateurs humains pour certains types d'apprentissage) passe par une meilleure compréhension du processus d'apprentissage humain. Notre approche procède de l'heuristique suivante : chercher à reproduire entre l'apprenant et la machine une relation naturelle. Dans notre optique, la relation naturelle de référence n'est pas la relation enseignant-enseigné (qui est une invention sociale récente), mais un comportement biologiquement ancré : celui de la conversation.

### 3. Les approches "classiques" de l'EAO

Pendant de nombreuses années, deux approches de l'Enseignement Assisté par ordinateur se sont opposées : l'approche *tutorielle* et l'enseignement par simulation. La première est issue de l'Enseignement Programmé, la seconde a été illustrée par des produits comme LOGO [Papert 1980]. A ces deux approches classiques s'est ajoutée l'approche hypertexte, à laquelle nous avons contribué en développant notre système SAVANT1.

#### 3.1. L'EAO tutoriel et le behaviorisme

Pendant longtemps, le paradigme dominant en EAO de type tutoriel était issu du cadre behavioriste et de ses prolongements comme la pédagogie "par objectifs". Il est difficile de parler d'E.A.O. tutoriel sans avoir à l'esprit les méthodes d'enseignement prônées par B.F. Skinner, qui est l'un des principaux inspirateurs de l'Ecole de pensée behavioriste fondée avant 1920 par John Broadus Watson.

Le Behaviorisme se préoccupait au départ de psychologie animale. Son objectif était de rendre compte du comportement animal et humain, notamment de la faculté d'apprentissage, en rejetant tout "mentalisme". Le Behaviorisme va adopter en fait une position extrême : toute hypothèse sur le fonctionnement interne de l'organisme étudié est rejetée *a priori*. On ne prend en compte que ce qui est "visible" objectivement, c'est-à-dire les stimuli (les entrées du système) et les réponses (les sorties du système). Par exemple, un behavioriste qui considère l'activité langagière du petit enfant essaiera d'évaluer le plus précisément possible l'ensemble des mots et des phrases auxquels l'enfant a pu être exposé d'une part, et d'autre part les productions de l'enfant étudié (phonèmes, mots, constructions syntaxiques). C'est à partir de ces éléments qu'il chercherait à expliquer l'apparition du langage, en utilisant un seul paramètre : le "plaisir" éprouvé par le jeune enfant qui voit ses productions vocales suivies d'un effet, et le "déplaisir" qu'il est censé éprouver lorsqu'on ne le comprend pas ou lorsqu'on le corrige (la mémorisation n'intervenant que dans le premier cas) [Naville 1942 p. 248].

La thèse centrale du behaviorisme est que tout apprentissage, qu'il soit le fait d'un invertébré ou d'un mammifère supérieur, procède d'un *conditionnement*. Le conditionnement "opérant" défini par Skinner est un mécanisme de sélection comportementale. L'organisme produit un comportement (au départ spontané et aléatoire). Ce comportement a des conséquences qui se révèlent en retour positives (plaisir) ou négatives (déplaisir, douleur):

"Dès que nous avons établi le type particulier de conséquence que nous appelons un renforcement, nous sommes en mesure de modeler le comportement d'un organisme pour ainsi dire à notre gré. [...] Si nous prenons soin de présenter de la nourriture à un pigeon affamé à certains moments précis, [...] nous pouvons lui apprendre à tourner en rond, à décrire des huit sur le sol de sa cage, à se tenir immobile dans un coin, à tendre le cou ou à frapper du pied." [Skinner 1969 p.16]

---

<sup>1</sup> Certains travaux utilisant la technique dite du magicien d'Oz tendent au contraire à montrer que la machine présente elle aussi des attrait relationnels [Nicolle 92].

On comprend que la récompense puisse augmenter la fréquence de production de comportements élémentaires, après qu'ils se sont produits, au départ, spontanément. Mais qu'en est-il de l'apprentissage des comportements complexes souhaités par le dresseur, puisqu'il est tout à fait improbable qu'ils soient effectués spontanément pour être ensuite récompensés. Pour expliquer leur acquisition, les behavioristes montrent comment tout comportement apparemment complexe peut être en fait décomposé en une succession de comportements élémentaires accessibles à l'apprentissage par conditionnement : le comportement global sera appris par la récompense de ses composants.

L'E.A.O. de type tutoriel était, dans ses débuts, l'exacte mise en oeuvre des principes behavioristes dans le domaine de l'enseignement, d'autant que Skinner avait lui-même tracé la voie dans ce qui s'appelait l'enseignement programmé [Skinner 1969]. L'élève est toujours mis dans une situation pour laquelle on est certain qu'il produira un *comportement observable*. C'est ce comportement qui est jugé, puis le jugement est retourné à l'élève aussi souvent que possible sous la forme d'une félicitation ou au contraire d'une invitation à recommencer. On retrouve le même souci d'utiliser l'observation de comportements visibles pour prouver une acquisition dans la Pédagogie par Objectifs [Mager 1977].

L'approche behavioriste a été beaucoup critiquée d'un point de vue scientifique. L'interdiction de formuler des hypothèses sur ce qui n'est pas directement observable va à l'encontre des principes mêmes de la démarche expérimentale. Par ailleurs, en ce qui concerne l'apprentissage animal, les éthologues ont montré que de nombreux comportements ne devaient rien à un système de récompense / punition [Lorenz 1974; Eibl-Eibesfeldt 1967]. Enfin, l'application du Behaviorisme à l'apprentissage des connaissances nous semble présenter deux obstacles majeurs. Le premier vient des limitations dues à son caractère statistique (voir chap. II.1). Le deuxième vient de ce que, dans cette théorie, l'existence même des connaissances est niée !

Dans le chapitre II.2, nous avons montré qu'il n'était pas possible d'expliquer des comportements humains comme celui des interlocuteurs sans postuler l'existence objective de *connaissances* dans l'esprit du locuteur. Nous aurons ici l'occasion d'insister sur une distinction, qui nous semble essentielle, entre *connaissances conceptuelles* et *connaissances procédurales*. Mais pour les Behavioristes, le "concept" de concept n'est pas pris en compte, puisqu'ils ne font aucune hypothèse sur l'état interne de l'organisme étudié. Ils tentent donc à chaque fois de remplacer les connaissances conceptuelles par un *comportement observable*.

Par exemple, imaginons que l'on veuille enseigner les concepts d'électron et de fusible grâce à un tutoriel. Si on s'en tient au dogme, on tentera par des renforcements appropriés de stabiliser le comportement souhaité chez l'élève. Mais quel est le comportement, ici ? Si quelqu'un qui veut savoir l'heure entend qu'il est cinq heures, on peut penser qu'il a compris. Le Behavioriste, lui, exigera un comportement de sa part (par exemple qu'il se précipite vers la gare). Quel comportement peut-il bien attendre dans le cas de l'électron et du fusible ? En fait, il attendra que l'élève donne les réponses attendues à un certain nombre de questions du genre "quelle est la valeur de la plus petite charge électrique que l'on peut isoler ?". Mais un certain nombre de critiques sautent aux yeux :

- L'élève a des chances d'avoir "compris" (au sens que nous avons donné à ce terme) *avant* de répondre aux questions, ces dernières servant alors de preuve de l'acquisition. On lie donc deux choses qui sont essentiellement différentes, au point de finir par les confondre : la transmission du concept, et l'évaluation de cette transmission. Or répondre sans arrêt à des questions non motivées se révèle vite fastidieux pour l'apprenant<sup>(2)</sup>.
- Les questions ne sont pas censées contribuer à la transmission du concept. Elles ne sont là que comme révélateur. Les renforcements (félicitations, réprimandes, invitations à recommencer) ne véhiculent pas d'information relative au concept. Alors où l'élève trouve-t-il cette information ? Parfois, l'élève est obligé d'effectuer une tâche bien plus complexe et insoupçonnée de l'auteur pour deviner ce que ce dernier attend de lui avec ses questions. Mais le plus souvent, l'auteur évite le

---

<sup>2</sup> Dans notre système SAVANT3, qui sera présenté dans le chapitre III.2, on ne cherche pas à vérifier directement l'acquisition des notions, et toutes les questions sont motivées.

problème en donnant une définition préalable du concept (ce qui n'est pas prévu dans la théorie behavioriste de l'apprentissage animal où l'on part de comportements spontanés). Il est donc normal de se demander si l'effort pédagogique ne devrait pas se porter sur cette définition du concept plutôt que sur l'évaluation.

- Pour enseigner un comportement, on peut parfois le fractionner, comme cela est prescrit dans la théorie skinnérienne. Mais que signifie fractionner un concept comme celui d'électron? Cela consiste-t-il à parler de particule élémentaire? Et dans quelle mesure ce nouveau concept sera-t-il compris plus facilement? Comment fractionner le concept de fusible?

L'approche tutorielle de l'enseignement assisté par ordinateur a subi certains aménagements, rendus nécessaires par le souci de rendre l'interaction acceptable. Mais aucun paradigme nouveau n'est venu remplacer l'ancienne vision behavioriste selon laquelle l'essentiel de l'acte d'enseignement réside dans la rétroaction, le jugement de valeur, que l'on renvoie à l'apprenant à propos d'une *tâche* qu'il vient d'effectuer.

### 3.2. *L'enseignement par environnement simulé*

Depuis longtemps, mais surtout, dans le contexte de l'E.A.O., avec le développement du système LOGO par S. Papert [Papert 1980], on s'est rendu compte du fait que les individus pouvaient apprendre par eux-mêmes. L'enfant qui interagit avec la tortue LOGO construit ses théories et peut les tester [Grumbach 1990]. Cette idée de construction provient des thèses de Piaget, qui ont d'ailleurs fortement inspiré la démarche de Papert. Les simulations se prolongent dans les environnements réactifs et les Micro-mondes [Thompson 1987].

Contrairement à l'approche tutorielle classique, les environnements simulés sont bien acceptés par les apprenants, principalement parce qu'ils sont perçus comme une activité ludique. Peut-on imaginer de les utiliser pour l'enseignement des concepts ?

En fait, ces environnements permettent avant tout l'acquisition de *procédures*. Certes, dans le cas de simulations complexes, par ex. STEAMER [Hollan et al. 1987], l'utilisation de certaines procédures évoluées nécessite des connaissances conceptuelles spécifiques. Ceci a été parfaitement mis en évidence par J-F Richard dans le cas de l'apprentissage d'un traitement de textes [Poitrenaud et al. 1990]. L'acquisition de ces connaissances conceptuelles motivées par l'accomplissement d'une procédure est soit totalement spontanée et laissée au hasard (cas de LOGO), soit obtenue grâce à des aides spécifiques (systèmes d'aide contextuels) dont le principe diffère d'une application à l'autre. Autrement dit, ce ne sont pas les systèmes de simulation, les micro-mondes, etc. qui peuvent, en tant que tels, fournir à eux seuls une réponse suffisante au problème de l'apprentissage des connaissances conceptuelles.

### 3.3. *Les hypertextes et les hypermédias*

En réaction contre la démarche extrêmement directive de l'EAO tutoriel, nous cherchions en 1980 un système dans lequel l'élève garde un contrôle total de la machine. Les élèves visés étant habitués à travailler seuls, nous avons imaginé de leur offrir un accès aussi rapide que possible aux définitions des concepts contenus dans leur cours, et aux raisonnements qui lient ces concepts, autrement dit de leur fournir une encyclopédie conceptuelle interactive, qui fut baptisée SAVANT1<sup>(3)</sup>.

Notre scénario était le suivant: l'élève est en train de lire un document papier. Il rencontre un problème de compréhension (de telles difficultés peuvent bloquer l'élève pendant près d'une heure !). Il est capable de localiser son problème de compréhension sur une notion ou un résultat prérequis qu'il ne possède pas parfaitement, pense à se connecter sur l'encyclopédie, et parvient très rapidement au concept ou au résultat cherché. Ce dernier est exposé *en un seul écran* très concis, et cet écran est *autonome*: les concepts prérequis sont explicitement signalés (vidéo inverse), et les notations employées sont toutes explicitées. L'élève prend

<sup>3</sup> S.A.V.A.N.T.: Système d'Assistance Visuelle à l'Apprentissage de Notions Théoriques. Deux autres systèmes portent ce nom : un serveur multimédia (SAVANT2) et un système d'enseignement par l'argumentation (SAVANT3).

connaissance de la définition, consulte des prérequis, consulte la démonstration s'il s'agit d'un résultat, puis retourne satisfait à son document écrit.

**Rapport de vraisemblance**

Par définition:

$$V(z) = \frac{p(z|H1)}{p(z|H2)}$$

Dans le cas de deux hypothèse,  $V(z)$  mesure la plus grande vraisemblance de l'Émissi du 1er symbole par rapport à celle du 2ème, après l'observation  $z$ .

---

$V(z)$ : rapport de vraisemblance de H1 par rapport à H2  
 $H1$ : hypothèse "observation" du premier symbole"  
 $z$ : Élément de l'espace d'observation supposé de dimension  
 $p(z|H1)$ : densité de probabilité conditionnelle

figure 8: exemple d'écran de l'encyclopédie SAVANT1

Nous avons ainsi constitué au cours des années une encyclopédie de plus de 3000 écrans, accessible par plus de 850 mots clés. Pour des raisons ergonomiques et économiques, nous l'avons réalisée en vidéotex, ce qui nous permettait de confier un terminal Minitel à chaque élève pour qu'il puisse utiliser le système depuis sa résidence<sup>(4)</sup>.

Cette exigence de l'écran unique autonome, adoptée dans SAVANT1, s'impose aussi pour la présentation des raisonnements. Ceux-ci sont présentés selon différents niveaux de détail, chaque niveau comportant des fenêtres donnant accès à un niveau plus fin. Mais là encore, un niveau du raisonnement doit tenir dans l'espace d'un écran.

SAVANT1 diffère quelque peu de ce que l'on a appelé ensuite un hypertexte, parce que l'information est complètement retraitée pour figurer dans l'encyclopédie, et selon des règles précises. De plus, l'encyclopédie n'est pas "feuilletable". Nous avons en effet été contraints, à la suite des premières expérimentations, de séparer totalement l'information technique, celle de l'encyclopédie, accessible seulement par mot clé, d'une information "narrative" [Dessalles 1984a, 1984b]. C'est cette dernière que nous avons développée pour la délinéariser (selon des techniques navigationnelles originales) et la rendre argumentative [Dessalles 1992g].

SAVANT1 a été utilisé pendant plusieurs années par les élèves de TELECOM-Paris. Nous avons voulu, dès 1984, prolonger l'expérience avec le développement d'un serveur multimédia, SAVANT2 [Dessalles 1986]. Ce système, original sur le plan technique (serveur de vidéodisques, réseau en fibres optiques, commutations vidéo automatiques) a été exploité jusqu'en 1989 à TELECOM-Paris<sup>(5)</sup>. La vidéo interactive était utilisée pour amplifier l'effet de la partie narrative. Ayant reconnu l'importance de l'argumentation dans l'apprentissage, nous avons utilisé SAVANT2 pour réaliser une première forme de conversation interactive. L'utilisateur assiste à une conversation entre trois personnages, et après chaque réplique, choisit le prochain locuteur. Les

<sup>4</sup> Le serveur SAVANT1 était en 1981 l'un des tout premiers serveurs vidéotex, en particulier dans le secteur de l'enseignement. Son développement a été voulu par Daniel Cadé, et il a été développé à l'ENST par Jean-Pierre Bach, du Centre de Calcul.

<sup>5</sup> SAVANT2 constituait une innovation sur le plan technique. Il s'agissait de l'un des premiers serveurs multimédia en France. Son développement a été décidé dès 1982 par Daniel Cadé, à l'époque Directeur adjoint de l'ENST. Il est le fruit d'une collaboration entre le Centre de calcul de l'ENST (Jean-Pierre Bach), le labo vidéo de l'ENST (Alain Biaggi), et le service NTE (Nouvelles Techniques pour l'Enseignement) dont j'assurais la responsabilité. Le projet dans son ensemble (mise en place du réseau fibres optiques, système de multiplexage vidéo/données, pilotage des vidéotelecteurs et des commutations, réalisation des vidéodisques, exploitation) a bénéficié de la collaboration de nombreuses personnes de TELECOM-Paris.

différents enchaînements possibles étaient conçus à l'avance. Suivant ses choix, l'utilisateur pouvait assister à des argumentations différentes. Mais au total, il avait profité de la conversation et des concepts qui y auront été évoqués. Malheureusement, la conception du graphe de la conversation était malaisée. Il s'agit a priori d'un arbre, qu'il fallait "mailler" pour des raisons d'économie, tout en conservant la pertinence de chaque réplique.

Il existe maintenant de nombreuses utilisations des hypertextes et des hypermédias pour l'apprentissage. Toutefois, comme pour les simulations, leur utilisation pour l'enseignement des connaissances conceptuelles n'entre pas dans un cadre théorique clair. Avec SAVANT1 puis SAVANT2, nous voulions donner un accès *optimal* à l'information conceptuelle, en contrôlant les effets de contexte : suppression du contexte dans l'encyclopédie SAVANT1 (par le fait de rendre les écrans autonomes), construction d'un contexte argumentatif dans la partie narrative des deux systèmes. C'est grâce à ces expériences que nous avons compris l'importance du contexte argumentatif, ce qui nous a conduit au développement du système SAVANT3 que nous décrivons plus loin (chapitre III.2).

## 4. Connaissances conceptuelles et connaissances procédurales

Au cours de notre étude des conduites argumentatives spontanées, nous avons acquis la conviction que, contrairement à ce que l'approche tutorielle classique tenait pour admis, les concepts et les savoir-faire ne pouvaient pas être enseignés de la même façon et que l'argumentation devait jouer un rôle essentiel dans la transmission des concepts. Nous commencerons par clarifier cette séparation entre les connaissances procédurales et les connaissances conceptuelles, car l'essentiel de notre démarche repose sur elle.

### 4.1. Une distinction trop souvent négligée

Notre objectif étant d'enseigner des concepts en utilisant un système capable d'argumenter, nous allons nous intéresser principalement aux systèmes capables d'aider à l'acquisition de connaissances conceptuelles, par opposition aux connaissances de nature procédurale. Cette distinction, niée dans la conception behavioriste, est pour nous fondamentale. L'analyse des conversations nous a amplement prouvé que le traitement et l'apprentissage des connaissances conceptuelles, en l'occurrence à travers le langage, était très différent de ce que proposent les théories d'inspiration behavioristes qui donnent le primat à l'action. Nous retrouvons notre point de vue chez de nombreux auteurs qui, sous des dénominations diverses, font bien cette distinction essentielle entre concepts et savoir-faire :

"There are two basic forms of learning : *knowledge acquisition* and *skill refinement*. When we say that someone learned physics, we mean that this person acquired significant concepts of physics, understood their meaning, and understood their relationships to each other and to the physical world. The essence of learning in this case is the acquisition of new knowledge, including descriptions and models of physical systems and their behaviors, incorporating a variety of representations. [...] A second kind of learning is the gradual improvement of motor and cognitive skills through practice, such as learning to ride a bicycle or to play the piano. [...] The bulk of [this] learning process consists of refining the learned skills, whether mental or motor coordination, by repeated practice and by correcting deviations from desired behavior." [Carbonell, Michalski & Mitchell 1983]

"On peut par exemple distinguer les apprentissages selon la nature de ce qui est acquis au terme du processus. On mettra d'un côté les dispositions 'propositionnelles', relevant du 'savoir que' (knowing that), de l'autre les dispositions "pratiques", relevant du 'savoir-faire' (knowing how). On aura donc d'une part des capacités purement inférentielles, de l'autre des aptitudes (skills)." [Anderl 1987]

On voit se dessiner l'idée que les connaissances conceptuelles sont par essence liées à la logique et que leur acquisition est de nature *qualitative*, alors que les procédures s'acquièrent de manière graduelle. Par exemple Inhelder et Piaget opposent les procédures aux structures, ces dernières étant caractérisées par leur insertion dans des inférences :

"C'est au sein des inférences du sujet que se manifestent les structures tandis que les procédures comportent beaucoup plus d'empirisme; et, en ces inférences, ce sont les éliminations des contradictions et des incompatibilités qui sont les plus révélatrices de leur nature structurale." [Inhelder & Piaget 1979]

Ces auteurs non seulement accordent un statut logique aux "structures", mais de plus adoptent le point de vue qui est le nôtre, selon lequel la nature conceptuelle d'une connaissance se révèle dans sa capacité à entrer dans une relation d'incompatibilité (voir chapitre II.3). Nous sommes donc très près de l'idée d'utiliser l'*argumentation* comme outil privilégié pour enseigner des connaissances conceptuelles, comme le suggère S. Ohlsson :

"We are moving from a concern with mainly procedural knowledge to concerns with conceptual, declarative, theoretical, abstract knowledge" [...] "There is something different between explaining something or arguing or justifying and solving the Tower of Hanoi or solving a physics problem or whatever. I think that those activities - explaining, arguing, justifying - are closely connected with the notions of conceptual knowledge, belief systems and world view." [Ohlsson 1991]

Cette distinction fondamentale entre concepts et procédures, même si elle commence à être reconnue par un nombre croissant de chercheurs, semble encore être négligée par la plupart des auteurs, tant elle est contraire au schéma longtemps dominant de l'apprentissage par l'action. Par exemple, elle n'apparaît pas du tout clairement dans la théorie psychologique d'Anderson [1983] qui vise pourtant à décrire l'architecture générale de la cognition<sup>(6)</sup> ! La raison de cette négligence vient selon nous du fait que la recherche en psychologie de l'apprentissage s'est surtout attachée à décrire l'apprentissage de *tâches*. Or une grande partie de notre apprentissage est indépendant de toute tâche. De plus, si certaines tâches sont purement procédurales (par ex. faire du vélo), il n'en est pas de mêmes de certaines tâches intellectuelles, dans lesquelles la connaissance des concepts intervenant dans les procédures est essentielle. Nous nous proposons de commenter ces aspects des relations concepts / procédures, car ils sont essentiels pour la suite.

## 4.2. *Concepts et procédures : deux apprentissages radicalement différents*

Il est important de comprendre que l'acquisition d'un concept est d'une nature qualitativement différente de celle d'une acquisition de savoir-faire. Prenons deux exemples: la notion de protéine et le calcul d'une intégrale. Le bon sens dicte que le premier est un concept et le second un savoir-faire. En EAO tutoriel classique, enseigner la notion de protéine n'a pas de sens. En revanche, on peut par exemple mettre l'élève en situation de discriminer parmi différentes formules celles qui décrivent des protéines, de même qu'on le mettra en situation de calculer une intégrale donnée. Pourtant, deux différences fondamentales séparent ces deux exemples d'apprentissage.

**Performance graduelle vs. qualitative** : la première différence concerne la performance de l'élève. Un élève qui fait des changements de variables inutiles ou qui intègre par parties de manière maladroite peut apprendre comment éviter les essais peu prometteurs. Il peut donc progresser. Cela a donc un sens d'évaluer sa performance, comme on le fait systématiquement en EAO tutoriel. En revanche, on ne peut en aucun cas affirmer que l'élève qui fait moins d'erreurs en discriminant des formules chimiques de protéines a progressé dans la compréhension de ce qu'est une protéine. Une protéine, c'est la mise bout à bout de différents acides aminés. L'élève qui se trompe identifie peut-être mal les acides aminés, ce qui est une autre histoire : cela ne veut pas dire qu'il n'a pas compris ce qu'est une protéine. Ou alors il n'a effectivement rien compris à ce qu'est une protéine. Mais en aucun cas il n'a compris qu' "à moitié". En résumé, mesurer quantitativement l'acquisition d'un concept n'a pas de sens, et pour cette raison la démarche théorique de l'EAO tutoriel classique devient caduque dans le cas des concepts.

**"Insécabilité" des concepts** : la deuxième différence concerne la "sécabilité" des concepts. Dans la théorie skinnérienne qui sert de base à l'EAO tutoriel classique, un comportement trop complexe pour être

---

<sup>6</sup> Anderson, certes, consacre beaucoup d'efforts à décrire l'apprentissage qu'il appelle procédural. Les connaissances qu'il désigne comme procédurales correspondent tout à fait à ce que nous désignons sous ce terme (même si Anderson les représente avec des règles de production qui, lorsqu'elles sont explicites, sont justement le moyen informatique standard pour exprimer des connaissances déclaratives !). En revanche, ce qu'Anderson appelle "connaissances déclaratives" est très éloigné des règles de logique (par ex. règles d'incompatibilité) qui relient les concepts, et recouvre plutôt des éléments mémorisés de manière associative.

acquis en une fois peut être fractionné en unités plus simples juxtaposables. Pour parvenir à calculer des intégrales, l'élève peut commencer par intégrer des fonctions simples, puis appliquer chaque méthode (intégration par parties, changement de variables) séparément, puis reconnaître les cas favorables (fonctions polynomiales ou sinusoidales pour l'intégration p.p.), etc. Malheureusement, on voit mal comment découper le concept de protéine: le sous-concept d'acide aminé est bien plus complexe que le concept de protéine. Il en est de même de leur accrochage par la liaison C--O N--H.

Selon notre point de vue, l'acquisition d'un concept est un processus de nature logique, alors que l'acquisition d'un savoir-faire est de nature procédurale. Une procédure peut être mise au point grâce à l'évaluation d'une performance, c'est-à-dire d'une distance par rapport à un objectif idéal. Ce n'est pas le cas pour un concept. Nous allons au contraire essayer de montrer que la transmission d'un concept peut être modélisée par un processus de nature logique, ce qui explique que la performance, dans ce cas, soit nécessairement qualitative.

### 4.3. *Connaissances conceptuelles et résolution de problème*

De même que certains apprentissages de savoir-faire ne nécessitent aucune connaissance de nature conceptuelle <sup>(7)</sup>, la plupart des traitements conceptuels se font en l'absence de toute situation de résolution de problème. Notre étude des conversations nous a permis d'observer de nombreux apprentissages de connaissances en l'absence de toute tâche à résoudre. Stellan Ohlsson imagine une base de données capable de fournir des explications à un patient, concernant sa maladie, lorsqu'il se trouve dans la salle d'attente de son médecin :

"[...] the system will then look into its knowledge base and it will generate natural language explanations which are tailored to the particular patient. [...] So there will be no task here for the patient, the patient has no problem solving task, we are not out to teach him any cognitive skills. The goal is instead to improve his understanding of why he is having such and such symptom and why the doctor is prescribing such and such a therapy and possibly as the therapy goes on why such and such side-effects occur ; that is a first step towards a computer based training system that is going to operate on the user's belief system rather than upon his cognitive skills and his procedural knowledge." [Ohlsson 1991]

Cependant, l'apprentissage des procédures et celui des connaissances n'est pas à ce point cloisonné. Plusieurs auteurs ont montré que l'acquisition de certaines procédures nécessitait la formation d'une représentation conceptuelle. Par exemple Richard [1983], Moyse [1989] et Grumbach [1989] montrent que des descriptions purement fonctionnelles du type *but* → *actions*, qui traduisent généralement des savoir-faire procéduraux (cf. [Anderson 1983]), ne permettent que les actions prévues dans les règles. Toute adaptation à une situation nouvelle nécessite de la part des individus la mise en place d'un modèle structurel. Dans [Grumbach 1989], il s'agissait d'une notice décrivant le chauffage d'un véhicule. Dans une notice, on pouvait lire des descriptions purement fonctionnelles : "*Si vous désirez désembuer le pare-brise, alors tournez la manette M vers la gauche et positionnez l'interrupteur I en position I*", alors qu'une autre notice pour un véhicule similaire donnait seulement des indications structurelles : "*La manette M, tournée vers la gauche, ouvre le circuit d'arrivée d'eau chaude*".

Ces études tendent à montrer qu'une représentation structurelle logique des processus devient nécessaire dès que l'individu est en situation de devoir s'adapter, au point qu'il est parfois amené à construire cette représentation si elle ne lui a pas été donnée. Toutefois, les descriptions fonctionnelles, procédurales, semblent être utiles au novice, car la seule connaissance d'une description structurelle lui posera des problèmes :

"Ce qui est difficile pour le novice, c'est probablement de savoir ce que le dispositif permet de réaliser en fait d'actions qui soient à la fois réalisables à l'aide des commandes disponibles et envisageables comme des objectifs possibles, comme sous-buts qu'il puisse se donner dans une tâche complexe. Même s'il connaît l'effet des commandes, cette connaissance n'est pas utilisable immédiatement car en général l'effet

<sup>7</sup> Un jour quelqu'un expliquait "qu'il y avait deux façons de tourner en moto : soit on se penche sur le côté, soit on tourne le guidon". Son modèle conceptuel grossièrement erroné (il faut faire les deux actions conjointement) ne l'empêchait pas d'utiliser son véhicule de manière correcte ! Dans de tels cas, nous avons des savoir-faire "purs", qui tolèrent l'absence de modèle conceptuel, ou même la présence de modèles conceptuels erronés.

d'une commande ne correspond pas au résultat d'une action qui soit équivalente à un sous-but dans la programmation habituelle de l'action." [Richard 1983]

Malgré cette restriction, on constate que des connaissances de nature conceptuelle, par ex. une description structurelle relative à un dispositif, peut être utilisée pour accomplir une tâche, au même titre que des connaissances de savoir-faire. Le mécanisme est toutefois fort différent. Nous allons observer les conséquences de ces relations concept / savoir-faire lorsqu'il s'agit d'enseigner.

Cauzinille-Marmèche et Mathieu [1988] distinguent les systèmes d'EAO selon leur objectif : enseigner des compétences élémentaires, ou des stratégies de résolution, ou encore des concepts et les relations qui les lient. Nous rejoignons cette classification dans les descriptions qui vont suivre. Pour commencer, nous allons nous attacher à décrire les approches E.A.O. dont l'objectif est d'enseigner des connaissances purement conceptuelles, pour ensuite évoquer les approches qui prennent en compte les connaissances conceptuelles dans l'enseignement des savoir-faire.

## 5. Des systèmes qui enseignent des concepts

Nous évoquerons quelques systèmes, dont SCHOLAR et WHY, d'une part parce qu'ils sont sans doute parmi les premiers systèmes qui tentent d'enseigner grâce à la richesse de leur dialogue, et d'autre part parce que même s'ils sont souvent cités, ils ont été peu imités par la suite. Or leur objectif est le même que le nôtre : enseigner des concepts, et la méthode est très proche : utiliser le dialogue.

### 5.1. Le "Socratic System"

L'une des premières tentatives pour utiliser l'ordinateur afin de gérer un dialogue socratique dans un but didactique est due à Wallace Feurzeig, dont le système "socratique" est ici présenté par R. Bolt :

"The Socratic System and a student together 'discuss' a problem to be solved or a body of information to be learned. The system states a problem, sets the appropriate conditions upon its solution, asks questions, provides data requested, and replies to questions and assertions put forth by the student. The student, in turn, asks for clarification or for data he believes to be relevant, proposes tentative answers, and works his way through the best solution he can find, which he reports to the system." [Bolt 1968]

Voici un exemple d'interaction avec le Socratic System donné par R. Bolt :

[. . .]  
 system: *The overtime amounts to \$60,000 for seasonal production and none for level. These new data should change your pro formas and have very significant effects on your decision...*  
 student: *Seasonal production*  
 system: *Now that you have made a decision, let's consider it in the light of the relevant facts... Consider these corrected pro formas carefully. Then tell me - are you still happy with your current decision ?*  
 student: *Yes*  
 system: *Tell me again, just to be sure that I understand, WHICH production schedule do you recommend ?*  
 student: *Seasonal production.*  
 system: *Let's consider a hypothetical question. If you were guaranteed that you would not be caught with a lot of unsaleable inventory, would you change your decision to one in favor of level ?*  
 student: *No.*  
 system: *Then let's consider whether or not we are agreed on amount of savings that would result. Assuming all the inventory were sold, how many thousands of dollars in savings would result from going to a level production schedule ?*

*student: 110.*  
*system: That's about right. The actual value, determined from direct labor costs, overtime costs, and machine maintenance costs, is \$108,000. That's a lot of money. Doesn't this change your decision ?*  
*student: No.*

Ce que nous retenons du dialogue socratique, c'est principalement son aspect argumentatif. Dans cet extrait, le système ne juge pas l'action de l'étudiant (seasonal production) comme mauvaise ou comme bonne. Il donne des arguments *contre* le choix de l'élève, et amène l'élève à se justifier ou à changer ses choix. C'est cet échange d'arguments que nous avons cherché à systématiser dans SAVANT3.

Dans le Socratic System, les arguments semblent préétablis. Le moment où ils sont délivrés dépend du dernier choix fait par l'étudiant, mais aussi de tout le passé de l'interaction (les informations que l'étudiant a obtenues, sa stratégie de résolution du problème, etc.). Les systèmes ultérieurs essaient dans une plus grande mesure de *calculer* leurs arguments.

## 5.2. SCHOLAR

J.R. Carbonell, avant 1970, réalisa ce qui est sans doute le premier système intelligent qui tente de soutenir un dialogue didactique avec l'apprenant. Son système, SCHOLAR, utilisait des connaissances stockées dans des réseaux sémantiques (qui venaient d'être imaginés par Quillian) et il possédait quelques capacités pour interagir en langue naturelle (mot clés créés dynamiquement à partir du réseau). Il gérait un dialogue à initiative partagée [Wenger 1987] :

*SCHOLAR : The eastern and central plain of Argentina is ---*  
*Student : Chaco*  
*SCHOLAR : No. The northern plain of Argentina is the Chaco. The eastern and central plain of Argentina is the Pampas.*  
*Student : Is the Chaco in Chile ?*  
*SCHOLAR : No. The Chaco is in Argentina, Paraguay, and Bolivia, not in Chile.*

SCHOLAR savait analyser les réponses de l'étudiant, il essayait de les évaluer avec son réseau sémantique, donnait une réponse élaborée, puis, dans le cas où l'élève n'essayait pas de prendre l'initiative, il recherchait un sujet associé pour poser une nouvelle question.

Ces recherches visant à mettre au point un système capable d'apprendre par le dialogue furent poursuivies au sein de la même équipe, et donnèrent naissance au système WHY.

## 5.3. WHY

Le système WHY avait pour objectif de reproduire une relation de type "socratique" entre l'élève et la machine. Il utilisait des connaissances stockées sous forme de scripts (chaînes temporelles et causales) hiérarchisés qui servaient à la fabrication des questions et à l'analyse des réponses [Stevens et al. 1979]. Ce système se caractérise par des stratégies de dialogue qui ont été formalisées à partir de l'observation de dialogues tutoriaux réels [Collins 1976] :

*WHY : Do you think the Amazon jungle has heavy rainfall or light rainfall ?*  
*Student : Heavy rainfall*  
*WHY : Why does the Amazon have heavy rainfall ?*  
*Student : Because it's near the Atlantic*  
*WHY : Yes, the Amazon jungle is near the equatorial Atlantic. How does that affect rainfall in the Amazon jungle ?*  
*Student : The water in the current is warm*  
*[. . .]*  
*WHY : Do you think that any place with mountains has heavy rainfall ?*

*Student : Yes*

*WHY : Southern California has mountains. Why doesn't Southern California have heavy rainfall ?*

*[...]*

Dans cet extrait (pris dans [Wenger 1987]), on observe que WHY ne se contente pas d'une réponse correcte d'un élève. Il demande la cause du fait affirmé par l'étudiant (d'où sans doute le nom du système). Lorsque la cause lui paraît insuffisante, WHY essaie de raisonner par l'absurde et laisse ainsi entendre que des contre-exemples existent. C'est ce que les auteurs appellent la stratégie de piège (*entrapment strategy*).

A. Collins [1976] analyse la méthode socratique comme composée de plusieurs stratégies qu'il essaie de reproduire dans WHY :

[1] Ask about a known case	[10] Pose a misleading question	[18] Question a prediction made without enough information
[2] Ask for any factors	[11] Specify how the variable depends on a given factor	[19] Point irrelevant factors
[3] Ask for intermediate factors	[12] Probe for a necessary factor	[20] Point out an inconsistent prediction
[4] Ask for prior factors	[13] Probe for a sufficient factor	[21] Ask for consideration of a possible value
[5] Form a general rule for an insufficient factor	[14] Probe for differences between two cases	[22] Test for consistency with a given hypothesis
[6] Pick a counterexample for an insufficient factor	[15] Request a test of the hypothesis about a factor	[23] Ask for consideration of relevant factors
[7] Form a general rule for an unnecessary factor	[16] Ask for a prediction about an unknown case	[24] Trace the consequences of a general rule
[8] Pick a counterexample for an unnecessary factor	[17] Ask what are the relevant factors to consider	
[9] Pick a case with an extreme value		

Notre propre démarche, avec SAVANT3, se retrouve proche de cette stratégie particulière, dite stratégie de piège, qu'elle systématiser. Les 24 stratégies analysées par Collins nous semblent participer de cette stratégie de piège (str. [6], [8], [20]), ou en être une conséquence.

Dans SAVANT3 comme dans SCHOLAR et WHY, l'objectif est d'enseigner des connaissances conceptuelles, et le moyen employé est le dialogue. En fait, nos présupposés théoriques rejoignent en partie ceux de ces auteurs lorsqu'ils analysent les erreurs commises par les étudiants dans ce genre de dialogue comme étant dues à des conceptions erronées (*misconceptions*) :

"Our analyses of dialogues show that tutors spend a good part of their time diagnosing conceptual bugs from errors manifested in the dialogue." [Stevens et al. 1979]

Le paradigme du dialogue socratique repose ainsi sur le fait que ces conceptions erronées peuvent se manifester dans le dialogue, et que le dialogue permet de les corriger :

"When the student makes an error, the [human] tutor will sometimes correct it immediately, but in many cases will ask other questions until the misconception underlying the error is isolated." [Stevens et al. 1979]

Comme eux, nous estimons que le dialogue socratique, parce qu'il est argumenté, est le moyen par excellence grâce auquel de telles conceptions erronées peuvent être mises en évidence et corrigées. Nous avons cherché à systématiser ce point de vue dans notre approche. Simplement, notre système SAVANT3 va effectuer le calcul de ses arguments, et des stratégies comme celles qui sont listées par Collins émergeront de ce calcul, au lieu d'être imposées en tant que telles au système.

Afin de mieux situer cette préoccupation argumentative pour l'enseignement des concepts, nous allons l'opposer à des approches qui visent, elles, à enseigner des savoir-faire "purs".

## 6. Des systèmes qui enseignent des savoir-faire procéduraux simples

John S. Brown, Richard Burton et Kurt VanLehn ont étudié de manière minutieuse les erreurs d'enfants en train d'acquérir des savoir-faire scolaires comme les opérations arithmétiques. Ils concluent à l'existence de bogues procéduraux : leurs observations tendent à montrer que les erreurs ne sont pas dues simplement à du bruit ou à de des problèmes de "charge" cognitive, mais correspondent à l'application correcte de procédures erronées. Voyons cela sur un exemple :

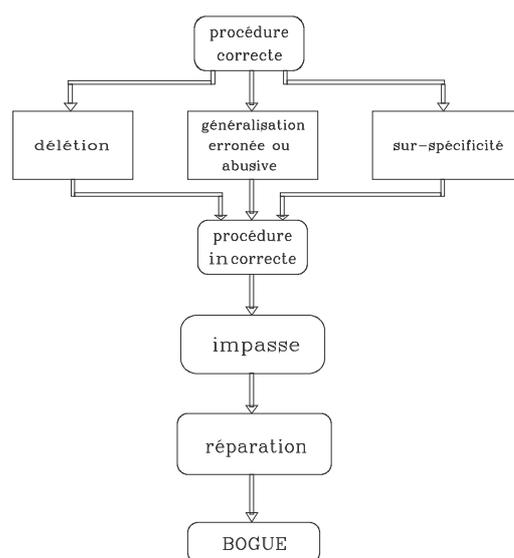
$$\begin{array}{r} \phantom{+} \phantom{1} \phantom{2} \phantom{7} \\ \phantom{+} \phantom{1} \phantom{2} \phantom{7} \\ + \phantom{1} \phantom{2} \phantom{7} \\ \hline \phantom{+} \phantom{1} \phantom{2} \phantom{7} \\ \phantom{+} \phantom{1} \phantom{2} \phantom{7} \\ \phantom{+} \phantom{1} \phantom{2} \phantom{7} \\ \phantom{+} \phantom{1} \phantom{2} \phantom{7} \end{array}$$

L'addition de droite montre que l'enfant porte "correctement" la retenue en haut de la colonne la plus à gauche. Malheureusement, cette procédure qui était juste pour les additions à deux chiffres se révèle erronée dans le cas général.

Le problème qui se pose est d'expliquer pourquoi certaines erreurs sont observées, et pas d'autres, et de rendre compte de leur genèse. En fait, dans la théorie développée par K. VanLehn [1990], les bogues (erreurs plus ou moins systématiques) ne sont pas la conséquence directe des procédures erronées. Dans l'exemple ci-dessus, l'élève qui pendant plusieurs mois n'a fait que des additions à deux chiffres a acquis une procédure sur-spécifique. Il porte la retenue dans la colonne qui est à la fois adjacente et la plus à gauche. Lorsqu'il passe à l'addition à trois chiffres, sa procédure est bloquée et il rencontre ce que VanLehn appelle une *impasse*. L'enfant devient alors inventif. Il peut choisir comme ici de relâcher une contrainte de sa procédure.

C'est à ce moment-là que l'enfant crée ce qui peut devenir une bogue, si l'action incorrecte est répétée et mémorisée. En effet, afin de poursuivre son exercice, il cherche à appliquer une procédure, celle qu'il voulait appliquer et dont il affaiblit les prérequis, ou une autre procédure indépendante. C'est ce que VanLehn appelle une *réparation*. Ce phénomène explique la genèse des bogues. De manière assez surprenante, il semble que le processus qui crée la bogue soit largement indépendant de la cause de l'impasse !

Derek Sleeman [1982] a lui aussi étudié la genèse des erreurs procédurales, notamment dans le cas de l'apprentissage de l'algèbre élémentaire. Ses conclusions sont assez proches, dans la mesure où il caractérise le comportement erroné des élèves par l'application correcte non pas d'une bogue répertoriée, mais d'une règle d'action déviante (*mal-rule*) insérée dans un système de règles censé représenter la connaissance procédurale que l'élève doit acquérir.



D. Sleeman s'est demandé quel était le type de correction approprié pour permettre aux élèves de corriger leur procédures. Le réflexe d'un enseignant est généralement d'explicitier les *raisons* qui font que la procédure appliquée par l'élève est erronée. A l'issue d'une expérience étonnante [Sleeman 1989] au cours

de laquelle la correction pouvait comporter des éléments d'explication (*reteaching*) en plus de la correction procédurale ("voici comment il faut faire"), il montre que l'explication n'a aucun effet sur les progrès des élèves ! Par exemple, à un enfant qui change un terme de membre sans en changer de signe, il vaut apparemment mieux se contenter de montrer qu'il faut changer le signe plutôt que de perdre son temps à donner la sémantique de l'opération. Il en est de même pour l'enfant qui remplace  $3x+5$  par  $8x$  !

Un tel constat aurait de quoi décourager la plupart des pédagogues. D. Sleeman s'est avoué lui-même fort surpris du résultat de l'expérience, et l'attribue, selon nous avec raison, au fait que ces procédures sont enseignées pour elles-mêmes, avec peu de références au modèle conceptuel qui leur donne leur signification. Les enfants apprennent la procédure, et la procédure seulement. Un peu comme dans une situation où l'on enseigne l'algorithme de la division, alors qu'on est incapable de justifier (sans quelque réflexion) le détail des actions accomplies (pourquoi suis-je en train de soustraire, pourquoi abaisser le chiffre suivant ?).

Il semblerait d'après ces travaux que la correction des bogues et des procédures erronées puisse être obtenue par un feed-back assez limité. Ceci n'est toutefois possible que dans le cas de procédures simples, qui doivent devenir des automatismes. Dans le cas de l'apprentissage des procédures exigeant des aspects stratégiques, il ne s'agit plus de corriger les erreurs, mais au contraire de faciliter l'acquisition d'une expérience par la présentation de cas appropriés. De nombreux systèmes reposent sur cet objectif. On peut citer par exemple le système APLUSIX [Nicaud et al. 1991] qui donne à l'élève des suggestions appropriées pour factoriser des polynômes.

Nous constatons ainsi que certains apprentissages peuvent se passer de l'acquisition de tout concept. Ce sont les apprentissages d'automatismes comme ceux qui nous permettent de résoudre une équation du premier degré. L'acquisition de stratégies, en revanche, commence à impliquer quelques concepts (par ex. les identités remarquables dans la factorisation des polynômes, ou le changement de variable dans le calcul intégral). Certaines procédures complexes sont, elles, indissolublement liées à des concepts, et leur acquisition passe nécessairement par une *compréhension* de ces concepts. C'est le point que nous abordons maintenant.

## 7. Des systèmes qui enseignent des savoir-faire procéduraux complexes

Comme l'a montré J-F. Richard [Poitrenaud, Richard et al. 1990], certaines procédures complexes ne peuvent être appliquées que lorsque l'on connaît conceptuellement les objets sur lesquelles elles opèrent. Il n'est pas évident, pour un débutant qui apprend à se servir d'un traitement de textes, qu'une image ou un saut de paragraphe puissent, pour certaines procédures comme la sélection, être considérés comme des caractères<sup>(1)</sup>. L'utilisation d'un traitement de texte comme celui qui a servi à l'édition de ce document<sup>(2)</sup> suppose connus de nombreux concepts (style, section, séquence nommée, champs, signets, fusion, etc.). La non compréhension de ces concepts interdit l'accès à de nombreuses actions, même par imitation (comme le suggère E. Visalberghi [1989]).

De nombreux systèmes visent à l'enseignement de procédures complexes, et il sont donc amenés par la force des choses à présenter certains concepts sur lesquels opèrent les procédures, et même parfois à présenter un modèle conceptuel des procédures enseignées.

L'apprentissage de la programmation nous offre un exemple d'imbrication entre procédures et concepts. Dans PROUST [Johnson & Soloway 1987], le système évoque bien sûr les concepts de la programmation en

---

<sup>1</sup> J-F Richard cite le cas, fréquent chez les personnes habituées à la machine à écrire, d'une procédure compliquée pour supprimer un espace inséré par erreur au milieu d'un mot. Ces sujets vont par exemple sélectionner la première moitié du mot, puis réaliser un copier-coller pour l'amener de l'autre côté de l'espace parasite ! Ce n'est que plus tard qu'ils avouent découvrir qu'un espace est un caractère comme les autres, qu'il peut être détruit par une action sur la touche [Del]. L'espace était pour eux un vide, et il ne pensaient pas à détruire quelque chose de vide.

<sup>2</sup> Le traitement de texte employé est Microsoft-Word.

langage Pascal (boucle *while*, affectation, etc.), mais il introduit aussi des concepts comme le "test sentinelle", qui ne fait pas partie du langage, mais dont l'objectif est de protéger l'entrée dans la boucle. Nowé et Jonkers [1991] disent utiliser des "concepts abstraits de programmation" pour formuler des diagnostics de haut niveau (en Lisp ou Pascal). Dans LISP-TUTOR [Anderson et al. 1989] les diagnostics utilisent des concepts comme "variable de comptage", "communication d'une valeur d'initialisation à une fonction". INTELLIGENT-PROLOG-TUTOR [Lee 1990] signale à l'apprenant des erreurs comme des définitions circulaires. LISP-CRITIC [Fischer et al. 1991] est capable de suggérer des transformations d'un programme Lisp en donnant des explications relatives à l'efficacité ou à la lisibilité.

Dans tous ces systèmes qui ont pour objectif d'aider l'étudiant à acquérir un savoir-faire en programmation, on constate que les diagnostics sont généralement conceptuels puisqu'ils font intervenir des notions de programmation (par ex. PATAT [Nowé & Jonkers 1991] peut écrire : "*variable result must be initialised to 0 because result is the running total variable and the unity element for the sum operator is 0*", par opposition à un diagnostic non conceptuel qui pourrait être : "*cette clause est erronée, voici la forme correcte*"). Autrement dit l'effet pédagogique de tels diagnostics nécessite un examen par l'étudiant de ses propres connaissances.

Nous décrirons (chap. III.4) notre propre tentative pour développer un système capable d'aider à l'acquisition de savoir-faire complexes par un diagnostic conceptuel.

## 8. Conclusion

Nous avons tenu ici à insister sur la différence, essentielle de notre point de vue, entre les apprentissages conceptuels et les apprentissages procéduraux. Les exemples de systèmes que nous avons cités indiquent, comme nous le suggérons par ailleurs, que l'argumentation est un moyen privilégié pour la transmission des concepts. Elle joue notamment un rôle essentiel pour l'élimination des fautes conceptuelles (*misconceptions*). C'est ce principe que nous avons exploité dans notre système SAVANT3 (voir chapitre III.2).

L'argumentation peut, si l'on en croit certaines études, n'être d'aucune utilité dans certains contextes où les apprenants doivent acquérir des automatismes. En revanche, lors de l'acquisition de procédures complexes, une compréhension des concepts du domaine est essentielle pour acquérir une expertise. Là encore l'argumentation, qui se présente alors comme une critique des actions entreprises par l'apprenant, peut jouer un rôle essentiel.

Il y a donc place pour des systèmes que l'on présentera plus volontiers comme des critiques que comme des tuteurs. Dans le cas d'apprentissages conceptuels, il s'agit de repérer les conceptions erronées par leurs conséquences incohérentes. Dans le cas des savoir-faire complexes, c'est un peu la même chose, à part que l'incohérence doit être repérée au niveau des actions de l'apprenant sur son environnement, et pas seulement au niveau de ses déclarations. SAVANT3 s'inscrit résolument dans cette optique qui consiste à utiliser la critique argumentée pour faire passer et pour corriger des connaissances conceptuelles.

## **Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage**

### **1. Introduction**

Nous décrivons ici le système SAVANT3 que nous avons développé à TELECOM-Paris. Dans son principe, ce système utilise certains des mécanismes argumentatifs que nous avons pu analyser en observant des conversations spontanées, afin de tester la cohérence des connaissances de l'étudiant. Il se caractérise par une grande économie de mise en oeuvre : une quinzaine de relations logiques suffisent à SAVANT3 pour soutenir une conversation sur un sujet donné. Le caractère naturel de l'interaction que ce système tente de reproduire permet d'espérer qu'il sera bien accepté par les étudiants.

### **2. L'importance pédagogique d'un contexte argumentatif**

#### **2.1. L'apprentissage au cours de la conversation**

Nous avons mis en évidence dans le chapitre III.1 le fait que la plupart des systèmes dont l'objectif était l'enseignement de connaissances conceptuelles utilisaient le dialogue dit "Socratique" comme mode d'interaction avec l'élève. La justification de ce type d'approche utilise généralement comme situation modèle la relation professeur élève, comme on peut l'observer par exemple dans des cours particuliers [Collins 1976]. Ce genre de justification a été critiqué [Petrie-Brown 1989]. Pour notre part, nous cherchons à ancrer l'intérêt didactique du type de dialogue que nous essayons d'instaurer entre l'élève et la machine, non dans la relation institutionnelle enseignant-enseigné, mais dans sa contrepartie naturelle, la conversation spontanée.

Nous avons laissé entendre dans les chapitres II.3 et II.4 que la perception d'une situation incohérente pouvait jouer le rôle de déclencheur pour l'apprentissage d'un nouveau concept ou pour la modification de relations conceptuelles. Prenons un exemple conversationnel particulièrement clair de ce point de vue :

[ex\_fond]

contexte : A s'étonne d'améliorer certaines de ses performances sportives alors qu'il atteint la trentaine.

A1- *Moi, avec l'âge, je peux tenir des trucs en fond, mais dès que c'est le sprint...*

B1- *Normal.*

A2- *Je tiens mieux le fond qu'avant. C'est dingue, ça!*

B2- *Normal, normal. Mais ça c'est normal, hein. On est plus endurant que résistant. Avec l'âge.*

### C1- Quelle est la différence?

[...]

L'incohérence est clairement exprimée en A2. Sa solution vient de l'introduction des concepts d'endurance et de résistance: l'endurance culmine plus tard que la résistance. La situation de A est ainsi présentée par B en B2 comme un cas particulier d'une loi plus générale que A et C ne semblaient pas soupçonner.

Nous pouvons considérer qu'un apprentissage conceptuel, comme ici la distinction endurance / résistance et son lien avec l'âge, intervient comme l'*explication* d'une situation perçue comme incohérente (voir chap. III.4)<sup>(1)</sup>. Mais pourquoi chercher à reproduire une situation de type conversationnel entre l'étudiant et la machine ? Après tout, on pourrait imaginer que le fait de donner de simples définitions des concepts pourrait avoir une efficacité équivalente !

La réponse à cette question est essentielle pour notre propos. Nous retenons comme critère d'efficacité pédagogique, en ce qui concerne l'apprentissage des connaissances conceptuelles, les deux aspects suivants :

⇒ **compréhension**

⇒ **mémorisation**

## 2.2. Comprendre un nouveau concept grâce à un contexte problématique

Que veut dire précisément *comprendre* ? En prolongement du chapitre II.3, nous suggérons le fait qu'un concept sera compris dès que les relations logiques qui le relie à d'autres concepts seront correctement connues. Cette définition nous permet de faire les remarques suivantes :

- un concept compris permet de s'étonner à bon escient, et réciproquement
- un concept compris permet d'expliquer les situations qui le mettent en jeu, et réciproquement
- la compréhension d'un concept n'est pas un phénomène graduel<sup>(2)</sup> (cf. chap. III.1).

En effet, si le concept  $c$  entre dans des relations logiques avec d'autres concepts (nous représentons ces relations par des clauses négatives, dites aussi clauses paradoxales) :

$$[ c \ \& \ c_1 \ \& \ c_2 \ \dots \ \& \ c_m ] \Rightarrow \mathbf{F}$$

$$[ \mathbf{non} \ c \ \& \ c'_1 \ \& \ c'_2 \ \dots \ \& \ c'_n ] \Rightarrow \mathbf{F}$$

alors ces relations sont autant d'occasions d'étonnement. De plus, *non c* et  $c$  sont respectivement des explications (invalidations indirectes, comme négations de prémisses omises, cf. chap. III.4) relatives aux étonnements suivants, si un interlocuteur les manifeste :

$$[ c_1 \ \& \ c_2 \ \dots \ \& \ c_m ] \Rightarrow \mathbf{F}$$

$$[ c'_1 \ \& \ c'_2 \ \dots \ \& \ c'_n ] \Rightarrow \mathbf{F}$$

Compte tenu de ces remarques, on comprend l'importance du contexte argumentatif pour la compréhension. Toute relation logique qui lie un nouveau concept  $c$  à des concepts déjà connus  $\{c_i\}$  peut donner lieu à un étonnement. Supposons qu'un étonnement ait effectivement lieu, et appelons *clause de référence* (CR) la relation logique qui traduit cet étonnement. Un contexte argumentatif est l'ensemble formé

<sup>1</sup> L'invalidation d'une situation improbable ou d'une situation (in)désirable peut donner lieu, de même, à un apprentissage conceptuel. Nous n'en parlerons pas, car le mécanisme est exactement le même que pour l'incohérence (paradoxe).

<sup>2</sup> Nous nous situons là dans le cas où le contexte est spécifié. Il serait certes possible de voir une gradualité dans le fait qu'entre tous les contextes imaginables, une proportion variable des relations logiques incluant le nouveau concept est connue.

par la CR et les clauses qui permettent de relier les autres concepts  $\{c_i\}$  de la CR à des concepts tenus pour vrais dans la situation (comme expliqué dans le chapitre I.2).

Par exemple, le contexte argumentatif de l'extrait précédent peut être représenté ainsi :

$\forall$  Personne ,  $\forall$  Sport [ plus-âgé(Personne) & meilleur(Personne, Sport) ]  $\Rightarrow$  F

$\exists A$  ,  $\exists$  nage-de-fond plus-âgé(A) & meilleur(A, nage-de-fond)

Le concept *endurance* vient expliquer la situation, puisqu'il intervient sous forme négative dans une version plus complète du contexte :

$\forall$  Personne ,  $\forall$  Sport [ plus-âgé(Personne) & non endurance(Sport)  
& meilleur(Personne, Sport) ]  $\Rightarrow$  F

La présence d'un contexte argumentatif assure que le nouveau concept sera *compris*, dans la mesure où il permet à l'apprenant d'expliquer une situation. Cette explication insère le nouveau concept dans un réseau de relations avec les concepts déjà connus. Comprendre un concept, c'est donc accepter des relations conceptuelles (que l'on peut représenter comme des relations d'incompatibilité) qui lient ce concept aux concepts connus. Mais qu'en est-il de la mémorisation ?

### 2.3. Mémoriser un nouveau concept grâce à un contexte problématique

Cette question de la mémorisation de connaissances conceptuelles a été abordée d'une manière tout à fait intéressante pour notre propos par John Bransford et Barry Stein [1982]. Ces auteurs ont réalisé l'expérience suivante. Ils présentèrent à des étudiants des phrases comme celles-ci :

1<sup>ère</sup> série

- |                                    |  |
|------------------------------------|--|
| ◆ The tall man bought the crackers | (l'homme grand a acheté les biscuits)  |
| ◆ The bald man read the newspaper  | (l'homme chauve lisait le journal)     |
| ◆ The funny man liked the ring     | (l'homme amusant a bien aimé la bague) |
| ◆ The hungry man purchased the tie | (l'homme affamé a acheté la cravate)   |
| ◆ The short man used the broom     | (le petit homme a utilisé le balai)    |
| ◆ The strong man skimmed the book  | (l'homme musclé feuilletait le livre)  |

Ces phrases sont construites de telle manière que la relation entre la caractéristique de l'homme qui fait l'action et l'action elle-même soit arbitraire. On demande ensuite aux étudiants de ranger ces phrases dans l'ordre de leur facilité de compréhension. Mais le but du test n'est pas là. Sans qu'ils soient prévenus, les étudiants subissent un test de mémorisation : "Quel est l'homme qui acheté la cravate ?", "Quel est celui qui aimait la bague ?", etc. Le résultat obtenu par les étudiants fut médiocre : 4,2 phrases sur 10 en moyenne.

Il semble qu'un *lien* entre les éléments d'une phrase à mémoriser puisse favoriser la mémorisation. Mais quelle est la nature de ce lien ? Pour le savoir, Bransford et Stein ont présenté à un autre groupe d'étudiants les phrases suivantes :

2<sup>ème</sup> série

- |   |
|---|
| ◆ The tall man bought the crackers that had been lying on the top shelf           |
| ◆ The bald man read the newspaper in order to look for a hat sale                 |
| ◆ The funny man liked the ring that squirted water                                |
| ◆ The hungry man purchased the tie so that he could get into the fancy restaurant |
| ◆ The short man used the broom to operate the light switch                        |
| ◆ The strong man skimmed the book about weight lifting                            |

Cette fois le taux de mémorisation atteint 7,4 / 10. Pour les auteurs, l'interprétation ne faisait aucun doute :

"[...] memory was enhanced because the elaborations helped people understand the significance or relevance of linking a particular type of man to a particular activity." [Bransford & Stein 1982]

Pourtant d'autres interprétations ont été émises, notamment le fait qu'une élaboration reliée sémantiquement aux éléments de la phrase aidait à la mémorisation. Autrement dit, le fait d'ajouter des éléments du même contexte *sémantique* va permettre aux étudiants de se souvenir des phrases. Bransford et Stein ont voulu tester cette idée en essayant des phrases comme celles-ci :

### 3<sup>ème</sup> série

- ◆ The tall man bought the crackers from the clerk in the store
- ◆ The bald man read the newspaper while eating breakfast
- ◆ The funny man liked the ring that he received as a present
- ◆ The hungry man purchased the tie that was on sale
- ◆ The short man used the broom to sweep the porch
- ◆ The strong man skimmed the book before going to sleep

"These statements included elaborations that make sense semantically, but the elaborations did not help understand why it might be relevant for a particular type of man to perform a particular activity. The results indicated that students who received sentences such as these exhibited poorer memory than students who received the first list of base sentences (the list without any additional elaborations). It seems clear, therefore, that the semantic congruity of elaborations does not necessarily lead to increase retention." [Bransford & Stein 1982]

Une simple élaboration sémantique ne favorise donc pas la mémorisation. Il faut davantage. Il faut un lien "pertinent". Notre propre interprétation de cette expérience consistera simplement à préciser ce terme de pertinence (*relevance*) employé par les auteurs pour désigner les "élaborations" qui facilitent la mémorisation. Le rôle de l'information additionnelle, dans la série de phrases bien mémorisées (2<sup>ème</sup> série), est de faciliter l'insertion de l'action de chacun des personnages dans un contexte "problématique". Voici des exemples de contextes qui peuvent venir spontanément à l'esprit de celui qui lit la deuxième série de phrases :

- ◆ Nobody could see these crackers, but fortunately this man did.
- ◆ The bald man needs a hat, otherwise he may have a cold
- ◆ Some people may take offence at having their face splashed, not this funny man
- ◆ The hungry man was refused the entrance into the restaurant
- ◆ The short man could not reach the light switch
- ◆ Nobody can be interested in such a book, but a strong man

Nous pensons que la deuxième série de phrases a été bien mémorisée parce que les sujets ont imaginés ces contextes problématiques<sup>(3)</sup>. Pour la 3<sup>ème</sup> série, les sujets ont pu imaginer une scène pour chaque phrase, mais l'établissement d'un contexte problématique est, en revanche, beaucoup moins immédiat<sup>(4)</sup>.

L'important est de constater, et cette expérience nous invite à le faire, qu'un *lien logique* entre les éléments d'une phrase facilite la mémorisation. Bransford et Stein suggèrent eux-mêmes les conséquences pédagogiques d'un tel constat :

"[...] even biology novices who know the general terms *vein* and *artery* may have difficulty remembering which ones have particular properties because the relationships between each concept (vein, artery) and its properties (e.g. elastic, thick) seem arbitrary." [...] "biology experts who are presented with long lists of simple facts (e.g. arteries are elastic) versus elaborated facts (e.g. arteries are elastic, thick and have no valves) should produce [...] better free recall for the concept (e.g. artery) when it is more elaborated [...]"

<sup>3</sup> Ces exemples jouent sur la modalité désirable / indésirable, mais le même effet serait obtenu avec les modalités paradoxale ou improbable (cf. note p. 230).

<sup>4</sup> Dans les deux séries de phrases "élaborées" présentées par Bransford et Stein, la dernière phrase est sans doute la moins intéressante. Le lien entre l'homme fort et le livre d'haltérophilie est surtout associatif, alors que le fait de feuilleter le livre peut être facilement lié logiquement à l'envie de dormir.

For biological novices, however, the relationships between concepts and properties are arbitrary [...] Under these conditions, semantically congruous elaborations may hurt retention rather than help." [Bransford & Stein 1982]

Nous allons adopter le paradigme suivant :

*Un contexte argumentatif, qui comporte une situation problématique, facilite la compréhension et la mémorisation de nouvelles connaissances.*

Notre objectif, avec le développement de SAVANT3, est de nous assurer que tout nouveau concept présenté aux étudiants sera inséré dans un contexte argumentatif. Les étudiants peuvent alors plus facilement relier logiquement le nouveau concept à d'autres concepts déjà connus. Le moyen par lequel ces liens logiques seront rendus manifestes, dans notre application, consiste dans la présentation de situations potentiellement incohérentes.

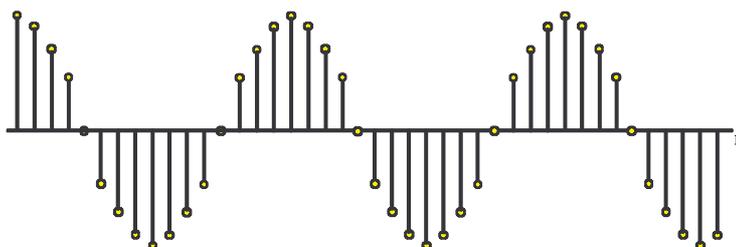
### 3. Principe de fonctionnement de SAVANT3

SAVANT3 repose sur un principe inspiré de l'observation des conversations : présenter à l'étudiant un contexte argumentatif organisé autour d'une contradiction. Nous venons d'indiquer l'importance de ce type de contexte pour la compréhension et la mémorisation des connaissances. Dans le chapitre III.1, nous avons rappelé qu'un principe similaire était à l'oeuvre dans la stratégie de piège du célèbre système WHY. Nous retrouvons une stratégie similaire dans le système NF-Tutor de Thomas Rätz [1992]. T. Rätz note que les explications fournies dans le cas où une contradiction est détectée sont pédagogiquement plus efficaces, et qu'elles peuvent être calculées (par opposition à une situation où les explications seraient issues de "canevas" (*pattern*) prévus à l'avance). Toutefois, son système repose sur l'existence d'une liste de conceptions erronées que l'auteur est obligé de fournir explicitement. L'explication par la mise en évidence d'une contradiction est déclenchée lorsqu'une erreur commise par l'étudiant peut être rapprochée d'une de ces conceptions erronées.

Dans notre système, les conceptions erronées ne sont pas explicites au départ. C'est l'occurrence d'une contradiction qui est détectée en premier. Cette contradiction va provoquer un étonnement de la machine, puis conduire à un dialogue. C'est ce dialogue qui va permettre à l'étudiant de corriger la conception erronée si celle-ci existe.

Au début d'une interaction avec SAVANT3, la situation n'est généralement pas incohérente. Le système va donc poser des questions ou émettre des répliques qui visent en fait à amener l'étudiant à préciser (instancier) la situation jusqu'à ce que le système puisse établir une incohérence.

Nous pouvons observer cette stratégie dans l'extrait suivant, qui porte sur la périodicité d'une sinusoïde discrète  $\cos[2\pi F_1 n]$ . Ce signal est périodique si et seulement si  $F_1$  est rationnelle.



Cette propriété mathématique peut donner lieu à un dialogue comme celui-ci entre la machine (M) et l'étudiant (E) (les réponses entre crochets indiquent un choix fermé) :

M1: Le signal  $\cos[2\pi F_1 n]$  (pour  $F_1$  donnée) est-il périodique ?  
 E1: oui  
 M2: D'après ce que j'ai compris: le signal  $\cos[2\pi F_1 n]$  (pour  $F_1$  donnée) est périodique.  
 On pourrait prendre par exemple  $F_1 = 1/2\pi$ .  
 E2: oui  
 M3: mais encore ?...  
 E3: ça fait  $\cos[n]$   
 M4: Peut-on dire que  $\cos[n]$  est périodique?  
 E4: [non]  
 M5: Attendez... D'après ce que vous m'avez dit:  
 - le signal  $\cos[2\pi F_1 n]$  (pour  $F_1$  donnée) est périodique  
 -  $\cos[n]$  n'est pas périodique  
 Je m'avoue quelque peu surpris !  
 E5:  $F_1$  doit être entière  
 M6: D'après ce que j'ai compris:  $F_1$  est un entier.  $\cos[2\pi F_1 n]$  risque de ne pas beaucoup ressembler à une sinusoïde !  
 E6: pourquoi?  
 M7: Mais cela pose un problème... sauf si vous considérez que  $\cos[2\pi F_1 n]$  est un signal constant  
 ...  
 E7:  $F_1$  doit être l'inverse d'un entier  
 M8: oui... mmm... D'accord...

Nous fournissons en annexe 1 d'autres extraits, dont un qui est obtenu à partir de la même connaissance, pour un élève qui fait des choix différents.

Afin de commenter le fonctionnement de SAVANT 3 pendant cet extrait, nous donnons ci-dessous les principales règles dont dispose le système sur ce sujet. Ces règles sont présentées (et utilisées par le système) sous forme d'incompatibilités. Par exemple la règle n°2 indique qu'il est impossible que  $F_1$  soit rationnelle et que la sinusoïde ne soit pas périodique. Le chiffre entre parenthèses indique si la règle doit être considérée comme évidente (règle de "bon sens", valeur 1) ou non (valeur 0).

règle 1 (0):

le signal  $\cos[2\pi F_1 n]$  (pour  $F_1$  donnée) est périodique  
 $\cos[n]$  n'est pas périodique  
 la fréquence  $F_1$  n'est pas rationnelle

règle 5 (0):

$F_1$  est un entier  
 $\cos[2\pi F_1 n]$  n'est pas un signal constant

règle 2 (0):

la fréquence  $F_1$  est rationnelle  
 le signal  $\cos[2\pi F_1 n]$  (pour  $F_1$  donnée) n'est pas périodique

règle 6 (0):

l'échantillonnage a lieu dans les conditions de Nyquist  
 la fréquence d'échantillonnage  $f_e$  est diviseur de  $f_0$

règle 3 (0):

$\cos[2\pi F_1 n]$  vient de l'échantillonnage de  $\cos(2\pi f_0 t)$   
 la fréquence  $F_1$  n'est pas rationnelle  
 la fréquence d'échantillonnage  $f_e$  est multiple de  $f_0$

règle 7 (1):

$F_1$  est un entier  
 la fréquence  $F_1$  n'est pas rationnelle

règle 4 (1):

$\cos[2\pi F_1 n]$  vient de l'échantillonnage de  $\cos(2\pi f_0 t)$   
 $F_1$  n'est pas un entier  
 la fréquence d'échantillonnage  $f_e$  est diviseur de  $f_0$

règle 8 (1):

la fréquence d'échantillonnage  $f_e$  est diviseur de  $f_0$   
 la fréquence d'échantillonnage  $f_e$  est multiple de  $f_0$

règle 9 (0):

$\cos[2\pi F_1 n]$  ne vient pas de l'échantillonnage de  $\cos(2\pi f_0 t)$

règle 10 (0):

cos[2πF<sub>1</sub> n] vient de l'échantillonnage de cos(2πfo t)  
F<sub>1</sub> est un entier  
la fréquence d'échantillonnage fe n'est pas diviseur de fo

règle 11 (1):

cos[n] est périodique  
cos[0] est égal à 1  
cos[n] n'est pas égal à 1 pour n non nul

règle 12 (1):

cos[2πF<sub>1</sub> n] est un signal constant  
F<sub>1</sub> n'est pas un entier

règle 13 (0):

cos[n] est égal à 1 pour n non nul

règle 14 (1):

cos[0] n'est pas égal à 1

règle 15 (1):

la fréquence F<sub>1</sub> est rationnelle  
la fréquence F<sub>1</sub> est quelconque

règle 16 (1):

le signal cos[2πF<sub>1</sub> n] (pour F<sub>1</sub> donnée) n'est pas  
périodique  
la fréquence F<sub>1</sub> n'est pas rationnelle  
cos[n] est périodique

SAVANT3 utilise un moteur en chaînage arrière écrit en Prolog. Actuellement, ce moteur fonctionne de la manière suivante. Pour démontrer une proposition  $p$ , il recherche  $\text{non } p$  dans une règle de la base. Si ( $\text{non } p$ )  $\Rightarrow F$  est une règle de la base, la démonstration réussit. Si  $\text{non } p$  est trouvée dans une règle en conjonction avec d'autres termes, le moteur cherche à démontrer ces termes sans utiliser  $p$  ni  $\text{non } p$ . Enfin si  $\text{non } p$  n'est pas trouvée, la démonstration échoue.

Ce moteur travaille donc en monde ouvert (pour une terme donné  $p$ , la démonstration de  $p$  et la démonstration de ( $\text{non } p$ ) peuvent toutes deux échouer). Les faits appris en cours d'interaction avec l'étudiant sont ajoutés au fur et à mesure à la base de règles (dans la mesure où ils sont compatibles avec celle-ci). Le fonctionnement du moteur est décrit avec davantage de détails en annexe 2.

Nous décrivons maintenant le fonctionnement de SAVANT3 dans le dialogue que nous avons reproduit.

M1: Le signal cos[2πF <sub>1</sub> n] (pour F <sub>1</sub> donnée) est-il périodique ?	$\Rightarrow$ SAVANT3 émet la question associée par l'auteur à la proposition "le signal ... est périodique". L'auteur a imposé que ce terme soit connu dès le démarrage.
E1: oui	
M2: D'après ce que j'ai compris: le signal cos[2πF <sub>1</sub> n] (pour F <sub>1</sub> donnée) est périodique. On pourrait prendre par exemple F <sub>1</sub> = 1/2π .	$\Rightarrow$ Tentative pour saturer la règle 1. Reprise des termes appris de la règle en cours de saturation. $\Rightarrow$ SAVANT3 émet la "question" associée par l'auteur à la proposition "cos[n] est périodique"
E2: oui	
M3: mais encore ?...	$\Rightarrow$ oui n'est pas un mot clé local prévu par l'auteur pour la proposition attendue, et n'est un mot clé global pour aucun autre prédicat. Mais il est reconnu comme une réponse "vide" (comme certes, bien sûr, etc.)
E3: ça fait cos[n]	$\Rightarrow$ réponse non reconnue
M4: Peut-on dire que cos[n] est périodique?	$\Rightarrow$ SAVANT3 pose la question directement, sans passer par la question prévue par l'auteur
E4: [non]	$\Rightarrow$ l'élève répond dans un choix fermé oui/non
M5: Attendez... D'après ce que vous m'avez dit: - le signal cos[2πF <sub>1</sub> n] (pour F <sub>1</sub> donnée) est périodique - cos[n] n'est pas périodique Je m'avoue quelque peu surpris !	$\Rightarrow$ La règle 1 n'est pas encore saturée, car le statut de son troisième terme est inconnu. SAVANT3 feint la surprise.
E5: F <sub>1</sub> doit être entière	$\Rightarrow$ Cette réponse (reconnue par les mots clés F <sub>1</sub> et entier) invalide l'étonnement grâce à la règle 7.

M6:	D'après ce que j'ai compris: $F_1$ est un entier. $\cos[2\pi F_1 n]$ risque de ne pas beaucoup ressembler à une sinusoïde !	$\Rightarrow$ SAVANT3 tente de saturer la règle 5 dont il connaît un terme sur deux. Il émet la question associée par l'auteur à la proposition "cos... signal constant"
E6:	pourquoi?	$\Rightarrow$ L'étudiant ne voit pas le lien
M7:	Mais cela pose un problème... sauf si vous considérez que $\cos[2\pi F_1 n]$ est un signal constant ...	$\Rightarrow$ SAVANT3 reconnaît une demande d'explication. Il explicite le terme suivant de la clause en cours de saturation.
E7:	$F_1$ doit être l'inverse d'un entier	$\Rightarrow$ SAVANT3 assimile cette réponse à "F1 rationnelle"
M8:	oui... mmm... D'accord...	$\Rightarrow$ Toutes les clauses candidates sont invalidées

L'auteur a la possibilité d'imposer à SAVANT3 de démarrer l'interaction en recherchant la valeur de vérité d'une proposition donnée. Ainsi, en M1, le programme énonce la phrase associée par l'auteur à la proposition à valider : "*le signal  $\cos[2\pi F_1 n]$  (pour  $F_1$  donnée) est périodique*".

La stratégie employée par SAVANT3 pour gérer le dialogue consiste à essayer de valider une incompatibilité. Pour cela, il considère des règles "candidates", c.à.d. des règles qui ne sont pas marquées comme "évidentes", qui sont encore valides (une règle est valide si aucun de ses termes n'est démontrable comme étant faux) et qui ont au moins un terme vrai. Il va sélectionner, parmi ces règles candidates, la plus "prometteuse", c'est-à-dire celle qui contient le plus grand pourcentage de termes connus. Cette règle va provisoirement devenir la clause de référence.

Ainsi, en M2, c'est la règle 1 qui est sélectionnée. Le système connaît en effet le premier terme, qu'il vient d'apprendre, mais aussi le deuxième terme ( *$\cos[n]$  n'est pas périodique*), car il peut le démontrer avec les règles 11, 13 et 14.

Pourtant, bien qu'il sache le démontrer, SAVANT3 va émettre la phrase associée à ce deuxième terme pour amener l'élève à le valider (deuxième partie de M2). En effet, ce deuxième terme est connu du programme, mais sa démonstration fait intervenir des clauses qui ne sont marquées comme évidentes (règle 13). M2 a donc pour objectif d'amener l'étudiant à préciser si  *$\cos[n]$  est périodique*, ou mieux, à lui permettre d'anticiper la contradiction.

SAVANT3 détecte des mots clés dans les réponses de l'étudiant, ainsi que la présence d'une négation. Il ne reconnaît pas E3, et propose le choix fermé M4. Après E4, tous les termes de la clause de référence sauf un sont validés. SAVANT3 peut donc s'étonner en M5. Mais l'explication E5 invalide la clause de référence 1 grâce à la règle 7.

Le programme change alors de clause de référence, et se porte sur la règle 5. Dans la deuxième partie de M6, il tente de connaître la validité de la deuxième proposition de la règle 5. Après E6, perçue comme une demande d'explicitation, le programme livre le terme suivant de sa clause de référence en M7. Mais l'élève change d'avis en E7. Le programme comprend que  $F_1$  est rationnelle non entière, et il arrête la conversation car toutes les règles candidates (comme définies plus haut) se trouvent invalidées.

L'argumentation dépend des répliques de l'élève, qui a toujours plusieurs solutions pour rester cohérent. Ici par exemple l'élève aurait pu assumer le choix " *$F_1$  entière*", mais la machine aurait argumenté à propos de la fréquence d'échantillonnage. Il aurait pu aussi faire le choix " *$\cos[2\pi F_1 n]$  non périodique*", etc. (voir annexe 1).

#### 4. L'acceptabilité d'une conversation avec SAVANT3

On peut être frappé en premier lieu par la "rigidité" de la manière de s'exprimer de SAVANT3. Cette rigidité est facile à corriger, par l'introduction d'une variabilité des messages utilisés. Nous avons choisi de ne pas effectuer cette modification tant que SAVANT3 est en phase de développement. En effet, notre espoir est que la pertinence de l'argumentation repousse au second plan ces problèmes d'expression.

SAVANT3 a déjà été testé dès 1988, ce qui a permis de cerner certaines limitations. Il a subi depuis des modifications assez profondes. La principale difficulté rencontrée concerne la clause de référence (voir p. 200). Le type d'interaction qu'il essaie d'instaurer repose en effet entièrement sur la *perception par l'élève de la contradiction* exprimée dans la clause de référence. C'est cette perception qui garantit la pertinence logique, pour l'étudiant, des répliques de la machine (ce qui n'est pas le cas pour M2) et de ses étonnements (ce qui est le cas pour M5, mais pas pour M6). Il s'agit là d'un problème délicat qui nécessite des expérimentations avec des étudiants. L'un des principaux moyens utilisés actuellement pour tenter de remédier au problème consiste à rappeler les termes connus de la clause de référence (comme en première partie de M2, M5, M6).

Puisque l'efficacité pédagogique de SAVANT3 repose entièrement sur la perception, par l'étudiant, de l'étonnement de la machine, nous pouvons aisément comprendre que le dialogue entre eux n'est possible que si tous deux manipulent les mêmes concepts. L'étudiant doit connaître les propositions contenues dans les règles de la base de connaissances (*i.e.* le sens des mots qui les désignent), ainsi qu'une partie de la théorie du domaine (*i.e.* certaines des règles de la base, avec éventuellement des termes manquants). De telles conditions exigent que le contexte de l'interaction pédagogique soit fortement cadré. Si elles ne sont pas remplies, l'interaction échoue inmanquablement.

Le dialogue peut échouer pour d'autres raisons. Il peut arriver que l'étudiant ne partage pas les objectifs du dialogue, parce que le sujet ne l'intéresse tout simplement pas. Il n'entrera pas dans le dialogue, et refusera de prendre en compte le premier étonnement de la machine.

SAVANT3 ne se préoccupe pas de la négociation des objectifs du dialogue, contrairement à d'autres systèmes comme KANT [Baker 1991a] ou le système de M. Joab [Joab 1990]. De tels systèmes reconnaissent les questions de l'utilisateur qui portent sur l'opportunité de la tâche elle-même (ce sont des "méta-interventions"). Notre système néglige cet aspect, pour deux raisons principales. Premièrement, nous ne sommes pas ici en situation de résolution de tâche, et la notion d'objectif est donc moins évidente. Deuxièmement, le pari qui est à la base de l'approche est que l'élève sera motivé par le sentiment que ses déclarations sont contradictoires, comme le sont les interlocuteurs dans les conversations spontanées (M. Baker [1991b] identifie aussi le "conflit interne" comme générateur d'actes dialogiques dans des dialogues de résolution de problème). Toute la difficulté, comme on l'a dit, est donc de parvenir à faire ressentir cette contradiction le plus rapidement possible.

Rappelons enfin que, de par sa conception même, SAVANT3 a un domaine d'emploi bien délimité. Comme nous l'avons dit dans le chapitre III.1, il est destiné à l'acquisition de concepts (ou à l'élimination de fautes conceptuelles), ce qui exclut les situations d'acquisitions purement procédurales. Les tentatives pour faire traiter à SAVANT3 des sujets incluant des savoir-faire ne vont pas de soi. Nous envisagerons tout de même dans le chap. III.4 l'utilisation de SAVANT3 comme un système critique.

## 5. Conclusion

L'objet de ce chapitre était de présenter SAVANT3, qui est un système d'Enseignement Assisté par Ordinateur bâti sur des principes originaux issus de notre théorie de la pragmatique conversationnelle. SAVANT3 est conçu pour converser avec l'étudiant à propos de sujets techniques. Lorsque la conversation s'instaure, l'étudiant apprend : il modifie sa théorie du domaine, jusqu'à ce que celle-ci soit compatible avec celle du système.

Notre pari est que ce type d'interaction, basé sur la résolution de contradiction, sera agréable pour l'étudiant. Il est inspiré de l'observation des conversations naturelles, et nous pouvons ainsi espérer que l'étudiant prendra plaisir à argumenter avec la machine comme il le fait dans la vie courante.

L'un des intérêts principaux de SAVANT3 vient du fait qu'il repose sur peu de principes. Si SAVANT3 parvient à être validé comme moyen d'enseigner, alors il sera possible de faire remonter cette validation vers les principes qui gouvernent son fonctionnement. Ceux-ci sont par ailleurs suffisamment précis pour que l'auteur n'ait pas à gérer le dialogue. SAVANT3 est sans doute l'un des premiers systèmes qui évite à l'auteur

de se soucier de l'interaction. L'auteur se contente idéalement de fournir une connaissance au système, et c'est le système qui prend en charge la relation avec l'apprenant.

Bien entendu, SAVANT3 doit être perçu comme une partie d'un système d'EAO autosuffisant. Dans nos expérimentations, il était inséré dans un ensemble comportant un programme de présentation "narrative" des concepts, un système multimédia interactif (SAVANT2), une encyclopédie conceptuelle (SAVANT1), et des programmes de simulation. Les dialogues n'interviennent qu'aux moments où l'auteur veut s'assurer que l'étudiant a une théorie correcte du domaine qui vient de lui être présenté.

Nous décrivons, dans le chapitre III.3, le système auteur associé à SAVANT3. Ce système auteur fait encore l'objet de recherches. Notre objectif est de le baser lui aussi sur les principes de la conversation, de manière à faciliter le travail de l'auteur et ainsi à parvenir à un système viable sur le plan économique.

## Annexe 1 : Exemples d'interaction avec SAVANT3

L'extrait suivant utilise la même connaissance que celle de l'extrait donné dans le texte du chapitre. Le dialogue qui s'ensuit est assez différent de celui qui a été analysé. Noter que grâce au dialogue, l'étudiant (fictif) de ce dialogue prend conscience de l'insuffisance de la contrainte  $F1$  non entière.

*le signal  $\cos[2\pi F1 n]$  (pour  $F1$  donnée) est-il périodique ?*  
 ----- non  
*D'après ce que j'ai compris: le signal  $\cos[2\pi F1 n]$  (pour  $F1$  donnée) n'est pas périodique. tout cela a des conséquences sur  $F1$  !*  
 ----- oui  
*mais encore ?...*  
 -----  $F1$  ne peut pas être entière  
 Ah, Ok... oui...mmm...  
*D'après ce que j'ai compris: le signal  $\cos[2\pi F1 n]$  (pour  $F1$  donnée) n'est pas périodique.  $F1.n$  peut être entier.*  
 -----  $F1$  ne doit pas être rationnelle  
 oui... D'après ce que j'ai compris: la fréquence  $F1$  n'est pas rationnelle.  
*le signal  $\cos[2\pi F1 n]$  vient de l'échantillonnage à la fréquence  $f_e$  d'une sinusoïde continue  $\cos(2\pi f_o t)$ .*  
 ----- certes...  
 Attendez... D'après ce que vous m'avez dit:  
 -  $\cos[2\pi F1 n]$  vient de l'échantillonnage de  $\cos(2\pi f_o t)$   
 - la fréquence  $F1$  n'est pas rationnelle  
 Je m'avoue quelque peu surpris !  
 ----- pourquoi?  
 Mais cela pose un problème... Si on considère que la fréquence d'échantillonnage  $f_e$  est multiple de  $f_o$  ...  
 ----- non,  $f_e$  n'est pas multiple de  $f_o$   
 D'accord...

L'extrait suivant provient d'un cours de communications numériques. Il s'agit de définir un type de codage pour transmettre une page dactylographiée. Dans ce dialogue (inspiré d'une interaction réelle), l'étudiant choisit de faire du TELETEXTE. Il aurait pu choisir un codage par pixels de type FAX. Dans les deux cas, il doit rester cohérent. C'est d'ailleurs l'exigence de cohérence qui permet au dialogue de progresser.

*est-ce qu'un texte dactylographié peut constituer un message numérique?*  
 ----- oui  
*D'après ce que j'ai compris: la page de texte peut être considérée comme un message numérique. il faut choisir des symboles pour décrire chaque détail de la page*  
 ----- les caractères  
*oui...mmm...D'après ce que j'ai compris: les points de l'image ne sont pas pris comme symboles. que pourrait-on coder dans la page, en dehors des caractères typographiques?*  
 ----- s'il y avait des figures ?  
*D'après vous, on veut coder les figures (schémas, photos, signature).*  
*C'est surprenant, dans la mesure où l'on suppose que les points de l'image ne sont pas pris comme symboles!*

----- mais c'est une page dactylographiée !  
 D'après ce que j'ai compris: on réalise un codage de type TELETEXTE.  
 il n'y a pas que les caractères typographiques à coder...  
 ----- non  
 mais encore ?...  
 ----- la mise en page, les marges  
 oui...mmm... Attendez... D'après ce que vous m'avez dit:  
 - on veut coder les positionnements (paragraphes, etc...)  
 Je m'avoue quelque peu surpris !  
 ----- il faut des codes spéciaux  
 Ah bon!... D'accord...

La connaissance utilisée pour cet extrait est donnée ci-dessous :

*règle 1 (1):*

les points de l'image sont pris comme symboles  
 les ordres de traitement de texte font partie des symboles  
 on ne réalise pas un codage mixte (texte + images)

*règle 2 (1):*

la page de texte ne peut pas être considérée comme un message numérique  
 les points de l'image sont pris comme symboles

*règle 3 (1):*

la page de texte ne peut pas être considérée comme un message numérique  
 les caractères typographiques sont pris comme symboles

*règle 4 (1):*

on réalise un codage de type FAX (télécopie)  
 les points de l'image ne sont pas pris comme symboles

*règle 5 (0):*

On réalise un codage de type TELETEXTE  
 on veut coder les attributs des caractères (police, taille, lignage,...)  
 les ordres de traitement de texte ne font pas partie des symboles

*règle 6 (1):*

on ne réalise pas un codage de type TELETEXTE  
 les caractères typographiques sont pris comme symboles

*règle 7 (0):*

la page de texte peut être considérée comme un message numérique  
 on ne réalise pas un codage de type FAX (télécopie)  
 on ne réalise pas un codage de type TELETEXTE  
 on ne réalise pas un codage mixte (texte + images)

*règle 8 (0):*

la page de texte ne peut pas être considérée comme un message numérique

*règle 10 (0):*

on veut coder les figures (schémas, photos, signature)  
 les points de l'image ne sont pas pris comme symboles

*règle 11 (1):*

on réalise un codage de type FAX (télécopie)  
 on réalise un codage de type TELETEXTE  
 on ne réalise pas un codage mixte (texte + images)

*règle 12 (0):*

on réalise un codage mixte (texte + images)

*règle 13 (1):*

on réalise un codage de type TELETEXTE  
 les caractères typographiques ne sont pas pris comme symboles

*règle 14 (1):*

on ne réalise pas un codage de type FAX (télécopie)  
 les points de l'image sont pris comme symboles

*règle 19 (0):*

on réalise un codage de type TELETEXTE  
 on veut coder les positionnements (paragraphes, etc...)  
 les ordres de traitement de texte ne font pas partie des symboles

L'extrait suivant fait aussi partie d'un cours de Communications Numériques. Il traite de l'interférence entre symboles (IES). L'élève finit par réaliser que l'on peut annuler l'IES (chevauchement des impulsions à

l'instant de l'échantillonnage) sans exiger que les impulsions ne se chevauchent pas du tout. Noter la surprise prématurée manifestée à la fin de l'extrait par le système (les réponses entre crochets indiquent une réponse à un choix fermé).

*Les termes qui provoquent l'IES:  $p((n-k)T)$ , peuvent-ils être tous nuls (pour  $n$  différent de  $k$ ) ?*  
 ---- *oui, je crois*  
*D'après ce que j'ai compris: on ne modifie pas les paramètres de l'émetteur.*  
*On peut peut-être jouer sur la durée des impulsions MIA, pour résoudre ce problème d'IES ?*  
 ---- *on peut raccourcir les impulsions*  
*D'après vous, on choisit des impulsions  $p(t)$  plus courtes.*  
*C'est surprenant, dans la mesure où l'on suppose que on ne modifie pas les paramètres de l'émetteur !*  
 ---- *pourquoi ne peut-on pas changer l'émetteur?*  
*On ne modifie pas les paramètres de l'émetteur parce que cela fait partie des données du problème.*  
*D'après vous, on choisit des impulsions  $p(t)$  plus courtes. C'est surprenant, dans la mesure où l'on suppose que on ne modifie pas les paramètres de l'émetteur!*  
 ---- *on ne les prend pas plus courtes*  
*D'après ce que j'ai compris: les termes d'IES:  $p((k-n)T)$  sont nuls (avec  $k \neq n$ ).*  
*On suppose au départ que les impulsions se chevauchent. Comptez-vous supprimer ce chevauchement ?*  
 ---- *il faut écarter les impulsions*  
*Oui... mmm... D'après vous, on ralentit l'émission des symboles. C'est surprenant, dans la mesure où l'on suppose que on ne modifie pas les paramètres de l'émetteur!*  
 ---- *Ok*  
*Mais cela pose un problème... D'après ce que j'ai compris:*  
 - *on ralentit l'émission des symboles*  
 - *on ne modifie pas les paramètres de l'émetteur*  
*Que proposez-vous comme solution ?*  
 ---- *[ on ne ralentit pas l'émission des symboles ]*  
*Attendez... D'après ce que vous m'avez dit:*  
 - *les termes d'IES:  $p((k-n)T)$  sont nuls (avec  $k \neq n$ )*  
 - *on ne ralentit pas l'émission des symboles*  
*Je m'avoue quelque peu surpris !*  
 ---- *on peut changer la forme des impulsions*  
*Ah bon!... Attendez... D'après ce que vous m'avez dit:*  
 - *on change la forme  $p(t)$  de l'impulsion pour annuler l'IES*  
 - *on ne modifie pas les paramètres de l'émetteur*  
*Je m'avoue quelque peu surpris !*  
 ---- *si on filtre, ça change bien leur forme*  
*Ah bon!... D'accord...*

Le dialogue suivant met en scène un étudiant qui aurait des conceptions fortement erronées sur la notion de filtrage. Noter que le système n'émet jamais de jugement de valeur. Il se contente de montrer l'absurdité de certaines réponses (e.g. "passe-haut").

*si vous considérez la suspension de votre voiture comme un filtre, quel genre de filtre est-ce ?*  
 ---- *passe-haut*  
*D'après vous, votre suspension est un filtre passe-haut. C'est surprenant, dans la mesure où l'on suppose que le système de suspension est transparent aux très basses fréquences!*

----- il n'est pas transparent aux basses fréquences

D'après vous, le système de suspension n'est pas transparent aux très basses fréquences. C'est surprenant, dans la mesure où l'on suppose que la voiture suit les variations lentes du terrain!

----- ah oui

Mais cela pose un problème... D'après ce que j'ai compris:

- la voiture suit les variations lentes du terrain

- le système de suspension n'est pas transparent aux très basses fréquences

Que proposez-vous comme solution ?

----- [ le système de suspension est transparent aux très basses fréquences ]

D'après vous, votre suspension est un filtre passe-haut. C'est surprenant, dans la mesure où l'on suppose que le système de suspension est transparent aux très basses fréquences!

----- c'est un passe-bande

Donc, vous considérez maintenant que votre suspension n'est pas un filtre passe-haut. Est-ce bien le cas?

----- [ votre suspension n'est pas un filtre passe-haut ]

D'après ce que j'ai compris: votre suspension peut entrer en résonance. votre suspension n'est pas en très bon état, n'est-ce pas ?

----- elle n'entre pas en résonance

Oui... mmm... D'après vous, votre suspension ne peut pas entrer en résonance. C'est surprenant, dans la mesure où l'on suppose que votre suspension est un filtre passe-bande!

----- Ok, ce n'est pas un passe-bande

D'après ce que j'ai compris: votre suspension est un filtre passe-bas. considérez-vous votre suspension comme "dure" ?

----- non

D'accord...

## Annexe 2 : Le moteur de clauses paradoxales

### 1. Principe du démonstrateur

Il n'y a pas de conversation possible sans raisonnement. Dans les applications comme SAVANT3 et PARADISE, ces raisonnements sont effectués sur des clauses négatives, que nous appelons clauses paradoxales. En effet, tout ensemble de formules logiques peut être mis sous la forme d'une conjonction de disjonctions. Nous écrivons les disjonctions comme des conjonctions fausses, ( $a$  **ou**  $b$ ) devenant [**non**  $a$  & **non**  $b$ ]  $\Rightarrow F$ .  $F$  désigne une proposition toujours fausse. La connaissance devient ainsi une conjonction de clauses négatives (ou clauses paradoxales) de ce type.

Nous devons maintenant expliquer comment sont menées les démonstrations. Le moteur que nous décrivons maintenant fonctionne en ordre 0 dans SAVANT3, et en ordre 1 dans PARADISE. Nous le décrivons avec des propositions pour alléger l'écriture.

Pour démontrer un terme  $p$ , la stratégie adoptée est la suivante :

- soit à prouver  $p$
- rechercher **non**  $p$  dans une clause ne contenant aucun terme (ni son contraire) qui soit en cours de démonstration (hormis bien entendu **non**  $p$ )
- **non**  $p$  trouvé dans la clause  $C$
- prouver les autres termes de  $C$  s'il y en a

Insistons sur les avantages et les inconvénients d'un tel moteur par rapport à un moteur classique comme celui de Prolog.

Dans un ensemble de clauses de Horn comme celles d'un programme Prolog, les termes démontrables sont désignés à l'avance, et sont dans la "tête" de la clause. De plus, ces termes ne peuvent pas être niés. Ceci entraîne certaines limitations. Par exemple, à partir de la clause Prolog :

$$a \leftarrow b, c$$

le moteur Prolog ne peut pas établir **non**  $b$  à partir de la connaissance de **non**  $a$  et de  $c$ , alors que c'est logiquement possible. En revanche, à partir de la clause paradoxale :

$$[ \text{non } a \ \& \ b \ \& \ c ] \Rightarrow F$$

notre moteur peut conclure à la négation de chacun des trois termes à partir de la connaissance des deux autres.

Bien entendu, ces possibilités plus étendues, dont nous ne pouvons pas nous passer dans la mesure où nous voulons analyser et reproduire les raisonnements conversationnels, entraînent un temps de réponse plus long. Le démonstrateur doit en effet essayer un nombre important de combinaisons. Nous avons donc dû consacrer nos efforts à diminuer le temps de réponse.

### 2. Optimisation

Nous avons implanté et optimisé ce moteur, de manière à ce qu'il fonctionne efficacement tant en ordre 0 qu'en ordre 1. L'optimisation consiste principalement à mémoriser les résultats partiels. Elle est compliquée par le fait qu'il existe plusieurs contextes de démonstration, car les raisonnements peuvent utiliser, ou ne pas utiliser :

- les clauses marquées comme évidentes
- les clauses avec la modalité "indésirable" (pour certaines applications)

- les déclarations de l'utilisateur
- des hypothèses "essayées" par le programme utilisant le moteur

### 3. Monde ouvert

Une caractéristique essentielle de notre moteur est qu'il fonctionne en monde ouvert. Un terme  $p$  peut être vrai parce que démontrable, ou faux parce que **non**  $p$  est démontrable, ou encore indéterminé parce que ni  $p$  ni **non**  $p$  ne sont démontrables. Cette propriété est absolument nécessaire dès qu'il s'agit de reproduire les raisonnements utilisés dans les conversations.

### 4. Révision

Le moteur peut prendre en compte les déclarations de l'utilisateur dans ses raisonnements, à condition que ces déclarations soient compatibles avec la connaissance antérieure au moment où elles sont affirmées. Lorsque ce n'est pas le cas, le programme mémorise une erreur que l'application doit gérer (par exemple SAVANT3 "s'étonne" lorsqu'une erreur est commise).

Lorsque l'utilisateur fait une déclaration qui est l'exacte négation d'une de ses déclarations antérieures, le programme ne constate pas une erreur. Il révisé la connaissance apprise, et retire de sa mémoire tous les résultats partiels établis avec les déclarations de l'utilisateur.

### 5. Complétude

Ce moteur présente un inconvénient : il n'est pas complet. En d'autres termes, il peut arriver que certains termes soient vrais (et logiquement démontrables) sans que le moteur soit capable de les valider. Considérons la connaissance suivante :

$[ a \ \& \ \text{non } b ] \Rightarrow \mathbf{F}$

$[ \text{non } a \ \& \ c ] \Rightarrow \mathbf{F}$

$[ b \ \& \ c ] \Rightarrow \mathbf{F}$

Il est possible de déduire **non**  $c$  de cette connaissance : si  $a$  est vrai, alors  $b$  est vrai (première clause), ce qui entraîne **non**  $c$  (troisième clause) ; si **non**  $a$  est vrai, alors **non**  $c$  est encore vrai (deuxième clause). Or même dans le monde ouvert dans lequel nous nous plaçons, on peut assurer que soit  $a$ , soit **non**  $a$  est vrai. Donc **non**  $c$  est vrai.

Pourtant, le moteur que nous utilisons dans nos applications est incapable d'établir **non**  $c$ . Il va adopter la démarche suivante :

- soit à prouver **non**  $c$
- rechercher  $c$  dans une clause
- $c$  trouvé dans la deuxième clause
- prouver **non**  $a$
- rechercher  $a$  dans une clause ne contenant pas  $c$ , **non**  $c$
- $a$  trouvé dans la première clause
- prouver **non**  $b$
- rechercher  $b$  dans une clause ne contenant pas  $c$ , **non**  $c$ ,  $a$ , **non**  $a$
- échec
- $c$  trouvé dans la troisième clause (backtracking)
- prouver  $b$
- rechercher **non**  $b$  dans une clause ne contenant pas  $c$ , **non**  $c$
- **non**  $b$  trouvé dans la première clause
- prouver  $a$
- rechercher **non**  $a$  dans une clause ne contenant pas  $c$ , **non**  $c$ ,  $b$ , **non**  $b$
- échec

La solution consiste à adopter une stratégie de preuve par réfutation. Nous avons développé un nouveau moteur, dont il est facile de montrer qu'il est complet, puisqu'il implante directement une stratégie de résolution correspondant à la technique de la résolvente [Haton et al. 1992]. La stratégie peut être décrite de la manière suivante :

- soit à prouver  $p$
- ajouter la règle  $p \Rightarrow F$  à la base de connaissances
- tenter de saturer une règle en prouvant tous ses termes
- en cas de "backtracking", ôter les règles qui ont été ajoutées

Cette stratégie est d'ailleurs tout à fait dans l'esprit conversationnel. Une proposition  $p$  est vraie parce que son contraire **non**  $p$  provoquerait une incohérence.

Nous avons implanté ce moteur, et commencé à l'optimiser (y compris en ordre 1). Mais nous ne l'avons pas encore intégré à nos applications. Il est à noter que nous n'avons jamais pu constater, dans les dialogues, un quelconque défaut qui serait dû à l'incomplétude de notre premier moteur.



# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

Dans ce chapitre, nous décrivons brièvement les éléments d'un système auteur que nous avons développé pour faciliter la création de cours destinés à SAVANT3. Nous présentons en particulier le système ATOL qui a pour rôle d'éliciter les connaissances de l'auteur. ATOL repose sur des principes conversationnels. Il nécessite encore des développements, car le type de dialogue qu'il offre n'est pas suffisamment acceptable. Néanmoins, nous croyons qu'il repose sur une démarche viable, et il est intéressant de le décrire ici en tant qu'application de notre modèle conversationnel.

Nous commençons par décrire l'environnement de SAVANT3 tel qu'il est présenté à l'élève. Ensuite nous décrivons les différents outils, destinés cette fois à l'auteur, qui rendent possible la création de ce type de cours.

## 2. Un cours basé sur SAVANT3

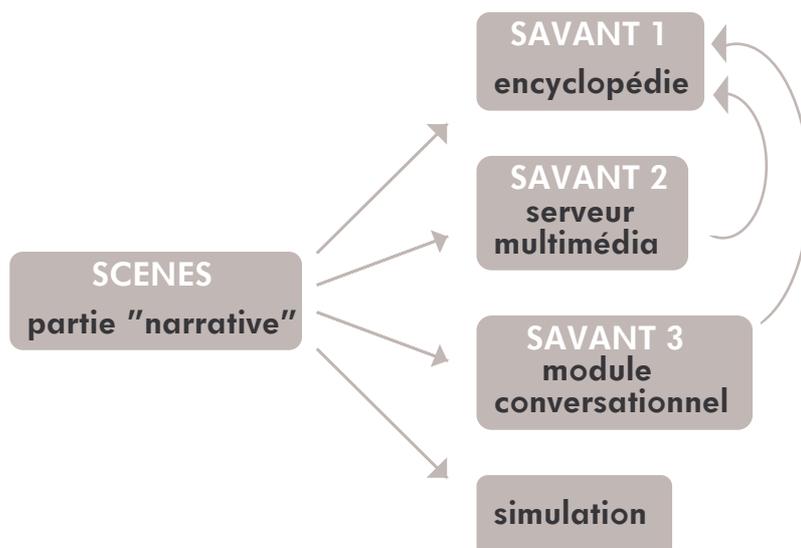
Les dialogues de type SAVANT3 (voir chapitre III.2) doivent intervenir dans des situations dont le contexte est parfaitement contrôlé. Leur succès, comme nous l'avons indiqué, exige avant tout que l'étudiant ait une certaine connaissance des concepts auxquels SAVANT3 fait référence, et connaisse de plus la plupart des liens logiques que le programme utilise. Ces conditions sont en général remplies dans les cursus traditionnels : l'auteur d'un livre de cours fait les mêmes suppositions lorsqu'il conçoit les exercices de la fin d'un chapitre.

C'est pourquoi les dialogues que nous proposons aux étudiants interviennent *en contexte*. Le contexte est construit selon un "scénario" pédagogique qui fait intervenir des moyens divers :

- une présentation "narrative" des concepts du cours (programme SCENES)
- une encyclopédie sur les concepts du cours (encyclopédie ou serveur SAVANT1)
- des dialogues conversationnels (programme SAVANT3)
- des simulations
- des ressources multimédia (serveur multimédia SAVANT2)

Les systèmes SAVANT1 et SAVANT2 ont été évoqués dans le chapitre III.1. Ils sont décrits par exemple dans [Dessalles 1992g].

Nous présentons plus loin des exemples d'écrans extraits d'un cours sur les systèmes linéaires. On y voit un écran de la présentation "narrative", un écran d'encyclopédie, et un écran de la simulation.



A part SAVANT3 qui est écrit en Prolog, tous ces programmes ont été développés en langage C. Dans certaines applications, les simulations sont en Pascal.

Le contenu de l'encyclopédie est conçu selon des règles strictes. Ces règles sont explicitées dans [Dessalles 1984].

Le développement d'une partie "narrative" sur un sujet donné obéit lui aussi à des contraintes que nous avons voulues systématiques. La principale consiste à toujours partir d'une situation paradoxale ou indésirable, en accord avec les principes conversationnels. Dans l'exemple qui sert d'illustration, l'auteur attire l'attention de l'étudiant sur le paradoxe apparent lié au fait que deux systèmes qui n'entretiennent aucune ressemblance ont pourtant rigoureusement le même comportement.

L'accès à l'encyclopédie SAVANT1 est généralement contrôlé par l'étudiant. En revanche, c'est le programme SCENES qui démarre les phases de simulation et de dialogue (ou les séquences multimedia SAVANT2 lorsqu'elles sont disponibles). Les dialogues de SAVANT3 peuvent utiliser la situation problématique établie dans la partie "narrative".

**UNE SUSPENSION DE VOITURE**

**UN CIRCUIT "LRC"**

Ces deux systèmes n'ont rien à voir entre eux, mais il se trouve pourtant qu'ils ont strictement le même comportement!

[appuyez sur une touche] (aide: '?')

Trois écrans extraits d'un cours sur les systèmes. L'un provient de la partie "narrative". Il est affiché par le programme SCENES. Les deux autres sont respectivement un écran de SAVANTI et un écran de simulation.

**REPOSE INDICIELLE**

temps

**REPOSE EN FREQUENCE**

Hz

Utilisez les flèches pour choisir LA POSITION pas de ZERO: [ESCI] FIN: [ESCI]

**Pôles et zéros**

La Fonction de transfert d'un système du second ordre peut s'écrire:

$$H(s) = K \frac{s - z}{(s - p_1)(s - p_2)}$$

où z est le "zéro" et p1, p2 les "pôles" de la fonction de transfert.

Les pôles peuvent être réels ou complexes, selon la valeur du coefficient d'amortissement.

[F1] [F2] [F3] [F4] [F5] [F6] [F7] [F8] [F9] [F10]  
 Suite Retour Rép. Guide Annul. Somm. Corr. Envoi Conn. -ok-  
 SAVANT - N.T.E. 1988 ===== Retour au programme: [F10]

### 3. L'environnement auteur de SAVANT3

Le développement de cours comportant des dialogues de type SAVANT3 nécessite un système auteur. Dans notre application, ce système auteur est constitué par un certain nombre d'outils qui sont destinés à aider l'auteur à concevoir des contenus pour les différentes composantes de l'environnement de l'élève (SCENES, SAVANT1, SAVANT2, SAVANT3). L'exploitation de SAVANT2 ayant cessé en 1989, nous ne l'intégrons pas dans cette description. Le tableau suivant donne la liste de ces outils<sup>(1)</sup> :

<i>programmes élève</i>	<i>programmes auteur</i>
SCENES	← outils de traitement d'images graphiques logiciel CRESCENE de création de scénarios "narratifs"
SAVANT 1	← logiciel élémentaire de composition vidéotex logiciel élémentaire de consultation de l'encyclopédie logiciel de connexion à un serveur distant
SAVANT 3	← logiciel NEGALOG de manipulation de base de connaissances logiciel ATOL d'extraction de connaissances

Nous avons développé un certain nombre d'outils pour le traitement des bases de connaissances destinées à SAVANT3. Rappelons en effet que la gestion des dialogues est du seul ressort de SAVANT3. L'auteur se contente de lui fournir la connaissance du domaine, ce qui représente généralement un ensemble de 10 à 20 règles. Le logiciel NEGALOG permet à l'auteur d'exprimer cette connaissance, sous forme de clauses négatives, par l'intermédiaire d'un certain nombre d'opérations :

- énoncer des propositions
- lier les propositions (ou leur négation) par des relations d'incompatibilité
- associer des questions (ou des *prompts*) à des propositions
- associer des mots clés à des propositions (avec un système de création de mots clés par défaut)
- créer des questions fermées

Par ailleurs, un logiciel a été conçu pour estimer le nombre de possibilités compatibles avec les règles d'une base de connaissances. En effet, un dialogue fructueux ne peut s'établir avec l'étudiant que si la base contient suffisamment de contraintes (incompatibilités) pour limiter l'univers des possibles et amener l'étudiant à entrevoir certaines incohérences (avérées ou potentielles). Mais l'ensemble des possibles ne doit pas être trop restreint, sinon l'étudiant n'a plus aucun choix et se sent contraint.

Nous décrivons maintenant le programme ATOL qui a pour fonction d'aider l'auteur à donner la connaissance du domaine d'une manière que l'on souhaite naturelle.

<sup>1</sup> Dans sa version initiale en tant que serveur vidéotex, SAVANT1 reposait sur un logiciel élaboré de consultation conçu par J-P Bach, du Centre de Calcul de TELECOM-Paris. Nous présentons ici la version "restreinte" de SAVANT1 telle qu'elle est couplée à SAVANT3 dans la version autonome sur PC.

## 4. Un système conversationnel pour l'auteur

### 4.1. Les limites de NEGALOG

Le développement d'une base de connaissances à l'aide du programme NEGALOG est relativement aisé si l'auteur a une idée précise des relations logiques qui lient les concepts du domaine. Il peut ainsi, en deux heures, créer un dialogue opérationnel. Toutefois, le dialogue conduit par SAVANT3 ne sera acceptable que sous un certain nombre de conditions :

- Les règles doivent être correctes.
- Les règles du contexte doivent toutes être présentes.
- Il est généralement préférable que les règles ne comportent pas de termes redondants. Par exemple,  $q$  est redondant dans la règle  $[p \ \& \ q \ \& \ r] \Rightarrow F$  si la règle  $[\text{non } q \ \& \ r] \Rightarrow F$  est présente dans la base de connaissances (voir définition de la redondance p. 222). L'argumentation de SAVANT3 en souffrira : SAVANT3 ne pourra pas utiliser son mécanisme d'étonnement prématuré (voir chap. III.2) si  $p$ , puis  $r$  deviennent vrais, car  $q$  se retrouvera automatiquement vrai. De plus, la présence de  $q$  gêne la perception de l'incohérence en cas de contradiction. Cette recommandation souffre cependant de quelques exceptions (e.g. si l'une des règles est de bon sens et l'autre non).
- Les règles redondantes doivent être supprimées. Si la base contient les trois règles :  $[p \ \& \ \text{non } q] \Rightarrow F$ ,  $[q \ \& \ \text{non } r] \Rightarrow F$ ,  $[p \ \& \ \text{non } r] \Rightarrow F$ , la troisième règle doit être ôtée. Sinon SAVANT3 risque de s'étonner deux fois "pour la même raison" : il peut s'étonner de  $p$  en utilisant la troisième règle, puis apprendre que  $r$  est vrai ; ensuite il s'étonne à nouveau de  $p$  en raison de la première règle, et l'étudiant a l'impression, en disant " $q$  vrai", de redonner le même argument puisque pour lui  $r$  est vrai par  $q$  (deuxième règle). Par ailleurs, les règles redondantes ralentissent le démonstrateur.
- Les auteurs débutants utilisent mal NEGALOG, en particulier parce qu'ils "cachent" de la logique dans leurs propositions. Des phrases comme :
  - "une fréquence de coupure inférieure à 100 Hz empêche les oscillations"
  - "la fréquence de coupure doit être inférieure à 100 Hz"
  - "pour supprimer les oscillations, il faut diminuer la fréquence de coupure"
  - "la réponse n'est atténuée que pour une fréquence élevée"
  - "H(z) possède des zéros à droite et pas de pôles à droite"

ne correspondent pas à des propositions élémentaires et ne peuvent pas être correctement manipulées par SAVANT3. Les quatre premières contiennent une implication, la dernière une conjonction. Il est par exemple impossible, pour SAVANT3, de nier correctement la dernière phrase. Les implications cachées dans les premières phrases sont inconnues du programme, qui risque ainsi de se contredire, ou d'exprimer des relations "niaises" qui traduisent un simple *modus ponens* (e.g. "il est surprenant que  $p$  et que  $q$  puisque  $p$  empêche  $q$ ", ou encore " $p$  et  $q$  sont vrais ; mais peut-être que  $p$  n'empêche pas  $q$  ?").

Pour toutes ces raisons, nous avons envisagé la création d'un outil qui puisse aider l'auteur à constituer facilement une base de connaissances acceptable, même s'il n'est pas logicien. Nos premiers efforts dans ce sens ont conduit à la création du programme ATOL.

### 4.2. L'élicitation des connaissances par le programme ATOL

Le programme ATOL est destiné à l'auteur. Il a pour but de lui permettre d'exprimer la connaissance du domaine par l'intermédiaire d'un dialogue, et de lui éviter ainsi d'avoir à manipuler directement le formalisme logique. ATOL reposait au départ sur un présupposé analogue à celui qui a présidé au développement de SAVANT3. L'étudiant, face à SAVANT3, est supposé être particulièrement sensible à ses

propres incohérences. L'idée est en gros la même lorsqu'il s'agit de gérer un dialogue avec l'auteur : jouer sur le couple incohérence / explication. Toutefois, dans sa version actuelle, ATOL joue davantage sur le sentiment d'*universalité*. Nous allons voir en quoi réside la différence d'approche.

#### 4.2.1. Vérification d'une règle

Pour vérifier la règle  $[p_1 \& p_2 \& \dots \& p_n] \Rightarrow F$ , ATOL propose le dialogue suivant :

- ATOL : "Sachant que  $p_2, p_3 \dots$  et  $p_n$ , il est absolument impossible que  $p_1$ "
- auteur : [possible lorsque *non*  $q$ ]
- ATOL : "Sachant que  $p_2, p_3, \dots p_n$  et  $q$ , il est absolument impossible que  $p_1$ "

etc.

Dans SAVANT3, *non*  $q$  apparaîtrait comme l'*explication* d'un étonnement lié à la cooccurrence de  $p_1, p_2, \dots p_n$ . Dans le dialogue tel qu'il est géré par ATOL, il n'y a pas d'étonnement. C'est donc de l'*universalité* d'un lien logique que l'auteur doit juger, et *non*  $q$  sera plutôt considérée par lui comme une *exception*.

Pour des raisons de simplicité linguistique, la première phrase est transformée lorsque  $p_1$  comporte une négation grammaticale (par ex. " $F_1$  n'est pas rationnelle") :

- "Sachant que  $p_2, p_3, \dots p_n$ , il est absolument certain que *non*  $p_1$ "

(dans ce cas *non*  $p_1$  ne comporte pas de négation grammaticale).

Lorsque l'auteur ne trouve pas d'exception, la règle est ajoutée à la base de connaissances, mais seulement après des tests de cohérence et de non-redondance.

#### 4.2.2. Test de cohérence et de non-redondance

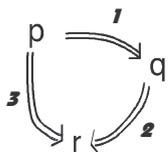
Chaque fois qu'ATOL mémorise une nouvelle règle, il teste la cohérence et la non-redondance de la nouvelle base de connaissances obtenue :

**test d'incohérence** : la base est incohérente si et seulement si il existe une règle dont tous les termes peuvent être prouvés avec les autres règles de la base (dans ce cas cette règle constitue l'incohérence).

**test de redondance de règle** : une règle est redondante si l'un de ses termes peut être infirmé à l'aide des autres termes de la règle et des autres règles de la base (dans ce cas la règle est superflue)

**test de redondance de terme** : un terme est redondant dans une règle s'il peut être prouvé à l'aide des autres termes de la règle et des autres règles de la base (dans ce cas le terme est superflu)

L'auteur est simplement prévenu de la redondance des termes superflus. En revanche, les règles redondantes sont tout simplement ôtées par ATOL. De plus, ATOL évite de proposer à la vérification (et donc potentiellement à l'enregistrement) des règles qui seraient redondantes ou en contradiction avec la connaissance déjà enregistrée. Mais le plus souvent, c'est l'enregistrement d'une nouvelle règle qui rend une ancienne règle redondante. Si l'on reprend l'exemple des trois règles  $[p \& \text{non } q] \Rightarrow F$ ,  $[q \& \text{non } r] \Rightarrow F$ ,  $[p \& \text{non } r] \Rightarrow F$ , on comprend que l'adjonction de la deuxième règle à une base contenant les deux autres



rend la troisième redondante. En effet, si l'on applique le critère de redondance de règle, le terme *non*  $r$  de la troisième règle peut être infirmé par la connaissance de  $p$  et des deux premières règles : la première établit  $q$  à partir de  $p$ , et la deuxième prouve  $r$  à partir de  $q$ . Cette situation de redondance n'existait pas avant l'adjonction de la deuxième règle.

### 4.2.3. Complétude de la base de connaissances

Lorsqu'une exception *non q* est donnée à la règle  $[p_1 \& p_2 \& \dots \& p_n] \Rightarrow F$ , ATOL soumet de nouvelles règles :

- $[p_1 \& p_2 \& \dots \& p_n \& q] \Rightarrow F$  (validité de la règle hors exception)
- $[p_1 \& p_2 \& \dots \& p_n \& \text{non } q] \Rightarrow F$  ("sauvetage" dans le cas de l'exception)
- $[\text{non } p_1 \& \text{non } q] \Rightarrow F$  (nécessité de *q*)
- $[\text{non } q] \Rightarrow F$  (validité de l'argument *q*)
- $[q] \Rightarrow F$  (validité de l'exception *non q*)

Cette stratégie conduit au type de dialogue suivant :

auteur : Arthur est en déplacement  
 ATOL : "il est absolument certain que Arthur est en déplacement"  

Vrai	Faux lorsque ...	Dépend de la situation	Abandon
------	------------------	------------------------	---------

 auteur : Arthur est au bureau  
 ATOL : "Sachant que Arthur n'est pas au bureau, il est absolument certain que Arthur est en déplacement"  

Vrai	Faux lorsque ...	Dépend de la situation	Abandon
------	------------------	------------------------	---------

 auteur : Arthur est en vacances  
 atol : "Sachant que Arthur n'est pas au bureau, et que Arthur n'est pas en vacances, il est absolument certain que Arthur est en déplacement"  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 ATOL : [enregistrement de la règle]  
 ATOL : "Sachant que Arthur n'est pas au bureau, et que Arthur est en vacances, il est absolument certain que Arthur est en déplacement"  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 ATOL : "Sachant que Arthur est en déplacement, alors il est absolument impossible que Arthur est en vacances"  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 ATOL : [enregistrement de la règle]  
 ATOL : On sait déjà qu'il est absolument certain que Arthur est en vacances... à condition d'admettre que  
     - Arthur n'est pas en déplacement  
     - Arthur n'est pas au bureau  
 Entrez un nouvel élément qui permet d'affirmer qu'il est absolument certain que Arthur est en vacances  
 auteur : []  
 ATOL : on sait déjà qu'il est absolument impossible que Arthur est en vacances ... à condition d'admettre que Arthur est en déplacement  
 Entrez un nouvel élément qui permet d'affirmer qu'il est absolument impossible que Arthur est en vacances  
 auteur : Arthur est au bureau  
 ATOL : "Sachant que Arthur est au bureau, alors il est absolument impossible que Arthur est en vacances"  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 ATOL : [enregistrement de la règle]  
 ATOL : Sachant que Arthur n'est pas au bureau, alors il est absolument impossible que Arthur est en vacances  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 [...]
 atol : Sachant que Arthur est au bureau, alors il est absolument certain que Arthur est en déplacement  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 ATOL : Sachant que Arthur est en déplacement, alors il est absolument impossible que Arthur est au bureau  

Vrai	Faux lorsque...	Dépend de la situation	Abandon
------	-----------------	------------------------	---------

 ATOL : [enregistrement de la règle]  
 [...]

Au total, au cours de cette interaction, ATOL enregistre quatre règles. Cela suffit déjà pour que SAVANT3 démarre une interaction :

SAVANT3 : *Arthur n'est pas au bureau ?*  
 utilisateur : *non*  
 SAVANT3 : *D'après ce que j'ai compris: Arthur n'est pas au bureau. Arthur est en déplacement ?*  
 utilisateur : *non*  
 savant3 : *Attendez... D'après ce que vous m'avez dit:*  
     - *Arthur n'est pas en déplacement*  
     - *Arthur n'est pas au bureau*  
     *Je m'avoue quelque peu surpris !*  
 utilisateur : *Il est parti en vacances*  
 SAVANT3 : *Ah bon !... D'accord*

### 4.3. Les limites de ATOL

Le dialogue proposé par ATOL est encore fastidieux pour l'auteur. Il y a un équilibre à respecter entre le nombre de suggestions faites par ATOL et le nombre de règles que l'auteur risque d'oublier. Ce programme doit donc faire l'objet de recherches visant à épurer sa stratégie et à lui donner les moyens de détecter les questions superflues. Nous avons toutefois jugé qu'il était utile de présenter ATOL, d'une part parce qu'il constitue un essai original pour doter SAVANT3 d'un système auteur potentiellement "intelligent", et d'autre part parce qu'il repose lui aussi sur des principes issus de la modélisation des conversations.

## 5. Conclusion

Les outils destinés au développement de cours pour SAVANT3 sont tous originaux. Nous avons mentionné ici les principaux. Certains d'entre eux sont encore au stade expérimental. C'est le cas d'ATOL. Dans l'état actuel des choses, la création d'un cours pour SAVANT3 est relativement aisée, si l'on compare aux autres systèmes d'EIAO. Nous pensons néanmoins que des progrès sont encore possibles pour faciliter le travail de l'auteur. D'autre part, l'augmentation des potentialités de SAVANT3, avec notamment l'intégration de fonctionnalités *critiques* (voir chapitre III.4) nécessitera la définition de nouveaux outils pour l'auteur. Nous poursuivons les recherches dans ces directions.

# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

L'utilisation des systèmes à base de connaissances (SBC)<sup>(1)</sup> rencontre des problèmes qui semblent liés à des limitations non des systèmes eux-mêmes, mais de leurs utilisateurs ! En effet, les SBC ont souvent la capacité théorique de produire des raisonnements qui pourraient être pertinents pour leur utilisateur à un moment donné. Malheureusement, soit le raisonnement pertinent n'est pas produit, parce que l'utilisateur ne sait pas exprimer son besoin dans des termes assimilables par le système, soit le raisonnement pertinent est effectivement produit, mais l'utilisateur est incapable de le comprendre à travers le résultat que lui fournit le système.

Il s'agit là d'un problème d'*explication*. Lorsque les utilisateurs font appel au SBC, ils expliquent (= exposent) leur problème d'une façon qui n'est pas nécessairement adaptée à ce que le système est capable de reconnaître. De même, lorsque le SBC leur fournit la trace d'un raisonnement (enchaînement de règles), ils ne sont pas en mesure de reconnaître l'explication (= solution) de leur problème.

Il est à peu près clair maintenant que la meilleure acceptabilité des SBC, qui est fortement attendue, passera par leur adaptation à l'utilisateur humain, puisque l'adaptation inverse semble impossible. Mais, comme nous le verrons, les solutions proposées divergent quelque peu.

Puisqu'il s'agit d'amener les SBC à communiquer d'une manière humainement acceptable, il est légitime de penser que notre étude de certaines des contraintes cognitives qui régissent les interactions spontanées entre humains doit pouvoir apporter un nouvel éclairage sur ce problème complexe d'interaction humain / machine.

Nous suggérons ainsi, comme d'autres, le fait que cette interaction doit prendre la forme d'un dialogue. Notre originalité consiste à dire que ce dialogue doit ressembler à une *conversation*, au sens restreint de ce terme, c.à.d. qu'elle doit reposer sur une *argumentation critique*.

A cet égard, nous nous rapprochons de certains travaux qui visent à concevoir des "Systèmes Experts Critiques". De tels systèmes ont pour fonction de critiquer des éléments de solution fournis par l'utilisateur, plutôt que de proposer des solutions globales peu assimilables. Par rapport à ces recherches, notre originalité consiste à proposer les critiques sous forme d'une argumentation.

Nous nous situons donc à l'intersection de deux courants : celui qui veut amener les SBC à gérer l'*explication* sous la forme d'un dialogue, et celui qui conçoit le rôle du SBC comme celui d'un *critique*. Notre contribution tire son originalité de la compréhension des phénomènes d'explication dans les relations conversationnelles.

---

<sup>1</sup> Cette dénomination tend à remplacer la dénomination, plus restrictive et souvent mal perçue, de *système expert*.

## 2. La problématique de l'explication dans les SBC

### 2.1. *Ajouter des connaissances spécifiques pour la génération d'explications*

Face à l'incapacité des systèmes à base de connaissances à *expliquer* leurs raisonnements de manière à les faire accepter, de nombreux auteurs ont incriminé l'absence dans le système de certaines connaissances nécessaires à la production d'explications. Les SBC disposent bien entendu de connaissances qui leur permettent de mener à bien leurs raisonnements, mais on s'est vite aperçu que la simple trace du raisonnement ne constituait pas une explication acceptable. Des auteurs comme W.J. Clancey [1981] et W.R. Swartout [1983] ont insisté sur la nécessité de doter le système de moyens pour analyser son propre raisonnement.

Clancey a doté son système NEOMYCIN de divers types de connaissances qui sont venues s'ajouter aux règles d'inférence du système expert MYCIN, notamment des règles stratégiques capables de justifier le déclenchement préférentiel de certaines règles d'inférence dans certaines situations. Ces règles stratégiques étaient inutiles lors de la production du raisonnement (dans le cas de MYCIN, il s'agissait de déterminer un agent infectieux responsable de maladies comme la méningite), mais elles servaient à expliquer le raisonnement (dans le cas de NEOMYCIN, il s'agissait d'enseigner, cf. [Clancey 1987]).

Alors que Clancey a dû adjoindre *a posteriori* les connaissances nécessaires à la génération d'explications, Swartout juge quant à lui plus "économique" de ne pas *perdre* ces informations. En effet, les connaissances qui permettront d'expliquer sont là au départ : ce sont les connaissances du domaine, et elles sont explicites lors de la conception du système. Malheureusement, elles sont généralement perdues lors de la "compilation" (pour reprendre un terme de Clancey) de la connaissance sous forme de règles d'inférence. Swartout prend l'exemple de la règle d'inférence suivante :

"Si on observe une baisse de potassium dans le sérum alors que de la digitaline a été administrée, alors baisser la dose de digitaline."

En fait, ce type de règle dérive d'une connaissance plus profonde, qui précise que la baisse du taux de potassium, associée à une dose trop élevée de digitaline, peut entraîner une fibrillation ventriculaire. Cette connaissance dérive elle-même d'un schéma plus abstrait, qui peut s'exprimer ainsi :

modèle : "toxicité d'une drogue" - Si une anomalie (e.g. anomalie de dosage) est associée à une augmentation de la dose d'une drogue, cela peut entraîner un dérèglement dangereux.

En décrivant son système XPLAIN, Swartout insiste sur l'intérêt de conserver toutes ces informations lors de la création du programme expert. Cela permet au système de fournir des explications comme celle-ci :

Please enter the value of serum potassium : **why** ?

The system is anticipating digitalis toxicity. Decreased serum potassium causes increased automaticity, which may cause a change to ventricular fibrillation. Increased digitalis also causes increased automaticity. Thus, if the system observes decreased serum potassium, it reduces the dose of digitalis due to decreased serum potassium.

L'adjonction au SBC de connaissances spécifiques aux besoins de l'explication constitue un progrès significatif. Grâce elles, des systèmes comme NEOMYCIN ou XPLAIN sont capables d'*expliquer* leur comportement en termes de buts et de démarche. Pourtant, quoique nécessaire, cette solution n'est peut-être pas suffisante.

### 2.2. *L'explication émergeant d'un dialogue*

B. Safar [Safar et al. 1992] souligne le fait que dans de nombreuses applications, les représentations internes du système et les détails de sa démarche n'intéressent pas l'utilisateur, et ne constituent donc pas une explication. Nous tournons là le dos à "l'explication-trace" ("*translate-the-code approach*"). De même, C. Paris [1992] constate que les modules d'explication sont généralement ajoutés après coup aux SBC, et fournissent des explications dérivées de motifs *ad hoc*, alors que l'on peut concevoir que la résolution de problème soit indissociable du processus explicatif lui-même : le SBC est alors perçu, non plus comme un

système chargé de résoudre un problème, mais comme un système chargé d'*expliquer* (un problème, une solution, un phénomène). On évolue ainsi inévitablement vers une conception *dialogique* de l'explication.

En effet, l'idée d'une explication dynamique constituant la totalité de l'interaction, par opposition à une phase explicative qui viendrait seulement après une phase de résolution de problème, commence à s'imposer pour plusieurs raisons :

- l'utilisateur est rarement capable d'exprimer ses buts d'une manière précise et complète
- le système ne peut pas satisfaire tous les buts de l'utilisateur
- la satisfaction complète ou incomplète de certains buts peut engendrer de nouveaux buts de la part de l'utilisateur

Ces raisons amènent à concevoir l'explication comme une *négociation*, caractéristique de certains types d'interactions dialogiques [Baker 1992]. Les phases de raisonnements, liées à la satisfaction de buts partiels, sont ainsi mêlées aux phases d'analyse de la requête et de propositions de compromis.

La problématique de l'explication a ainsi complètement bouleversé la conception du rôle du SBC. Perçu au départ comme un résolveur de problème, le système à base de connaissances devient avant tout un système capable de fournir des explications adaptées au besoin de l'utilisateur. Les capacités de raisonnement du SBC ne sont plus mises en avant, elles sont mises au service de l'explication [GENE 1993].

Dans cette conception, c'est tout de même le système qui résout le problème de l'utilisateur, ou un problème approchant, de manière à satisfaire l'utilisateur. Dans l'approche des systèmes critiques, à l'inverse, c'est l'utilisateur qui résout lui-même son problème : le système va simplement l'aider dans sa tâche en émettant des critiques appropriées.

### 2.3. L'explication vue comme une critique des actions de l'utilisateur

La motivation qui a présidé à la création des systèmes critiques trouve aussi son origine dans le souci de changer le rôle du SBC pour qu'il soit mieux accepté. Le système à base de connaissances, utilisé comme un critique, n'est plus chargé de résoudre un problème. Il est engagé dans une relation de *coopération* avec l'utilisateur :

"Cooperative problem-solving systems are examples of human-computer cognitive systems. They serve as cognitive amplifiers of the human. [...] The major difference between classical expert systems, such as MYCIN and R1, and cooperative problem-solving systems involves the roles of the user and computer. Most expert systems ask the user for input, make all decisions and then return an answer. In a cooperative problem-solving system, the user is an active agent empowered by the system's knowledge." [Fischer et al. 1991]

L'utilisateur du système est engagé dans une tâche, et entreprend des actions pour la résoudre. Le système critique observe ces tentatives, et intervient lorsqu'il est en mesure d'émettre une critique utile. Pour cela il évalue les actions de l'utilisateur en fonction des buts poursuivis (but de la tâche ou objectifs analysés comme des sous-buts de l'utilisateur), il envisage des solutions alternatives moins coûteuses, il teste l'éventuelle violation de contraintes par l'action entreprise par l'utilisateur (contraintes de validité et contraintes de préférence). Cette démarche est en accord avec ce qu'indique l'observation des concepteurs (*e.g.* ingénieurs) en situation de résolution de problème :

"L'élaboration de la solution s'effectue en interaction continue avec l'évaluation de la solution. La solution finale est élaborée progressivement, par adjonction d'éléments de solution successifs. Avant d'adopter un élément de solution, le concepteur l'évalue. Cette évaluation aboutit à un constat de respect - ou non - des contraintes fixées. Il en résulte un jugement d'acceptabilité de l'élément de solution et une orientation de l'attention du concepteur vers des éléments de solution plus spécifiques." [Bonnardel 1989].

Cette approche par la critique est donc conçue pour apporter une aide pertinente à l'utilisateur, comme dans le cas des SBC conçus pour l'explication. L'ambition est toutefois en apparence plus limitée. Il ne s'agit pas de fournir une solution complète à l'utilisateur, à la suite d'une éventuelle négociation du problème. Il s'agit simplement de donner ici et là des critiques pertinentes.

Pourtant, on peut voir là l'un des points forts des systèmes critiques. Ils n'ont pas besoin de posséder une expertise complète du domaine. Ils peuvent fonctionner utilement quel que soit leur niveau de compétence, en n'émettant que les critiques qu'ils ont réussi à calculer. Les systèmes critiques peuvent ainsi se contenter d'une expertise partielle, alors que les systèmes "experts" ne fonctionnent correctement que dans des situations où ils sont parfaitement compétents.

Notre observation des conversations et du contenu des répliques émises par les interlocuteurs nous amène à justifier une approche qui combine les avantages des deux courants que nous avons évoqués. En effet, les explications humaines d'une part sont dialogiques et nécessitent parfois une conversation entière, et d'autre part elles prennent la forme d'une critique entre des acteurs ayant un niveau d'expertise comparable sur le sujet de la conversation.

Nous proposons de caractériser l'explication telle quelle apparaît dans les conversations, puis de voir comment elle peut être transposée dans des situations de résolution de problème entre un humain et un système à base de connaissances.

### 3. Anatomie de l'explication conversationnelle

#### 3.1. Définir l'explication

L'explication dans les SBC est un concept qui, à notre avis, regroupe des problématiques différentes. Le mot *explication* a trois sens principaux :

- on peut expliquer un terme obscur (par ex. le mot *rhombe*)
- on peut expliquer par un long développement (par ex. la théorie de la relativité restreinte)
- on peut expliquer un phénomène étonnant (par ex. "*de quoi est-il mort si jeune ?*")

Ces acceptions correspondent toutes trois à des besoins réels des SBC. Toutefois, nous défendons l'idée que ces besoins ne seront pas satisfaits par les mêmes moyens. Nous choisissons de nous intéresser ici exclusivement au troisième sens de l'explication, celui qui est relatif à l'étonnement, et ceci pour trois raisons.

Tout d'abord, c'est celui pour lequel l'étude des conversations nous donne le meilleur éclairage. De plus, il correspond, dans le contexte des SBC, au besoin le plus difficile à satisfaire, alors que des moyens classiques (e.g. hypertextes) semblent pouvoir apporter une aide significative pour l'explication de termes obscurs ou de chapitres entiers de la connaissance (cf. notre expérience avec SAVANT1 [Dessalles 1984]). Enfin, nous pensons qu'il est possible d'utiliser les principes de l'explication basée sur la surprise non seulement pour satisfaire les requêtes ponctuelles des usagers ou pour les aider dans une résolution de tâche, mais aussi pour engendrer des développements plus longs destinés à leur *expliquer* (dans les deux premiers sens du mot) des termes ou des phénomènes (cf. chap. IV.1).

#### 3.2. L'explication après surprise

Le phénomène de l'explication est très fréquent dans les conversations et dans la vie courante. Prenons l'extrait suivant, dans lequel le mot *explication* est explicitement mentionné.

[télévision 21-02-92]

*commentateur A- Formidable Surya Bonaly, qui vient de réaliser, pour la première fois dans l'histoire du patinage, une qua-dru-ple boucle et qui ne s'est pas contentée de cela. Elle a passé [... quatre triples sauts]*

[...] [les notes techniques s'affichent]

*commentateur B- Quelle déception. Quelle déception.*

*commentateur A- Oh la la!*

*commentateur B- Quelle déception!*

*commentateur A- 5.1 de la part de la juge anglaise! Alors là, il y a une chose que l'on a du mal à comprendre.*

[silence 4 secondes]

*commentateur B- L'explication vient peut-être de la réception des sauts.*

[silence 6 secondes]

*commentatrice - de la réception et aussi de la préparation. C'est-à-dire qu'il y a toujours un petit blocage dans sa préparation.*

On perçoit clairement ici comment l'explication vient résoudre un étonnement ("il y a une chose que l'on a du mal à comprendre..."). Cet étonnement vient de la contradiction entre les prouesses techniques réalisées par S. Bonaly et la faiblesse de la note technique octroyée:

[ prouesses\_techniques & note\_technique\_faible ]  $\Rightarrow$  **F**

Mais l'explication (noter le mot "explication" spontanément employé par le deuxième commentateur) est produite ici par une **invalidation indirecte** (cf. chap. I.1) qui infirme une prémisse absente au départ. En effet, dans notre représentation, la formule précédente doit être remplacée après les deux dernières répliques par quelque chose comme:

[ prouesses\_techniques & **non** défaut\_majeur & note\_technique\_faible ]  $\Rightarrow$  **F**

L'incompatibilité de départ n'a donc plus cours, car une prémisse insatisfaite avait été omise. Mentionner *a posteriori* quelque chose qui apparaît comme un défaut majeur, apparaît ainsi pour les commentateurs comme un moyen de résoudre la contradiction.

Le fait que le type d'explication auquel nous nous intéressons soit défini par rapport à une incohérence, source de l'étonnement, fait de ce genre d'explication un phénomène fortement contraint. Si on exprime la contradiction de référence de la manière suivante:

[  $p_1$  &  $p_2$  & ... &  $p_n$  ]  $\Rightarrow$  **F**

alors une explication  $q$  est admissible si et seulement si:

- soit  $q = \mathbf{non} p_i$ , ou bien  $q$  permet de démontrer **non**  $p_i$  (invalidation directe)
- soit [  $p_1$  &  $p_2$  & ... &  $p_n$  & **non**  $q$  ]  $\Rightarrow$  **F** peut être acceptée par l'interlocuteur comme étant la règle correcte (invalidation indirecte)

Nous avons vu des exemples d'invalidation directe dans le chapitre I.1. Nous avons vu aussi que la distinction entre invalidations directe et indirecte, dans certains cas limites, pouvait dépendre de la représentation logique. Le point fondamental à retenir est que l'explication conversationnelle se traduit par la *suppression d'un étonnement*: soit par la réfutation de l'un des termes explicites de l'étonnement, ou tout au moins sa mise en doute (cf. le prochain extrait), soit par la réfutation d'un terme "oublié" (*i.e.* une hypothèse implicite).

L'invalidation indirecte peut être considérée à juste titre comme "plus riche", dans la mesure où elle amène des éléments nouveaux. Il est important de noter que ce qui la caractérise, c'est qu'elle mentionne un fait non seulement absent de la contradiction initiale, mais surtout qui puisse être **reconnu** et accepté par le premier locuteur comme une prémisse oubliée. C'est ce qui en fait une opération fortement contrainte, au point qu'il est envisageable qu'un système artificiel puisse reconnaître et émettre des invalidations même indirectes: soit une prémisse est volontairement omise par le système dans un premier temps (comme dans le cas de SAVANT3, cf. chap. III.2), l'explication étant ensuite reconnue comme une invalidation de cette prémisse; soit le système dispose de deux niveaux de connaissance d'accessibilité différente (cas de l'EBL), et une contradiction au niveau superficiel peut disparaître ensuite lors de l'accès au niveau profond, grâce à la présence de prémisses non satisfaites (cf. discussion dans le chap. I.3).

### 3.3. L'explication par explicitation

Les arguments précédents laissent supposer que toute explication (au sens où nous l'entendons, *i.e.* excluant les définitions de termes et les longs développements) pouvait être traduite par l'invalidation d'un étonnement. Ce serait oublier le cas de l'*explicitation*, qui consiste au contraire à expliciter l'étonnement lui-même (ou plus précisément la clause de référence, cf. chap. I.1 et I.3) :

A1: Tu es bien du Sud-ouest? Est-ce que la Dordogne fait partie de la Gascogne?  
 B1: Non. Ca fait partie de l'Aquitaine, mais la Gascogne, c'est plus au sud.  
 A2: Parce que dans [le film] "Cyrano de Bergerac", ils présentent Cyrano comme un Gascon.  
 B2: Ah, mais Bergerac, c'est différent. Ca ne fait pas partie du Périgord. Le Périgord n'est pas en Gascogne. Mais le Bergeracois, peut-être.

La réplique A2 est indéniablement une explication (noter le *parce que* qui l'introduit). Ce n'est pourtant pas l'invalidation, mais l'explicitation d'un contexte incohérent. Bergerac étant en Dordogne, A s'étonne de ce que Cyrano soit qualifié de gascon. Il y a contradiction logique entre le fait d'appartenir à la Dordogne et le fait d'être qualifié de gascon, sachant que pour A, la Dordogne est hors de la Gascogne.

En revanche, B2 est bien une explication (ou une tentative d'explication) qui agit comme une invalidation : B2 a pour effet de supprimer, d'invalider le caractère incohérent du contexte :

<i>notations</i>	<i>contexte</i>
cb: Cyrano est de Bergerac	[bd & <b>non</b> dg & bg] $\Rightarrow$ <b>F</b>
dg: la Dordogne est en Gascogne	[cg & cb] $\Rightarrow$ bg
bd: Bergerac est en Dordogne	cb
cg: Cyrano est gascon	<b>non</b> dg
bg: Bergerac est en Gascogne	bd
	cg

Les propositions *bd*, *bg* et [**non** dg] sont incompatibles, or elles se retrouvent toutes vraies. La première clause du contexte est alors saturée (*i.e.* tous ses termes sont vrais) et il y a donc bien une incohérence, qui pousse A à poser la question A1. L'explication A2 est une *explicitation* du contexte problématique via la deuxième clause.

B1 vient confirmer [**non** dg]. En revanche B2 est bien une explication invalidante. Nous pouvons modéliser le raisonnement de B en notant qu'il substitue le Périgord à la Dordogne dans le raisonnement de A. Il refuse donc la première règle<sup>(2)</sup>, et il la remplace par (en notant *bp* : Bergerac est en Périgord, et *pg* : le Périgord est en Gascogne): [*bp* & **non** *pg* & *bg*]  $\Rightarrow$  **F**. Or cette nouvelle règle ne s'applique pas car *bp* est fausse. Le statut de *bg* devient inconnu, de même que celui de l'incohérence.

Nous retiendrons donc qu'une explication conversationnelle consiste dans l'explicitation d'une clause saturée ou dans son invalidation (directe ou indirecte)<sup>(3)</sup>. Nous allons voir les conséquences de ce constat pour l'interaction avec les systèmes à base de connaissances.

<sup>2</sup> Ce refus de règle équivaut en fait à une invalidation du premier terme de la clause plus précise : [ $A \cap B = \emptyset$  &  $X \in A$  &  $X \in B$ ]  $\Rightarrow$  **F** où X est Cyrano, A est la Dordogne et B la Gascogne.

<sup>3</sup> On peut considérer que certaines explications consistent dans l'explicitation ou l'invalidation non pas d'une incohérence, mais d'une improbabilité ou d'une (in)désirabilité (voir chap. I.1). Il semble toutefois que le besoin d'explication passe, même dans ces cas-là, par un étonnement, c.à.d. par un changement de modalité dans la relation improbable ou (in)désirable. Ce point est de toutes façons tout à fait marginal pour ce qui nous occupe ici, et n'a aucune conséquence pratique sur ce qui est exposé dans ce chapitre.

## 4. L'importance de la pertinence des explications pour les SBC

### 4.1. Les explications conversationnelles sont fortement contraintes

Nous avons montré dans le chapitre I.1 à quel point les conversations humaines étaient contraintes d'un point de vue logique. Parmi ces conversations, bon nombre sont organisées autour de la perception d'une incohérence. Dans ce cas, les interlocuteurs ne tolèrent que quatre types de répliques :

- des "co-étonnements", qui manifestent la compréhension de l'incohérence
- des clarifications du contexte
- des invalidations, directes ou indirectes, de l'incohérence
- des accords marquant l'acceptation d'une invalidation

Cet aspect contraint des interactions humaines est fondamental lorsqu'il s'agit d'amener un système artificiel à interagir sur des sujets nécessitant une relation langagière. Pour bien montrer à quel point les humains sont exigeants en matière d'explication, nous reproduisons ici un extrait que nous avons déjà donné dans le chapitre I.1 :

[ex\_Goffman] (tiré de [Tannen 1984])

contexte: A, B et C parlent de sociologie, et il apparaît que B a une bonne connaissance des livres du sociologue Erving Goffman. A et C sont surpris, dans la mesure où ils pensaient que cet auteur n'était connu que des spécialistes.

A1- *But anyway. ... How do you happen to know his stuff?*

B1- *Cause I read it.*

C1- *What do you do?*

A2- *[??] are you in ... sociology or anything?*

B2- *Yeah I read a little bit of it.* [pronounced reed]

A3- *Hm?*

B3- *I read a little bit of it.* [pronounced red]

A4- *I mean were you... uh studying sociology?*

B4- *No.*

A5- *You just heard about it, huh?*

B5- *Yeah. No. I heard about it from a friend who was a sociologist, and he said read this book, it's a good book and I read that book 'n...*

A6- *I had never heard about him before I started studying linguistics.*

B6- *Really?*

A7- *Yeah.*

Cet extrait est particulièrement intéressant, car il montre à quel point une "explication" qui n'agit pas comme une invalidation peut être insatisfaisante. A et B sont étonnés que C ait une si bonne connaissance de l'oeuvre de E.Goffman alors que rien n'indique qu'il ait fait des études de sociologie. A et B détectent une incohérence, qui peut s'exprimer ainsi :

[ connaît( X, livres\_de\_Goffman ) & non sociologue ( X ) ] ⇒ F

Ils posent donc leur question A1. Mais, fait remarquable et pourtant prévisible dans notre modèle, la justification donnée à trois reprises par B (B1, B2, B3) ne les satisfait absolument pas. Effectivement, l'effet logique des répliques de B n'est pas d'invalider l'incohérence précédente :

$lu(B, livres\_de\_Goffman) \Rightarrow connaît(B, livres\_de\_Goffman)$

On observe alors que A et C vont en quelque sorte *exiger* une invalidation acceptable de ce qui a provoqué leur étonnement, en posant leurs questions C1, A2, A4, A5. Ils obtiendront ce qu'ils cherchent, car B5 agit bien comme une invalidation indirecte. En effet, B5 invalide une version modifiée du paradoxe, que A est supposée accepter :

[ *connaît* ( X, livres\_de\_Goffman ) & **non** *sociologue* ( X )  
& **non** *recommande*(Y, X, livres\_de\_Goffman) ]  $\Rightarrow$  F

Noter que l'information apportée par B5 permet de bloquer l'apparition d'une nouvelle incohérence qui résulterait de l'application de ce contexte non plus à B, mais à l'ami de B (noté D) :

*recommande*(D, B, livres\_de\_Goffman) & *sociologue*(D)

Les explications conversationnelles apparaissent ainsi comme très fortement contraintes. On peut considérer que ces contraintes modélisent la *pertinence logique* de l'explication *q* :

*Une explication n'est conversationnellement pertinente  
que si elle explicite ou invalide une clause saturée du  
contexte*

Rappelons qu'une clause saturée est une clause négative dont tous les termes se retrouvent vrais. S'il s'agit d'une clause paradoxale (comme dans les exemples cités), l'explication est relative à une incohérence (voir note 3 p. 230).

#### 4.2. *Génération d'explication dirigée par l'expertise vs. dirigée par les problèmes*

Nous défendons l'idée que les explications fournies par les systèmes à bases de connaissances doivent être calculées comme des invalidations ou des explicitations de contextes problématiques. Dans l'exemple d'interaction avec XPLAIN donné p. 226, le système donne une explication conversationnellement pertinente dans la mesure où il explicite un contexte indésirable (risque de fibrillation ventriculaire).

La situation est plus délicate lorsque la requête de l'utilisateur est provoquée par un problème que ce dernier a rencontré. Nous suggérons fortement le fait que pour être pertinent, le système doit absolument fournir une information qui va invalider le caractère problématique de la situation perçue par l'utilisateur. Or cela est rarement le cas.

En effet, la plupart des systèmes calculent les explications en fonction de leur propre perception de la situation et d'une éventuelle mesure des connaissances de l'utilisateur. Ils ne passent pas par une phase d'identification du problème de l'utilisateur. En d'autres termes, ils fonctionnent suivant un principe de **génération d'explications dirigée par l'expertise**, alors que nous suggérons une **génération d'explication dirigée par les problèmes**.

Par exemple, une architecture comme ESMERALDA [Lemaire 1992], dans laquelle tout est organisé pour adapter le niveau d'expertise de la machine à celui de son utilisateur, ne prévoit pourtant pas d'analyser le problème qui motive la requête de l'utilisateur. L'extrait suivant est relatif au jeu de bridge :

*ESM : . . . Declare one spade*

*User : why ?*

*ESM : Declare major first because fifthminor*

*User : what ?*

*ESM : Fifthminor is a five-card suit in minor. Let me resume. . . Just a definition. Fourthmajor is, for example, d, v, 8, 7 in spadesuit. Declare major first because fifthminor, fourthmajor biddable at level 1 and less than 11 pts.*

*User : why announce one spade ?*

*ESM : Just a definition. Strongopening is an opening with at least two honnors. Declare one diamond implies miss spade contract because the opene r may not have a strongopening with four spades.*

Nous observons ici qu'ESMERALDA produit finalement deux explications, l'une causale, l'autre qualifiée de négative, entre lesquelles elle avait hésité dans un premier temps pour finalement choisir la première. Ce choix entre les explications est effectué en fonction de contraintes de temps de calcul, de connaissances prérequis, etc., sans que soit entreprise une analyse de la motivation de la requête.

Le fait que cette stratégie soit dirigée par la seule expertise n'a sans doute pas de conséquence rédhibitoire lorsqu'il s'agit de donner une définition à un utilisateur novice. De plus, dans certaines situations où le système effectue des actions, l'explicitation des motivations sous-jacentes peut, avec une bonne probabilité, résoudre le problème détecté par l'utilisateur et donc constituer une explication pertinente.

Nous voulons toutefois insister sur l'insuffisance de tels principes, qui sont adoptés dans de nombreux modules d'explication. La génération d'explications dirigée par l'expertise ne peut, selon nous, conduire à des explications pertinentes que par chance. Si l'utilisateur se comporte comme il le fait dans les interactions humaines, il va solliciter le système chaque fois que la situation, telle qu'elle résulte de ses propres actions ou de celles du système, lui semblera incohérente. Quel que soit son niveau d'expertise, l'utilisateur a des attentes, et chaque fois que l'une d'entre elles est contredite, il s'attend à ce que le système lui fournisse une explication qui invalide logiquement son attente.

Dans le cadre de la recherche sur l'explication que nous menons en collaboration dans le projet GENE<sup>(4)</sup>, nous souhaitons parvenir à la spécification d'un système capable de fournir des explications adaptées à son utilisateur. Nous avons donc proposé que toute tentative d'explication de la part du système débute par une analyse logique de la requête, de manière à ce que le caractère problématique de la situation perçue par l'utilisateur soit correctement identifié. Cette identification nécessite généralement une interaction dialogique, comme on peut l'observer avec des experts humains. L'extrait suivant reproduit une interaction entre un lecteur (L) et un bibliothécaire (R) dans un service de prêt de livres :

(extrait de [Polity et al. 1990], dialogue 24 p. 43)

*L1: Excusez-moi, est-ce que vous avez passé tous les livres que j'ai rendus ?*

*R1: Oui, pourquoi, qu'est-ce qui s'est passé ?*

*L2: On me dit que j'ai trop de documents.*

*R2: Donnez-moi votre carte, je vais vous dire ce qu'il vous reste.*

*L3: Merci, je sais qu'il m'en reste deux.*

*R3: Ha, je ne vois rien ; c'est probablement parce que vous n'avez pas réglé votre cotisation à l'entrée.*

Le phénomène étonnant, pour le lecteur, vient de ce qu'on ne le laisse pas sortir avec les ouvrages qu'il vient d'emprunter. Il est pourtant certain de ne pas avoir dépassé le seuil d'emprunt. L'explication donnée en R3 constitue une invalidation indirecte :

contexte de L1

[non trop\_de\_livres & on\_me\_bloque] ⇒ F

on\_me\_bloque

restitution\_enregistrée ⇒ non trop\_de\_livres

contexte de R3

<sup>4</sup> GENE : Gestionnaire d'Explications NEgociées, est un groupe de recherche du PRC-IA réunissant six chercheurs du CNRS et des Universités.

[non trop\_de\_livres & on\_me\_bloque & cotisation\_à\_jour] ⇒ F

L'expert humain, R, est parvenu à résoudre le problème de L parce qu'il a compris que ce problème existait, et qu'il l'a identifié comme une incohérence. Un système qui n'aurait pas fait l'hypothèse d'une situation problématique se serait certainement contenté de répondre "oui" en R1, ce qui n'aurait pas beaucoup aidé le lecteur L. L'extrait suivant, qui est remarquable en raison de la similarité des situations, montre à quel point le calcul de l'explication dépend de l'identification du problème, et ne saurait être calculé *a priori* selon une méthode "dirigée par l'expertise" :

(extrait de [Polity et al. 1990], dialogue 21 p. 43)

L1: *Est-ce que vous avez passé tous les livres ?*

R1: *Oui, tout est passé, qu'est-ce qu'il y a ?*

L2: *Je ne peux pas passer, parce que j'ai trop de livres; pourtant j'ai regardé à la maison, et...*

R2: *Vous avez votre carte ?*

L3: *Je vais la chercher*

R3: *Comme livre ancien, il y a H. de Montrouge, et "Vous n'avez pas vu J. Duflo", ... y a aussi Pouchkine*

[...]

L9: *Je vais vérifier aussi; j'ai dû ne pas le rendre, je vérifierai.*

R9: *D'accord, ça doit être pour ça que ça coince.*

L10: *OK.*

Cette fois, l'explication est une invalidation directe de l'étonnement. Le lecteur a effectivement emprunté trop de livres.

### 4.3. La détection des requêtes problématiques par un SBC

Si l'on souhaite permettre à un SBC de situer ses explications par rapport au problème qui a motivé la requête de l'utilisateur, on doit envisager certaines répercussions sur la conception du système et sur son utilisation, car l'étonnement doit être partageable. Le système doit disposer pour cela de connaissances explicites dépassant celles qui lui sont strictement nécessaires pour conduire ses inférences (*e.g.* règles de cohérence et règles stratégiques). On peut même aller plus loin et ne pas discerner ces trois types de connaissances (inférentielle, cohérence, stratégique) dans le module d'explication, puisque toutes peuvent mener à des étonnements.

Notre système SAVANT3 (voir chap. III.2) peut produire des explications basées sur la détection d'une incohérence. Il s'agit toutefois d'un SBC un peu particulier. Par exemple, il ne comporte aucune distinction entre les règles d'inférence et les contraintes de cohérence (tout se passe comme si SAVANT3 ne comportait que des "règles de production de l'incohérence" (comme celle qui sont introduites dans le système COVADIS tel qu'il est décrit dans [Ayel & Rousset 1990]). SAVANT3 ne comporte pas non plus, comme on l'a vu, de règles stratégiques. Mais si l'on voulait en intégrer, il faudrait le faire explicitement, comme l'a montré Clancey [1987] en créant GUIDON, de manière à pouvoir s'étonner de choix stratégiques effectués par l'étudiant.

SAVANT3 repose sur le principe du partage de l'étonnement. Mais ce partage n'est possible avec un SBC que si le système se met à la portée de son utilisateur. L'étude logique de l'explication conversationnelle permet ainsi de confirmer dans le détail une idée intuitive souvent rappelée : une communication conceptuelle basée sur l'explication n'est possible qu'entre deux entités disposant de connaissances déjà très proches. Dans le cas de SAVANT3, l'étudiant doit absolument être conscient de la contradiction par rapport à laquelle le système se situe. Cela exige en particulier que le système manipule les mêmes concepts que l'étudiant (ce qui exclut qu'il y ait des termes obscurs pour l'étudiant), et que la plupart des règles soient partagées.

Dans le cas des SBC traditionnels, cette exigence a peu de chances d'être satisfaite, du fait de la distance entre règle d'inférence et règle logique. Certaines règles d'un SBC ont une justification heuristique ou statistique. Elles traduisent l'expertise et ont peu de chances d'être connues de l'utilisateur. De plus, l'explication d'un étonnement nécessite que toutes ses prémisses soient explicites. Or il est fréquent que des prémisses "toujours vraies" soient omises des règles d'inférence.

On peut imaginer toutefois de nombreuses situations dans lesquelles le type d'explication que nous prônons soit accessible à un SBC. Ce seront nécessairement des situations où les connaissances du système et celles de son utilisateur seront extrêmement proches: dialogue expert-SBC, système d'aide contextuelle avec modèle des connaissances de l'utilisateur, etc.

## 5. L'explication comme une critique argumentée

### 5.1. Critiquer plutôt que résoudre

Dans l'explication telle que nous venons de l'évoquer, le système à base de connaissance est généralement censé détenir la solution du problème de l'utilisateur. Si l'on conçoit l'explication non comme une solution, mais comme une critique, on obtient une conception beaucoup plus souple du rôle du système à base de connaissance.

Une critique n'est pas un jugement de validité définitif, mais une sorte de marque de désaccord potentiel. Les critiques portent sur les actions (ou les croyances) de l'utilisateur, et peuvent être déclenchées, par exemple, par la détection d'une situation anormale (par ex. une situation incohérente ou indésirable au cours de la résolution de problème). C'est la situation de critique *active*, dans laquelle le critique regarde comme on dit "par-dessus l'épaule" de l'utilisateur, par opposition à la critique *passive* déclenchée sur l'initiative de ce dernier.

De tels systèmes critiques existent, par exemple les systèmes ACTIVIST, LISP-CRITIC, FRAMER, etc. développés à Boulder par l'équipe de G. Fischer [Fischer et al. 1991].

ACTIVIST est un système d'aide actif pour éditeur de texte. Il observe les actions de l'utilisateur de l'éditeur, et en infère ses buts. Dès qu'il a pu observer trois actions qu'il peut interpréter comme l'exécution répétée d'une stratégie sous-optimale, ACTIVIST informe l'utilisateur de l'existence d'une meilleure procédure pour atteindre le même but.

LISP-CRITIC est un autre critique, conçu pour l'aide à la programmation en langage Lisp. Il suggère des transformations du code écrit par l'utilisateur qui rendent son programme plus lisible, plus concis ou plus efficace. Par exemple, il peut suggérer de remplacer une expression conditionnelle à choix multiples : (*cond* (*C a*) (*t b*)) par une expression plus simple (*if C a b*) dans un certain contexte.

Nous avons envisagé d'utiliser notre système SAVANT3 comme un critique pour aider à l'acquisition de savoir-faire complexes. Cela n'allait pas de soi, puisque SAVANT3 est avant tout conçu pour l'enseignement des connaissances conceptuelles, et non pour celui des savoir-faire. Nous avons cependant mentionné l'importance des concepts dans l'acquisition des savoir-faire complexes (chap. III.1).

Notre approche consiste donc à utiliser SAVANT3 pour apporter une critique de type conceptuel à l'étudiant. Nous allons en voir un exemple sous la forme d'une simulation argumentée pour l'apprentissage de Prolog. Mais notre démarche diffère de celle des autres systèmes critiques sur un point qui nous semble important. Dans son principe, la simulation argumentée que nous présentons ici ne consiste pas à fournir un simple diagnostic. Il s'agit, à partir de ce que SAVANT3 considère comme une anomalie de structure dans le programme de l'étudiant, de susciter un *échange argumentatif* avec ce dernier. On rejoint ainsi l'idée de base de SAVANT3 qui utilise une "conversation" pour tester la cohérence des connaissances de l'étudiant, et non une simple succession d'échanges question-feedback [Dessalles 1993c]. Cette approche semble

particulièrement indiquée dans le cas de l'apprentissage d'un langage de programmation, et ce pour plusieurs raisons :

- La possibilité de faire tourner le programme écrit par l'étudiant permet à celui-ci de constater des exécutions anormales, ce qui est censé le motiver pour solliciter la critique de SAVANT3<sup>(5)</sup>.
- Les critiques que l'on peut formuler sur un programme donné sont rarement définitives. L'étudiant peut avoir une motivation, non connue du critique, pour écrire par exemple une clause correcte mais très inefficace. L'échange argumentatif lui permet d'imposer son point de vue.
- Le diagnostic immédiat permet rarement de mettre la faute conceptuelle en évidence. C'est souvent lorsque l'élève se défend en donnant ses raisons que le dialogue remonte jusqu'aux présupposés erronés.

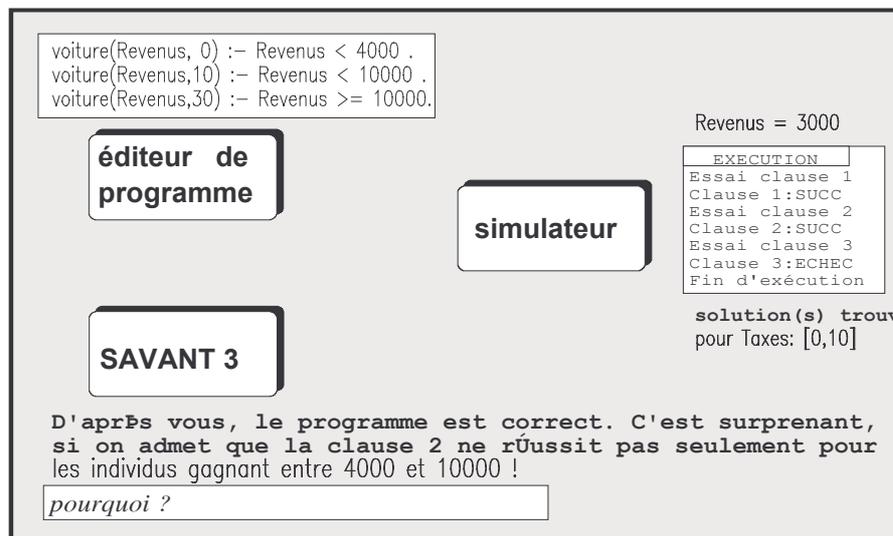
## 5.2. SAVANT3 dans le rôle d'un critique : une simulation argumentée pour l'apprentissage de Prolog<sup>(6)</sup>

Dans la maquette que nous avons mise au point, on demande à l'étudiant d'écrire un petit programme Prolog qui calcule la relation *voiture(X,Y)* suivante :

Pour **X** inférieur à 4000, **Y** doit être nul.  
 Pour **X** entre 4000 et 10000, **Y** doit être égal à 10.  
 Pour **X** supérieur à 10000, **Y** doit valoir 30.

Ce programme est supposé calculer la taxe (**Y**) qu'un individu doit payer sur sa voiture d'après le montant de ses revenus (**X**). Mais nous ne laissons pas à l'étudiant la possibilité d'écrire un programme quelconque. Nous lui fournissons un éditeur très limité qui impose des contraintes sur la formes des clauses qu'il peut écrire. La figure ci-dessous donne un exemple de programme.

A chaque instant l'étudiant peut exécuter son programme : il fournit la valeur de *Revenus* dans *voiture(Revenus, Taxes)*, puis l'interpréteur Prolog cherche toutes les solutions possibles pour *Taxes* pendant qu'une fenêtre de trace montre les clauses appelées et leur échec éventuel (voir figure).



<sup>5</sup> On peut imaginer que SAVANT3 intervienne automatiquement en cas d'anomalie (critique active), mais nous n'avons pas encore implanté cette possibilité.

<sup>6</sup> Ce travail a été réalisé en collaboration avec Pascal Meyers, de la Katholieke Universiteit Leuven, et publié dans [Dessalles & Meyers 1993].

Dans la maquette que nous avons mise au point, on demande à l'étudiant d'écrire un petit programme Prolog qui calcule la relation *voiture(X,Y)* suivante :

Pour **X** inférieur à 4000, **Y** doit être nul.  
 Pour **X** entre 4000 et 10000, **Y** doit être égal à 10.  
 Pour **X** supérieur à 10000, **Y** doit valoir 30.

Ce programme est supposé calculer la taxe (**Y**) qu'un individu doit payer sur sa voiture d'après le montant de ses revenus (**X**). Mais nous ne laissons pas à l'étudiant la possibilité d'écrire un programme quelconque. Nous lui fournissons un éditeur très limité qui impose des contraintes sur la formes des clauses qu'il peut écrire. La figure ci-dessous donne un exemple de programme.

A chaque instant l'étudiant peut exécuter son programme : il fournit la valeur de *Revenus* dans *voiture(Revenus, Taxes)*, puis l'interpréteur Prolog cherche toutes les solutions possibles pour *Taxes* pendant qu'une fenêtre de trace montre les clauses appelées et leur échec éventuel (voir figure).

```
voiture(Revenus, 0) :- Revenus < 4000 .
voiture(Revenus,10) :- Revenus < 10000 .
voiture(Revenus,30) :- Revenus >= 10000.
```

**éditeur de programme**

**simulateur**

**SAVANT 3**

Revenus = 3000

```
EXECUTION
Essai clause 1
Clause 1:SUCC
Essai clause 2
Clause 2:SUCC
Essai clause 3
Clause 3:ECHEC
Fin d'exécution
```

solution(s) trouvée pour Taxes: [0,10]

D'après vous, le programme est correct. C'est surprenant, si on admet que la clause 2 ne réussit pas seulement pour les individus gagnant entre 4000 et 10000 !

pourquoi ?

Cet exercice est utilisé pour enseigner à l'étudiant certains savoir-faire nécessaires pour l'écriture et la compréhension de programmes Prolog. L'objectif est d'attirer son attention sur la justesse et l'efficacité de son programme.

L'interpréteur Prolog ne peut pas mettre clairement en évidence le manque d'efficacité du programme. Il peut simplement montrer que le programme a un comportement anormal, mais il ne peut pas relier ceci à une faute conceptuelle éventuelle sur le fait, par exemple, qu'il y a backtracking et que le programme peut donner plusieurs solutions, ou encore sur la signification des cuts, etc. C'est en cela que l'intervention de SAVANT3 peut être utile.

A tout moment l'étudiant peut soumettre son programme à SAVANT3, qui s'efforce de découvrir une faille (*i.e.* de saturer l'une de ses règles) :

*extrait 2*

```
voiture(Revenus,0) :- Revenus < 4000, !.
voiture(Revenus,10) :- Revenus >= 4000, Revenus < 10000,
!.
voiture(Revenus,30) .
```

S3- D'après ce que j'ai compris, la clause 2 contient  $[Revenus \geq 4000]$ . Mais peut-être pensez-vous que le programme est le plus efficace possible ?

Et.- Pourquoi ?

S3- Il y a tout de même un problème... sauf si vous considérez que la clause 2 est appelée pour les individus gagnant moins de 4000

Et.- oui

S3- D'après vous, la clause 2 est appelée pour les individus gagnant moins de 4000. C'est surprenant si l'on admet que la clause 1 réussit pour les individus gagnant moins de 4000 !

Et.- pourquoi est-ce surprenant ?

S3- Il y a tout de même un problème... D'après ce que j'ai compris :

- la clause 2 est appelée pour les individus gagnant moins de 4000
- la clause 1 réussit pour les individus gagnant moins de 4000
- il y a un cut dans la clause 1

Que proposez-vous comme solution ?

Et.- [la clause 2 n'est pas appelée pour les individus gagnant moins de 4000 ]

S3- Mais attendez... D'après ce que vous avez dit :

- la clause 2 contient [Revenus  $\geq$  4000]
- la clause 2 n'est pas appelée pour les individus gagnant moins de 4000

Je m'avoue surpris !

Et.- j'enlève [Revenus  $\geq$  4000] de la clause 2

S3- Ah bon!... Ok...

Nous avons rencontré certaines difficultés techniques pour la réalisation de cette simulation argumentée. La principale tient à la constitution de la base de connaissances qui permet à SAVANT3 d'émettre ses critiques. Alors que 15 ou 20 règles suffisent habituellement sur des sujets purement conceptuels (cf. chap. III.2), il en faut presque dix fois plus pour gérer la critique de ce simple exercice de Prolog ! Nous perdons ainsi le principal intérêt de SAVANT3 qui est le faible coût de conception des contenus. Nous travaillons actuellement à une définition plus abstraite de la connaissance relative à la simulation. Une telle connaissance serait ensuite "compilée" au moment de l'interaction, en fonction de la situation créée par l'étudiant.

Le diagnostic d'une faute conceptuelle est difficile, voire impossible à partir d'une performance unique [Wenger 1987]. Le pari qui est à la base de notre approche est qu'un tel diagnostic n'est pas nécessaire pour que le système commence à argumenter. Le système détecte une anomalie. Si cette anomalie est la conséquence d'une faute conceptuelle, une argumentation bien menée doit permettre à l'élève de prendre conscience de l'origine de la faute.

## 6. Conclusion : le respect des contraintes cognitives de l'utilisateur

Le type de communication qui s'instaure dans les interactions humaines semble dépendre de celui qui contrôle les conditions de la communication. Plus précisément, lorsque la communication est contrôlée par celui qui possède la connaissance, comme c'est le cas par exemple dans un exposé magistral ou dans le cadre un peu formel d'une réunion de travail, nous avons pu montrer que le type d'explication fourni était organisé en fonction des contraintes cognitives de l'émetteur [Dessalles 1992b]. En revanche, dans les conversations spontanées, la communication semble être plutôt contrôlée par le récepteur, c'est-à-dire par celui qui ne possède pas l'élément d'explication. Or dans ce deuxième cas, les explications émises semblent systématiquement dépendre d'une situation problématique correctement perçue aussi bien par l'émetteur que par le récepteur de l'explication.

Cette constatation nous semble fondamentale, et nous permet de penser que l'exigence que toute explication réponde à une relation problématique est l'indice d'une contrainte cognitive propre à celui qui reçoit l'explication (cf. chap. II.3 et II.4).

Nous suggérons ainsi le fait qu'une meilleure acceptabilité des systèmes à base de connaissances peut être obtenue si les contraintes cognitives de l'utilisateur sont respectées, autrement dit si toutes les explications du système agissent correctement comme des invalidations (ou des explicitations) de relations problématiques perçues par l'utilisateur. Si l'on accepte ce point de vue, alors les systèmes à base de connaissances doivent être conçus de manière à pouvoir fournir des *explications de type conversationnel*.

L'explication de type conversationnel n'est pas *dialogique* par nature ! On peut imaginer que l'explication joue son rôle d'invalidation (ou d'explicitation) du premier coup. Mais ces cas sont rares dans la pratique. L'identification de la relation problématique exige souvent elle-même plusieurs échanges (cf. plus haut les extraits du bibliothécaire). Par ailleurs, une explication qui invalide la relation problématique initiale risque de créer un nouveau problème, et donc un nouveau besoin d'explication. C'est ce que l'on observe couramment dans les conversations. L'explication de type conversationnel est donc dialogique par nécessité pratique !

L'explication de type conversationnel n'exige pas, contrairement à une vision traditionnelle, que le système soit omniscient. L'intérêt de considérer le SBC comme un *critique* est justement de permettre au système de ne fournir que les explications qu'il peut fournir, c.à.d. celles qu'il peut raccrocher à une situation problématique. L'intervention du critique suppose que celui-ci a été capable de déceler une anomalie, et l'explication qu'il produira sera bien de type conversationnel : explicitation de l'anomalie, puis élément de solution permettant de supprimer l'anomalie. L'apport de la vision conversationnelle, par rapport aux systèmes critiques habituels, est de concevoir la critique sous forme *dialogique*.

Notre façon de concevoir l'explication dans les SBC comme devant se conformer aux contraintes de la conversation nous permet de proposer une convergence entre la vision *explicative* et la vision *critique* du rôle du SBC. Il reste bien entendu de nombreuses recherches à effectuer avant de pouvoir démontrer la faisabilité technique de l'explication conversationnelle pour des SBC réels, puis pour montrer que l'explication conversationnelle entraîne une meilleure acceptabilité des SBC.



## **IV- Applications et perspectives**

<b>IV.1 Des perspectives nouvelles</b>	<b>245</b>
<b>IV.2 Conclusion</b>	<b>257</b>



# Introduction à la partie IV

## Une nouvelle approche du traitement du langage et de la modélisation cognitive

### *L'importance de considérer les aspects logiques de l'argumentation*

Notre modèle de la communication spontanée, et ses implications cognitives, ne peuvent pas rester sans impact sur certains aspects de la recherche.

Le traitement du langage naturel, par exemple, est effectué généralement hors d'un contexte argumentatif. La machine est placée ainsi dans une situation beaucoup plus difficile que celle qui est la nôtre lorsque nous parlons et analysons ce qui nous est dit ! La prise en compte des aspects argumentatifs devrait faciliter de manière significative les autres aspects du traitement linguistique (traitements lexical, syntaxique, sémantique).

L'analyse du langage peut certes être facilitée, mais aussi sa génération. Notre modèle suggère qu'un texte sera mieux compris s'il est *argumenté*. Les techniques que nous avons explorées avec PARADISE nous permettent d'engendrer des textes qui ont cette propriété, essentielle à nos yeux.

Notre modèle a d'autres implications, et il n'est pas possible de les évoquer toutes. Nous avons choisi, tout d'abord, de parler des contraintes conversationnelles en les présentant comme le reflet des contraintes cognitives de celui qui *reçoit* l'information conceptuelle.

Ensuite, nous évoquons l'analogie entre les processus conversationnels et les processus de conception tels qu'ils peuvent être effectués par un ingénieur en bureau d'études.

Enfin nous abordons brièvement la question des liens entre certaines pathologies psychiatriques, qui se traduisent par des troubles de la pensée logique, et les dérèglements conversationnels qui semblent en résulter.

- *Nous terminons le chapitre IV.1 en soulignant une analogie qui nous semble particulièrement parlante, entre les mécanismes de la conversation quotidienne et ceux qui président à la découverte scientifique.*

- *Le chapitre IV.2 est la conclusion de ce travail. Le modèle conversationnel présenté dans la partie I et le modèle cognitif qui en est issu (partie II) sont le résultat d'un travail expérimental et théorique. Leur adéquation avec les données disponibles est excellent, et ils semblent féconds sur les plans théoriques et techniques. Pourtant, pourtant, l'accueil de ces modèles par les disciplines avec lesquels ils interfèrent n'est pas (encore) acquis...*

# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Introduction

L'étude de la communication spontanée, telle que nous l'avons menée, avec la mise en évidence de contraintes très fortes auxquelles les interlocuteurs se soumettent mutuellement, nous semble avoir des conséquences potentielles très importantes, et dans différents domaines.

L'existence de ces contraintes, et des compétences que celles-ci supposent chez les interlocuteurs, nous a permis de dessiner un modèle concernant "l'architecture" de la cognition. Elle nous a permis aussi de proposer un nouveau mode d'interaction humain / machine, basé sur l'évitement de la contradiction.

Nous voulons esquisser ici d'autres manières de tirer les conséquences de l'existence de "règles" incontournables dans la communication humaine. Nous évoquerons successivement quelques conséquences sur le plan de la cognition et de la psychologie, quelques suggestions concernant le traitement automatique des aspects supérieurs du langage, et enfin quelques réflexions de nature épistémologique visant à établir un parallèle entre le comportement conversationnel et le comportement scientifique.

## 2. Contraintes conversationnelles et contraintes cognitives : un lien étroit

Une fois analysé, le phénomène des contraintes conversationnelles apparaît comme assez spectaculaire, étant donné la simplicité et la rigueur de ces contraintes : dans les situations d'échange spontané, nous n'acceptons de traiter que des éléments qui se rapportent d'une manière particulière à un fait incohérent, improbable ou (in)désirable.

Face à cette constatation, on peut avoir deux attitudes. L'une consiste à dire que ces contraintes sont spécifiques à la situation de la conversation spontanée, qu'elles résultent d'une sorte d'équilibre entre les buts des participants. Dans cette façon de voir, elles ne seraient qu'une conséquence assez lointaine de notre architecture cognitive. Il ne semble pas nécessaire d'argumenter beaucoup pour montrer que cette hypothèse est erronée : elle ne permet pas d'expliquer la *spécificité* et la *systématicité* des contraintes que nous avons décrites.

Il nous reste donc la deuxième solution : *les contraintes conversationnelles sont le reflet direct de contraintes cognitives*. Ces contraintes cognitives sont liées au traitement *logique* des connaissances que notre environnement nous permet d'évoquer ou de créer.

Nous avons poussé cette hypothèse assez loin, allant jusqu'à proposer dans le chapitre II.4 une architecture neuronale pour la détection des incohérences. Nous allons évoquer maintenant quelques indices suggérant l'existence de ces contraintes cognitives de nature logique.

### 2.1. Contraintes cognitives de l'auditeur et "pédagogie conversationnelle"

Lors d'une étude comparative de types de communication non spontanés (réunion de travail, exposé oral ou écrit), nous avons observé que les contraintes conversationnelles étaient fort peu respectées [Dessalles 1992b].

Pour rendre compte de ce phénomène, nous avons supposé que le locuteur, dans une situation non spontanée, est parfois beaucoup plus libre de l'organisation de son discours. Celle-ci reflète donc sans doute la façon dont les connaissances sont organisées dans sa mémoire. C'est ainsi que l'on observe des locuteurs qui relatent les événements selon leur stricte succession temporelle, même si cela les amène à évoquer des éléments sans intérêt pour la problématique du moment. Ou encore, le locuteur en situation d'exposé "magistral" va dérouler le contenu de ce qu'il a à dire en suivant un *plan*. Or ce plan correspond souvent à une hiérarchie taxonomique, ou à un arbre *et/ou* parcouru "en largeur d'abord" [Dessalles 1992b], ce qui est très différent des arbres caractéristiques de l'argumentation spontanée (cf. chap. I.1 et I.3). Dans d'autres cas encore, l'organisation du discours est moins claire, et semble gouvernée par des chaînes associatives.

L'extrait suivant reproduit un passage d'une réunion de travail. Le style du locuteur principal (celui qui détient l'information attendue), B, change totalement selon qu'il est soumis aux objections de A (style conversationnel) ou qu'on le laisse parler (style "magistral", en B1 et B7).

[...]  
 B1- *Il s'agissait d'une demande de X. Bon alors Médiavec, si je ne m'abuse, j'ai peut-être les dates quelque part... Médiavec, c'est, je suis bien organisé quand même, du 8 au 13 avril.*  
 A1- *C'est-à-dire? Le 8 c'est quoi, quel jour?*  
 B2- *Samedi. C'est le 7, du 7 au 13 quoi. Du 7, du vendredi 7 au jeudi 13.*  
 A2- *Ca démarre le vendredi 7 ou c'est une préparation, le vendredi 7*  
 B3- *Préparation sans aucun doute, ça démarre, enfin*  
 A3- *A mon avis ça démarre le vendredi 7*  
 B4- *Oui moi aussi, je pense que ça démarre le vendredi 7,*  
 A4- *D'accord.*  
 B5- *Bien*  
 A5- *Donc, ça veut... euh, ça..., alors, ...*  
 B6- *Alors, non !*  
 A6- *Ouais vas-y, continue*  
 B7- *Je termine! J'ai eu X au téléphone. J'ai eu Y ultérieurement mais pour un autre problème. Alors le... il serait souhaitable de savoir si on participe, sous quelle forme, X propose un stand avec deux emplacements, si on a plusieurs prestations. Y ultérieurement a dit que on aurait un petit stand. C'est à voir. Bon c'est à discuter. Alors il faut savoir si on est présent, les jours de présence, qui sera présent, et comment on organise, et quelles sont les prestations que l'on propose.*  
 A7- *D'accord. Bon ça ça veut [...] ça veut dire deux personnes sur le stand, [...]*

La structure de B7 est très éloignée de ce que l'on peut observer dans une conversation spontanée. Par exemple, la mention "*J'ai eu Y ultérieurement mais pour un autre problème*", au moment où elle est délivrée, ne peut pas être raccrochée à la problématique par l'auditeur. Elle serait inadmissible dans une conversation. Il semble que la conversation téléphonique avec Y soit associée avec celle de X dans la mémoire de B, et associée au contexte. A ne peut percevoir que cette simple association, à l'exclusion de tout lien logique. Mais étant donné le caractère rituel de l'interaction (réunion à caractère professionnel), il n'exerce pas sur B la censure conversationnelle que l'on aurait observée dans une interaction spontanée.

En revanche, A parvient à faire valoir certains de ses droits d'auditeur, en A1 et A2 (noter qu'il se fait rappeler à l'ordre en B6). Il insère la mention des dates dans une problématique (il s'agit de savoir si l'installation du stand se fait un jour ouvrable ou non).

En résumé, on peut caractériser ce type d'observation ainsi :

- lorsqu'il n'est pas soumis aux contraintes de la conversation, l'émetteur d'une information semble avoir tendance à structurer son discours selon ses propres contraintes de stockage en *mémoire*.
- lorsqu'il est soumis aux contraintes de la conversation, l'émetteur d'une information respecte les contraintes de *compréhension* de l'auditeur, et structure systématiquement son discours selon une problématique.

Cette constatation, si on l'accepte, a des conséquences importantes en matière de communication des informations (faits et concepts).

A notre connaissance, la structuration du discours pédagogique a été fort peu étudiée. Il s'agit pourtant d'un domaine dont les enjeux sont particulièrement importants. Nous pensons qu'il est possible d'ancrer toute une pratique pédagogique sur les principes conversationnels. Ceux-ci, et nous l'avons expérimenté dans notre propre pratique, fournissent des règles pour structurer le discours "magistral", le discours écrit, le discours audiovisuel, etc. de manière à ce que les contraintes cognitives de l'auditeur, telles que nous les décrivons, soient respectées. Les efforts que nous avons déployés dans le développement de SAVANT3 (chap. III.2) sont, de même, motivés par l'espoir que le respect des contraintes cognitives de l'apprenant permette une meilleure efficacité et une meilleure acceptabilité du "discours" pédagogique.

## 2.2. *Contraintes cognitives du concepteur et "invention conversationnelle"*

L'étude des conversations spontanées offre des enjeux qui dépassent largement le cadre récréatif de la discussion "à bâtons rompus". L'observation de professionnels, en train de résoudre un problème à plusieurs, permet de constater l'étonnante identité entre la structure de leurs échanges et les aspects logiques de l'argumentation telle qu'on l'observe dans les conversations. La compréhension du processus de conception représente un enjeu industriel non négligeable : on peut en effet envisager de faciliter la conception en définissant des méthodes, ou en offrant des systèmes d'aide qui soient plus ambitieux sur le plan de la manipulation des concepts que ne le sont les actuels produits de CAO.

La ressemblance entre le processus de conception en résolution de problème et la conversation libre semble avoir échappé à la plupart des observateurs, tout simplement parce qu'il n'est pas facile d'observer des cas spontanés de conception collective. N. Bonnardel a pu réaliser des enregistrements de situations au cours desquelles les personnes observées se sont mises à concevoir collectivement et spontanément des solutions pour un problème réel. Dans ce genre de situation, les concepteurs évaluent les solutions en fonction d'un certain nombre de contraintes (contraintes construites, déduites et prescrites), et c'est la confrontation aux contraintes qui oriente la recherche de solution [Bonnardel 1989, 1990].

Ces contraintes de nature technologique ou relatives au coût ou aux délais sont parfaitement exprimables dans la logique conversationnelle que nous avons définie (avec notamment l'utilisation de la modalité **IND** (indésirable)). La conception dirigée par les contraintes devient alors très proche de ce que prévoit notre modélisation des conversations (chap. I.3). Le passage suivant, extrait de l'étude de N. Bonnardel, montre clairement le rôle de l'argumentation logique dans l'élaboration collective de la solution<sup>(1)</sup> :

C257: [...] Alors, au point de vue dessin, actuellement, on en est là, il faudrait que l'on dessine, un petit peu, une métallique non raidie, une métallique raidie et une...  
P258: sandwich.

<sup>1</sup> Cette étude a été menée dans le cadre d'un contrat INRIA / Aérospatiale. Il s'agissait d'étudier les situations de transfert de projet. Le passage cité est particulièrement intéressant pour ce qui nous occupe ici, car il fait partie d'un dialogue spontané au cours duquel les participants se sont mis à concevoir dans le cadre d'un problème réel, dépassant ainsi largement la consigne expérimentale que l'observatrice leur avait donnée.

C259: *et une sandwich, parce que moi, il faudrait que j'aie l'idée des désignations des cadres.*  
 R260: *Pas de monolithique ?*  
 C261: *Non, monolithique, a priori, on abandonne.*  
 P262: *Tu es sûre qu'en monolithique raidi, ce n'est pas mieux?*  
 C263: *Mais ça coûte un prix dingue, aussi.*  
 R264: *Raidi, si c'est des nervures intégrées.*  
 P265: *Ah non, des petits omégas.*  
 R266: *Et collés. On peut mettre des omégas collés.*  
 P267: *Remarque, c'est le même gros truc que pour le métallique, dès que tu vas vouloir venir mettre un équipement, ils vont être embêtés.*  
 R268: *Non, ce qu'il faut, si on fait du monolithique, c'est que ce soit plutôt pas raidi sinon ça coûte une fortune.*  
 P269: *Ouais.*  
 C270: *Non, le problème du monolithique, c'est que l'on va avoir un plan de qualification long parce que du monolithique, on a aucune structure qualifiée [opérationnelle] [...]*

Ici, P et R tentent de trouver une solution à base de matériau monolithique, face à C plutôt sceptique. Les tours de parole jouent des rôles logiques conformes à ce que l'on observe dans les conversations. Par exemple, C263, P267, C270 sont des réactions antagonistes (cf. chap. I.1). P262, R264, P265, R266 sont des invalidations indirectes. Noter la forme linguistique de R268 ("*ce qu'il faut, si..., c'est que... sinon...*") qui laisse transparaître la logique sous-jacente, et qui est presque identique à une forme utilisée dans PARADISE (chap. I.3).

### 2.3. Pathologie cognitive et troubles conversationnels

Certaines recherches visent à déceler des dérèglements objectifs dans le discours de patients atteints de troubles psychiatriques. Il semble en effet que l'un des problèmes fondamentaux de la psychiatrie soit la catégorisation des malades en fonction des symptômes observés, de manière à trouver éventuellement des causes organiques ou psychologiques distinctes selon les catégories, et en tout état de cause à prescrire des thérapies suffisamment spécifiques et adaptées.

Certains malades mentaux, notamment parmi les schizophrènes, présentent ce qu'il est convenu d'appeler "des troubles de la conversation". Nous avons cité (chap. I.1) un extrait recueilli par A. Trognon auprès d'une patiente en hôpital psychiatrique. Il serait tout à fait intéressant d'essayer de voir si l'observation des conversations spontanées que les malades sont susceptibles d'avoir entre eux ou avec le médecin ne peut pas contribuer à élaborer une catégorisation objective selon ce critère de compétence conversationnelle.

G. Bateson et les chercheurs qui ont travaillé avec lui ont développé une théorie de la schizophrénie qui est organisée autour des problèmes de troubles de la communication. Il est frappant de constater que ces troubles, qui semblent se traduire entre autres par une déstructuration de la logique du discours, sont expliqués par ces auteurs comme dus en partie à une incapacité à gérer certains paradoxes logiques dont la solution nécessite le passage à un niveau de métacommunication [Bateson et al. 1956; Watzlavick et al. 1967]. Leur théorie de la "double contrainte" (*double bind*) suppose l'existence de deux injonctions contradictoires émanant (le plus souvent) des parents. Par exemple une mère, qui se raidissait lorsque son enfant psychotique s'approchait et mettait le bras autour de ses épaules, lui dit après qu'il se fut éloigné : "Chéri, n'aie pas peur de montrer tes sentiments". Pour G. Bateson, l'enfant se retrouve perplexe : le raidissement de sa mère lui laisse supposer que ses marques d'affection sont inacceptables, alors que cette même mère laisse entendre qu'il doit les montrer. Un enfant qui devient schizophrène baignerait dans un contexte empli de ce genre de situations auto-contradictoires<sup>(2)</sup>, et serait incapable d'en sortir, par exemple en

<sup>2</sup> Ces auteurs parlent aussi d'injonction paradoxale. Une femme demandait à un mari passif: "je veux que tu me domines". Des parents jugeaient leur enfant trop dépendant : " Ne sois donc pas si docile". Un patient (P) se retrouve "piégé" par sa mère (M) qui déclare devant le praticien que P la contredit tout le temps. Si P s'oppose à cette déclaration, il la valide du même coup. Il la valide

dénonçant la contradiction. Cela exigerait de lui qu'il passe à un niveau de méta-communication, ce dont les schizophrènes (ou au moins certains d'entre eux) semblent incapables.

On ne peut s'empêcher de penser qu'il existe peut-être un lien entre cette incapacité à gérer ce genre de paradoxe logique, et les troubles conversationnels qui se traduisent entre autres par des erreurs dans le traitement logique des arguments. Quoi qu'il en soit, le modèle de Bateson rejoint nos préoccupations en montrant l'importance de la perception des rapports logiques dans la cognition humaine.

#### 2.4. *Styles conversationnels et styles cognitifs*

Pour conclure sur les aspects psychologiques qui peuvent être liés à la modélisation de la communication spontanée, nous voudrions livrer brièvement une réflexion concernant les "styles" cognitifs.

Il semble que notre comportement conversationnel soit organisé autour de trois compétences de base (cf. chap. I.1) :

- la capacité à détecter des incohérences et à les invalider (*i.e.* les expliquer)
- la capacité à repérer les événements *a priori* improbables et à en diminuer l'improbabilité
- la capacité à juger de la désirabilité ou de l'indésirabilité des éventualités, et à opposer ces deux aspects pour une même éventualité (*i.e.* peser le pour et le contre)

Nous semblons tous posséder ces trois capacités. Ceci ne veut pas nécessairement dire que nous les exerçons avec une égale propension. Il semble, mais cela nécessiterait une étude quantitative, que certains locuteurs aient une préférence pour l'un ou l'autre de ces modes conversationnels. On ne peut s'empêcher, dans ces conditions, de penser que cette préférence correspond à un style cognitif, et de suggérer, en caricaturant, que ces trois styles conversationnels se retrouvent respectivement dans trois orientations d'esprit : celle du scientifique (ou du policier), celle du journaliste (ou du conteur), et celle du gestionnaire (ou de l'homme politique).

### 3. Analyse et traitement des aspects supérieurs du langage

Parmi les perspectives que peut offrir notre modélisation conversationnelle, il nous faut bien entendu mentionner les implications en linguistique et dans le domaine du traitement automatique du langage.

#### 3.1. *Une nouvelle façon d'aborder la pragmatique*

La pragmatique est souvent perçue comme le "parent pauvre" parmi les domaines de la linguistique. Dans le chapitre I.2, nous avons montré en quoi notre modèle différait des modèles dominant en pragmatique. Nous voulons suggérer fortement le fait que la prise en compte des relations logiques dans l'argumentation est propre à renouveler notre vision du langage. A cet égard, la linguistique telle qu'elle est pratiquée nous offre une perception tronquée de la compétence langagière. A mesure que l'on s'élève dans les niveaux identifiés par la linguistique (phonologie, syntaxe, sémantique, pragmatique), on observe une complexité croissante, et un manque croissant d'outils théoriques. On en vient à penser que l'usage pragmatique du langage, qui inclut notamment ce qui touche à l'argumentation, est trop complexe pour être analysé de manière précise, qu'il est gouverné par le flou et par les particularités changeantes des situations. Si l'on tient cela pour admis, on repousse le phénomène argumentatif dans l'inconnaissable lié à la complexité du comportement humain.

Nous nous opposons à cette façon de voir. Il n'y a aucune raison *a priori* pour que les phénomènes pragmatiques soient encore plus complexes que ceux qui sont liés aux aspects sémantiques. Bien au contraire, nous avons essayé de montrer qu'il existe un certain nombre de contraintes qui régulent notre communication, et qui, à bien y réfléchir, la rendent possible.

---

aussi s'il se tait. Or son affection pour sa mère ne permet pas qu'il s'oppose systématiquement à elle. Lorsque ces injonctions paradoxales sont prises au pied de la lettre, elles constituent de véritables pièges pour l'esprit.

La langue n'est pas un moyen d'échanger des informations. Si tel était le cas, la pragmatique consisterait à décrire les protocoles que nous utiliserions pour communiquer de manière optimalement efficace, un peu comme ceux qui sont utilisés dans les réseaux de transmission de données. Nos tours de parole seraient de longues tirades contenant "en vrac" toutes nos connaissances sur le sujet abordé. Mais tel n'est pas le cas.

Les conversations sont bien autre chose. Elles sont pour nous le moyen de confronter nos connaissances. Pour reprendre la métaphore informatique, il ne s'agirait pas de communiquer des données, mais les règles d'un système à base de connaissances. On connaît les difficultés liées à l'adjonction d'une nouvelle règle dans une base de connaissances (tests de cohérence, de redondance, etc.). Si l'on devait amener deux systèmes à partager leur expertise, on serait peut-être amené à imaginer des protocoles qui ressembleraient étrangement à celui que les humains utilisent dans leurs conversations quotidiennes.

## 3.2. *Le traitement automatique des aspects supérieurs du langage*

### 3.2.1. *Analyse du langage et contraintes argumentatives*

Le problème de la compréhension du langage est un fantastique défi scientifique et technique. La prise en compte des contraintes argumentatives peut permettre de simplifier considérablement l'analyse des énoncés linguistiques dans certaines conditions.

Un cas extrême nous est fourni par les situations didactiques du type de celles pour lesquelles SAVANT3 a été conçu. Dans ces cas là, le système contrôle parfaitement le contexte, et le problème d'analyse du langage est transformé en un problème de *reconnaissance* de prédicat. Par exemple, SAVANT3 parvient à comprendre correctement, dans la majorité des cas, ce que l'étudiant a réellement voulu signifier, grâce à une simple reconnaissance de mots clés et à la détection d'une éventuelle négation.

Dans la plupart des applications, la situation est généralement moins favorable, car le contexte est rarement cerné avec autant de précision. On peut toutefois facilement imaginer que les contraintes du niveau pragmatique, telles que nous les avons analysées, permettent de lever la plupart des ambiguïtés des niveaux inférieurs de l'analyse (niveaux lexical, syntaxique, sémantique). Ceci n'est toutefois envisageable, répétons-le, que si *le contexte argumentatif de la situation est correctement circonscrit*. Le système d'analyse doit donc avoir à sa disposition une large base de connaissances, dans laquelle il doit reconnaître le caractère problématique de la situation évoquée par l'extrait à analyser. De nombreuses applications de l'analyse du langage peuvent satisfaire ces exigences.

### 3.2.2. *Génération de textes argumentés<sup>(3)</sup>*

Nous n'avons pas exploré ce problème de l'analyse des énoncés autrement que d'une manière marginale, à l'occasion du développement de SAVANT3 (chap. III.2) et de PARADISE (chap. I.3). En revanche, nous avons commencé à travailler sur le problème de la génération de textes.

La génération de textes présente des enjeux importants, soit pour valoriser les systèmes à base de connaissances (cf. la problématique de l'explication dans les SBC, chap. III.4), soit dans le cadre de traitements linguistiques (résumé, traduction, paraphrase, etc.). Les systèmes de génération de textes réalisés à ce jour fonctionnent généralement en deux temps : génération d'un "squelette" (structure profonde) à partir de principes de planification et d'une sélection des items à présenter, puis génération du texte en langue naturelle (structure de surface) à l'aide de traitements linguistiques et stylistiques. L'originalité de notre propre approche se situe plutôt dans la première phase.

Il existe peu de justifications pour décider qu'un texte doit avoir une organisation plutôt qu'une autre. On cite généralement la théorie RST (*Rhetorical Structure Theory*) [Mann & Thompson 1988] qui est censée offrir un cadre descriptif de la "macro-structure" d'un texte : organisation hiérarchique, relations fonctionnelles entre les parties. Par exemple voici, donnée par ces auteurs, la description de la relation "Evidence" (R = lecteur (*reader*) ; W = rédacteur (*writer*) ; N = noyau de la relation ; S = satellite) :

<sup>3</sup> Ce travail exploratoire a été mené en collaboration avec Oyvind Rastad (Division of Computer Science, The Norwegian Institute of Technology).

*relation name : Evidence*  
*constraints on N : R might not believe N to a degree satisfactory to W*  
*constraints on S : R believes S or will find it credible*  
*constraints on the N + S combination : R's comprehending S increases R's belief of N*  
*the effect : R's belief of N is increased*  
*locus of the effect : N*

La mise en oeuvre des principes de la RST suppose que le système de génération de texte possède des connaissances sur les croyances du lecteur et sur la capacité de tel ou tel argument à convaincre.

Dans notre propre démarche, nous nous interdisons toute connaissance "méta-argumentative" de ce type, portant sur le rôle ou l'effet des arguments. Ces effets peuvent être *calculés*, comme nous l'avons démontré en développant PARADISE (chap. I.3).

De plus, nous suggérons que le texte engendré soit organisé en priorité d'après un squelette *argumentatif*. Certes, certains systèmes de génération de textes prennent en compte quelques principes argumentatifs, notamment pour l'emploi correct des connecteurs "mais", "pourtant", etc. (e.g. [Cerbah 1992]). Mais ce souci intervient généralement au moment d'engendrer la structure de surface du texte. Nous suggérons que la structure profonde du texte soit elle-même organisée selon une trame argumentative. Voyons cela sur l'exemple suivant, tiré d'un magazine (Time, 25-01-93) :

*"The main problems for the big aircraft are likely to be on the ground ; in fact, a new generation of airports may be needed to serve them. While the plane should be able to use runways no longer than those required for 747's, ramps as well as runways may have to be strengthened and widened because of the plane's weight and size. Terminals might need extensive rebuilding despite the folding-wing scheme."*

Notre point de vue est que ce type de texte peut être dérivé d'une conversation fictive comme celle-ci :

*A1 : C'est ennuyeux, parce qu'il va falloir créer une nouvelle génération d'aéroports pour ces ultra-jumbos*  
*B1 : Pourquoi ? On n'aura pas besoin de rallonger les pistes, puisqu'ils ont une meilleure portance.*  
*A2 : Certes, mais ils sont plus lourds. Il faut renforcer les pistes. Et puis il faut les élargir, à cause de l'envergure. Sans compter les terminaux qu'il faut aussi refaire*  
*B2 : Mais je croyais qu'on pouvait replier les ailes pour l'accès aux terminaux*  
*A3 : Oui, mais chaque avion représente deux ou trois fois plus de passagers. Il faut pouvoir les absorber.*

La génération de texte ainsi conçue consiste à créer un arbre argumentatif à partir d'une base de connaissances, selon les principes explorés dans PARADISE (chap. I.3). Ensuite, cet arbre est parcouru selon une stratégie complexe (*i.e.* mixte en largeur et en profondeur) pour tenir compte de facteurs comme les répétitions (par ex. les ailes ou le poids interviennent à deux endroits différents dans l'arbre décrivant le texte ci-dessus) et la charge cognitive due à la profondeur des "backtrackings".

Nous donnons ci-dessous une forme linguistique proche de la structure profonde, engendrée à partir du contexte logique de l'extrait précédent :

*We need not to have newAirports, because we do not have longerRunways : Although bigAircraft implies highWeight, we do not have longerRunways, because of wideWings. Despite this, we need newAirports, because of strongerRunways, widerRunways, and terminalRebuilding.*

*First of all, we need newAirports because of strongerRunways. We have strongerRunways because bigAircraft implies highWeight. We also need newAirports because of widerRunways. We have widerRunways because bigAircraft implies bigSize. Lastly we need newAirports because of terminalRebuilding. In fact, although foldingWings, we need terminalRebuilding because bigAircraft implies bigSize which implies manyPassengers.*

Nous sommes encore loin d'une forme de surface, dont on pourrait imaginer qu'elle ressemble dans ce cas à l'extrait dont nous nous sommes inspirés. Mais il est important de noter que la constitution de l'arbre argumentatif ouvre ensuite beaucoup de possibilités et de souplesse pour la mise en oeuvre de mécanismes dont la fonction est d'assurer la lisibilité.

Nous pensons que le respect de la trame argumentative est essentiel pour la compréhension du texte engendré. Ce n'est bien sûr pas suffisant, et il y a beaucoup de travail à faire sur le parcours de l'arbre argumentatif, sur son élagage, et sur le traitement linguistique qui conduira à la structure de surface. Toutefois, le respect des contraintes logiques de l'argumentation, telles qu'on les observe dans la communication orale spontanée, nous semble fournir un cadre intéressant pour la génération de textes.

#### 4. Comportement conversationnel et comportement scientifique

Nous voulons livrer ici une réflexion d'ordre épistémologique : *la Science est une conversation*. Si l'on veut bien nous suivre un instant dans ce parallèle entre l'activité scientifique et l'activité langagière, nous pouvons même ajouter : la Science est une conversation sur le mode "paradoxal" (cf. chap. I.1). Autrement dit, nous proposons d'ancrer l'activité scientifique dans le comportement conversationnel. L'activité de modélisation, que l'on connaît surtout dans la recherche scientifique, serait une activité naturelle, à laquelle nous nous adonnons quotidiennement.

Mais si tel est le cas, si l'activité modélisatrice fait réellement partie du répertoire comportemental de base de chacun d'entre nous, comment expliquer que la Science (dans l'acception contemporaine de ce terme) soit née récemment dans l'histoire des civilisations, qu'elle soit étrangère à certaines cultures et à certains individus ? Peut-être que l'examen des conversations peut justement nous apporter une meilleure compréhension de la façon dont nous interprétons et expliquons scientifiquement le monde physique qui nous entoure.

Il serait en principe possible de traduire toute la connaissance scientifique par un gigantesque ensemble de formules logiques. Sur un domaine particulier, la Science produit des modèles, et un modèle n'est rien d'autre qu'un ensemble de formules logiques associé à une interprétation des symboles employés. Sur ce point déjà, on peut déceler une certaine ressemblance avec les conversations : au cours de leur interaction, les interlocuteurs modifient et établissent un contexte logique, qui lui aussi est une traduction logique d'un aspect (très) limité du monde qui nous entoure.

Bien sûr, il peut paraître osé de mettre en parallèle la futilité de certaines conversations avec la rigueur que la Science s'impose. Et il est vrai qu'une conversation peut paraître bien superficielle à un observateur extérieur qui connaît bien le sujet abordé. Il arrive aussi que certains arguments ne soient pas compris et que personne n'ose les invalider de peur de trahir son ignorance. Ou encore que les interlocuteurs affirment des faits dont ils ne sont en réalité pas du tout certains. Tout ceci ne contribue pas à faire des conversations un processus par lequel la connaissance avance aussi sûrement qu'elle le fait du fait de la recherche scientifique !

En fait, la question n'est pas là. Ce que nous cherchons à suggérer ici, c'est qu'il existe une analogie assez frappante entre la manière dont la recherche scientifique est menée et le déroulement d'une conversation, sans pour autant prétendre qu'il s'agit de processus superposables. Dans les deux situations, des modèles de la réalité sont transformés. Si l'on pense à une conversation sur le mode paradoxal, le corps de formules logiques qui comporte une incohérence est modifié et augmenté jusqu'à ce que l'incohérence disparaisse.

La science, de son côté, s'attache principalement à expliquer des phénomènes jusque-là inexpliqués, et donc par définition incohérents. Un phénomène apparaît comme bizarre (incohérent) s'il entre en contradiction avec des hypothèses admises. Penser à l'expérience de Michelson et Morley à la fin du siècle dernier, qui entraînait en contradiction avec l'hypothèse selon laquelle la lumière obéit à la loi de composition des vitesses (simple addition algébrique), ou encore à la découverte "par hasard" de la pénicilline par Fleming, qui fut surpris que les bactéries pussent mourir dans certaines régions de sa boîte de Pétri (à proximité d'une moisissure), bien que la nourriture fût suffisante.

Mais l'analogie entre la Science et les conversations sur le mode paradoxal va plus loin. Un des mérites du scientifique est parfois de s'étonner d'une anomalie là où les autres ne perçoivent que l'ordre banal des choses, et de commencer une recherche comme d'autres commencent une conversation. Et de même que la conversation va se poursuivre à coup d'invalidations, la recherche scientifique va progresser grâce à une succession de réfutations.

Ce dernier point nous paraît essentiel. Car ce qui, à notre sens, distingue le discours scientifique de tout autre discours de connaissance, n'est pas à chercher dans l'emploi des chiffres, des formules, ou dans l'énoncé de lois, mais réside dans la possibilité offerte à tous, par l'auteur d'un modèle, d'invalider celui-ci. Autrement dit, l'activité scientifique est caractérisée par un état d'esprit : celui d'autoriser et d'organiser sa propre réfutabilité (cf. [Popper 1959]). Certains voient dans Thalès l'un des premiers penseurs d'importance ayant adopté une attitude scientifique, quelques six siècles avant notre ère. Afin de résoudre le paradoxe entre la complexité de la matière et l'exigence d'une simplicité du monde postulée au départ, il émit l'hypothèse que toute matière avait une composition unique : celle de l'eau. Ce modèle n'a pas en soi beaucoup d'intérêt, mais c'est la façon dont il est énoncé qui est nouvelle pour l'époque, comme l'explique très bien J. Lalumia :

"Si l'on se penche sur la mythologie grecque, on découvre que beaucoup d'autres ont dit, *avant* Thalès, que le monde entier est composé d'une même matière et que cette matière est de l'eau. Mais ces hommes étaient des poètes ou des prophètes. Dans les sociétés anciennes, poètes et prophètes occupaient une position particulière. Ils étaient considérés comme des êtres pourvus d'un certain don de voyance, de sorte que, si l'un d'entre eux déclarait que l'univers est composé d'eau, ce don de voyance particulière était considéré comme étant à l'origine de cette déclaration. Il en résultait que le discours du "voyant" passait pour l'expression d'une vérité que lui seul avait la faculté de percevoir, de sorte que nul ne tentait d'en disputer avec lui ou de discriminer quelle pouvait être la part d'erreur ou de vérité. Ainsi, ce qui est important dans le cas de Thalès [...], c'est qu'il entendait s'appuyer sur des raisons susceptibles de convaincre quiconque se donnerait la peine de les examiner. Autrement dit, il invitait chacun à soumettre ses déclarations à l'épreuve de la critique. Il s'agit là d'un point important car, en nous référant aux origines de la science, nous entendons parler, non pas d'un soliloque, mais du début d'un dialogue."  
[Lalumia 1974]

La structure logique de ce dialogue dont parle Lalumia ressemble étonnamment à celle d'une conversation sur le mode paradoxal. D'ailleurs, le modèle de Thalès ne va pas tarder à être invalidé par Anaximandre avec des arguments qui nous paraissent tout aussi étonnants que le modèle lui-même, mais qui ont le mérite d'être des arguments. La condition première pour que le dialogue scientifique ait lieu, est que les interlocuteurs se considèrent, au niveau du principe, comme *égaux* en regard du sujet abordé : cette exigence est exactement la même que celle que nous avons posée au départ pour limiter le champ des conversations que nous prenions en considération (chap. I.1).

Ce qui semble avoir retardé l'arrivée, *a posteriori* inévitable, de la Science (notamment en occident), c'est bien l'interdiction imposée à l'immense majorité des individus par la société et la religion d'aborder certains sujets de manière critique.

C'est cette inhibition face aux problèmes qui dépassent notre quotidien qui mène encore de nos jours la plupart des individus à faire preuve de deux attitudes qui s'excluent mutuellement, et à passer de l'une à l'autre en fonction du problème posé : d'un côté, une attitude rationnelle (nous pourrions dire conversationnelle), que chacun adopte face aux problèmes concrets et sans gravité de la vie courante, et, d'un autre côté, une attitude a-critique qui se caractérise par l'adoption de modèles en bloc, sans volonté de remise en question, face à des sujets plus abstraits et moins palpables pour celui qui les considère. C'est ainsi qu'un commerçant trouvant qu'il manque 250F dans sa caisse ne manquera pas d'élaborer plusieurs modèles logiques pour expliquer que cette disparition a pu se produire en dépit de son étroite surveillance. Serait-ce

une erreur de comptabilité, le résultat d'un larcin ? Est-ce que par hasard quelque billet ne serait pas tombé derrière le comptoir ? Le commerçant va se lancer dans une recherche de type "scientifique" : il va chercher par terre, il va vérifier ses calculs, formuler d'autres hypothèses. Il va exposer l'incohérence de la disparition autour de lui, exposant ses différents modèles, en espérant que la conversation lui permettra d'en invalider certains, ce qui le fera progresser dans sa recherche. En revanche notre même commerçant, ayant à faire face à un grave problème de santé pour l'un de ses enfants, risque de ne pas adopter la même attitude rationnelle, et va peut être perdre tout sens critique au point de remettre la santé de son enfant entre les mains de guérisseurs qui ne peuvent offrir comme justification qu'une prétendue "autorité" comme celle à laquelle Lalumia fait allusion.

Cette attitude a-critique peut être adoptée lorsque l'individu est confronté à des domaines qu'il ne maîtrise absolument pas<sup>(4)</sup>. L'une des propriétés distinctives des modèles qui se soustraient à la critique est de s'intégrer dans une théorie très vaste, qui tente d'expliquer tout un domaine, et même parfois le monde, dans sa globalité, même si ce n'est pas dans le détail. Ces modèles décrivent une vérité non négociable et de portée générale. Nous sommes loin, avec cela, des modèles qui constituent le contexte logique de nos conversations, et qui peuvent être résumés en quelques formules logiques.

Il est vrai que quand on regarde l'ensemble des théories accumulées par les scientifiques en quelques siècles, on ne peut qu'être impressionné par le fait que ces théories ont tendance à se compléter comme les pièces d'un puzzle, le tout constituant ce que beaucoup vénèrent comme la Vérité scientifique, qu'il serait là aussi malvenu de remettre en question à la légère. C'est regarder les choses d'un peu loin, car vue de près et au quotidien, l'activité scientifique ressemble au contraire à une sorte de discussion passionnée, qui peut confiner parfois à l'empoignade, et où le souci de respecter une vérité établie n'effleure pas les protagonistes, bien au contraire, car rien ne stimule plus le chercheur que la perspective d'invalider un modèle reconnu [Kuhn 1962].

Cette exigence de remise en question continuelle a permis à la Science d'atteindre un fantastique degré de précision dans ses prévisions. Mais la rançon à payer pour cette efficacité se retrouve dans la perte d'un concept auquel nous tenons tous naturellement beaucoup : la vérité. La vérité, que nous voudrions complète et dépourvue d'ambiguïté, même si elle reste parfois provisoirement hors d'atteinte, n'existe plus dans le cadre de la Science. Ou plutôt, cette vérité est devenue locale dans l'espace et dans le temps, comme elle peut l'être dans les conversations.

Au début d'une conversation sur le mode paradoxal, les interlocuteurs savent bien qu'une partie de leur connaissance est erronée, et qu'il faut la modifier. Ils espèrent que la conversation leur permettra de résoudre le paradoxe, grâce à la modification d'un fait ou d'un lien logique. Si ce changement leur permet de rendre leur connaissance cohérente, il tiendront peut-être la connaissance modifiée pour vraie. Mais ils savent que la portée d'une telle vérité est limitée et provisoire.

La connaissance scientifique n'a pas un statut épistémologique qui la différencie qualitativement de la connaissance quotidienne acquise dans une conversation. Sa valeur vient de ce que son élaboration a impliqué de très nombreux individus, qui ont mis en oeuvre des moyens matériels et intellectuels considérables pour fabriquer leurs invalidations. Mais chacun de ses modèles reste en principe aussi critiquable que les déclarations de votre voisine. En revanche, les autres types de "connaissance" que nous avons évoqués, ceux qui se soustraient à la critique du vulgaire, ont, c'est le moins qu'on puisse dire, un statut épistémologique radicalement différent. Ils atteignent à la vérité parfaite, et le font savoir. Mais ils contiennent toujours une clause qui leur permet de se situer en dehors du champ de nos conversations quotidiennes, et on ne peut s'empêcher de penser qu'ils perdent ainsi en chemin une sorte de légitimité.

## 5. Le jeu conversationnel

L'étude de la logique des conversations, telle que nous l'avons menée dans la première partie de ce document, n'a pas besoin d'autre justification que celle d'essayer de faire progresser la connaissance que nous

---

<sup>4</sup> Il suffit de penser à la façon dont sont acceptées l'astrologie, l'homéopathie, l'influence de la lune sur le temps, etc.

---

avons du comportement humain. Nous avons montré que le comportement conversationnel était soumis à des contraintes, et nous comprenons maintenant que c'est à ce type de contraintes que la communication humaine doit une grande partie de sa richesse, de même qu'un jeu peut devoir l'infinie diversité de ses stratégies à un ensemble de règles posées au départ.

Nous avons voulu montrer que cette nouvelle façon de percevoir le langage était particulièrement féconde. Nous avons évoqué ses implications cognitives, psychologiques, technologiques, et même épistémologiques. Il existe d'autres conséquences théoriques et pratiques d'une telle approche qui reconnaît l'aspect spontané du traitement logique des connaissances. Simplement, leur évocation ici serait prématurée.



# Introduction. Introduction : appliquer les règles de la conversation à l'apprentissage

## 1. Une étude cognitive de la détection d'incohérence

Il serait possible de voir le travail présenté dans ce document comme une étude de la *détection d'incohérence* dans la communication spontanée et dans l'apprentissage. Ce serait résumer de manière un peu réductrice, mais cela aurait le mérite de rappeler le point fondamental sur lequel nous avons voulu fonder l'originalité de notre étude. Pourtant, cette description n'est pas conforme à la chronologie de la recherche qui a abouti au modèle présenté ici. Comme l'indique ce qui suit, l'importance centrale de l'incohérence dans la communication, dans l'apprentissage et dans la modélisation du fonctionnement cognitif ne s'est imposée qu'à la suite d'une rencontre entre plusieurs préoccupations.

## 2. Histoire d'un modèle

### 2.1. Comprendre la structure des conversations

J'ai commencé l'étude des conversations, il y a quelques années, avec une motivation de type éthologique. Il s'agissait d'étudier un comportement caractéristique de l'espèce dans ses conditions naturelles d'expression. J'ai donc enregistré des conversations de type familial, pour tenter ensuite d'en démonter les mécanismes.

Comme beaucoup d'autres chercheurs, j'ai commencé par repérer la polarité des arguments (orientation des arguments les uns par rapport aux autres), mais cela restait très largement insuffisant et n'avait aucun caractère prédictif. A la suite d'un examen plus poussé des quelques extraits dont je disposais à l'époque, la description sous forme logique (notamment l'incompatibilité) et les modalités (improbable et indésirable) ont fini par s'imposer comme outils naturels de description. L'organisation de tous les extraits autour d'une relation problématique (incohérence, improbabilité, (in)désirabilité) est alors devenue manifeste. A partir de là, de nombreux aspects de la structure des conversations se sont éclairés [Dessalles 1985].

Ce type de modélisation faisant appel à la logique n'était pas "à la mode" dans les années 80, d'autant que l'on a *a priori* tendance à considérer la conversation spontanée comme le comportement humain complexe par excellence. Même la compétence logique se voyait refusée aux êtres humains<sup>(1)</sup> !

---

<sup>1</sup> Il semble que toute une littérature se soit développée sur une soi-disant incapacité des humains à traiter des relations logiques. Or tout cela ne semble reposer que sur une interprétation peu parcimonieuse de tests comme celui des quatre cartes, dit test de Wason. Toutes les théories *contextuées* de la compétence logique (par opposition aux théories comme celle de Piaget qui supposent une compétence logique *générale*) reposent sur ce point sur une base expérimentale plutôt étroite : schémas pragmatiques de

Par ailleurs, rien ne semble devoir limiter ce que les interlocuteurs sont susceptibles de dire. Leur liberté semble absolue, et le contenu de leurs répliques est lié aux aléas de la situation d'énonciation. De plus, nous ne disposons même pas d'outils satisfaisants pour prédire le sens d'une phrase. Alors de là à prédire les répliques conversationnelles ! Dans ces conditions, il était difficile d'accepter qu'un ensemble limité de contraintes, comme celles que j'analyse dans le chapitre I.1, pût exister et diriger à ce point notre comportement, sans que nous en soyons conscients. Pour ceux qui n'ont jamais étudié le phénomène conversationnel sous l'angle logique, ces contraintes semblent trop simples, trop systématiques, sans compter qu'elles présupposent une compétence logique générale !

## 2.2. *Prévoir ce qui va être dit*

Il était donc nécessaire d'aller plus loin : non seulement le modèle que je propose "cadre bien" avec les données, mais il permet même de prédire le contenu et l'ordre des répliques à partir d'une connaissance logique. Pour le démontrer, j'ai développé PARADISE. Le programme PARADISE utilise en entrée une connaissance logique "locale" entièrement déclarative (*i.e.* un ensemble non ordonné de règles logiques), et reproduit en sortie la performance de l'interlocuteur. Le problème fondamental de la pragmatique du langage : expliquer pourquoi tel locuteur choisit de dire ce qu'il dit à un instant donné de la conversation, se retrouve ainsi ramené au problème, certainement beaucoup plus simple, qui consiste à isoler la connaissance utilisée par cet interlocuteur au moment de l'interaction.

## 2.3. *Enseigner en détectant des incohérences*

En parallèle, je menais des travaux en EAO (projet SAVANT1 et SAVANT2). J'ai compris rapidement que le fait de présenter une situation comme problématique, ainsi que le font systématiquement les interlocuteurs, était un moyen extraordinairement efficace pour capter l'attention de l'apprenant. J'ai expérimenté ce principe dans les séquences multimédia de SAVANT2 (sortes de "clips" didactiques). Mais ce n'est qu'après le développement de PARADISE que j'ai entrevu la possibilité de gérer tout un dialogue avec l'apprenant à partir de la seule gestion des incohérences. C'est ainsi que SAVANT3 a vu le jour [Dessalles 1990].

Ma motivation pour l'EAO provenait en fait d'un intérêt marqué pour les problèmes de l'apprentissage, notamment pour l'apprentissage des concepts. J'avais analysé au départ le problème de l'apprentissage comme un problème de chaîne de communication entre un enseignant et un apprenant. Si ce dernier ne comprend pas, c'est parce que la définition du concept est "noyée" dans un "bruit" formé d'éléments sans rapport avec ce qui doit être appris. Le but poursuivi dans SAVANT1 était d'offrir des définitions optimales de ce point de vue. Mais j'avais négligé un élément essentiel.

Devant les premiers échecs, j'ai compris l'importance du *contexte* pour l'apprentissage. J'ai donc introduit des éléments "narratifs" dans SAVANT1, et j'ai pu vérifier leur efficacité [Dessalles 1984b]. Mais ce n'est qu'en modélisant les conversations que j'ai compris quelle devait être la structure de ce contexte. Ce contexte devait être *problématique*. J'en ai donc tenu compte sur le plan technique, en créant SAVANT3. Mais ce n'est que beaucoup plus tard que j'ai pleinement compris les conséquences de ce constat sur le plan cognitif.

## 2.4. *Imaginer un cerveau capable de s'étonner*

La plupart des modèles de la cognition humaine soit ne traitent pas, soit échouent à expliquer, notre capacité (et notre propension) à détecter l'incohérence des situations que nous rencontrons. Les théories associatives, les théories à base de modèles mentaux, les théories sélectives, ne permettent pas de comprendre l'acharnement des interlocuteurs à rechercher une explication pour une situation qu'ils trouvent incohérente. Cette caractéristique du comportement humain, à elle seule, conduit à postuler l'existence d'un appareil conceptuel autonome. Une telle hypothèse est généralement rejetée *a priori*, car elle semble

---

raisonnement [Holland, Holyoak et al. 1986], détection de la tricherie [Cosmides 1989], schémas statistiques [Smolenski 1988], modèles mentaux [Johnson Laird et Byrne 1991]. Or la présence d'une compétence logique générale chez les humains ressort manifestement de l'étude des conversations, alors que le test de Wason est facilement interprétable au niveau linguistique (dans une langue comme le français, on signifie couramment une équivalence entre *A* et *B* par le fait de dire simplement "Si *A* alors *B*" sans mention supplémentaire).

neuralement invraisemblable. Mais grâce à la compréhension du rôle joué par la *détection* dans les phénomènes cognitifs [Dessalles 1992c], j'ai eu la chance de pouvoir montrer que ce préjugé était erroné : le fonctionnement de la "couche C" constitue un modèle neuronalement plausible de la détection d'incohérences.

### 3. L'avenir d'un modèle

L'approche que je propose est originale. Malheureusement, elle ne s'inscrit pas bien dans les cadres conceptuels classiques qui "dominent" dans les disciplines auxquelles elle pourrait être liée. Le tableau suivant résume la situation (au risque, peut-être, de la caricaturer) :

<i>Domaine abordé</i>	<i>Paradigme dominant</i>	<i>Concepts proposés</i>
Philosophie du langage	Actes de langage	Prise en compte des arguments Pas de prise en compte des acteurs
Analyse des conversations	Tours de parole	Prise en compte d'un contexte non observable. Prise en compte des arguments et non des tours de parole
Linguistique	Marqueurs d'orientation argumentative	Relations argumentatives hors de la langue
Pragmatique	Principe de coopération (Grice)	Répliques "destructives" (invalidations, banalisation, répliques antagonistes)
EAO	Enseignement de savoir-faire par essais-erreurs	Apprentissage de concepts
EIAO	Modélisation de l'apprenant	Dialogue dirigé par la contradiction
Théories de l'apprentissage	Apprentissage de nature statistique	Apprentissage par appariement
Modélisation neuro-mimétique	Extraction de régularités par réseaux connexionnistes	Détection par synchronisation neuronale Encodage spatio-ensembliste de la logique
IA - explication	Travail sur la trace d'un raisonnement	Reconnaissance de la problématique motivant la requête

Ce tableau montre que les concepts que j'ai été amené à proposer sont étrangers aux paradigmes qui sont actuellement dominants dans les différents domaines avec lesquels le modèle conversationnel interfère. Il y a même parfois opposition. L'étude de l'emploi du langage en situation (c'est-à-dire avec prise en compte du contexte) se retrouve ainsi marginalisée par les disciplines qui pourraient l'héberger. Cette situation est peut-être due à l'histoire du développement de ces disciplines. En revanche on peut concevoir ce type de recherche sur le langage comme un thème central de la Science Cognitive. C'est le point de vue que je défends.

La compréhension des mécanismes supérieurs du langage, leurs liens avec la pensée, avec l'apprentissage et avec la représentation matérielle des connaissances, tout cela est par essence au centre des préoccupations de la Science Cognitive. Quel thème de la recherche peut-il mieux concerner tout à la fois la psychologie, la linguistique, l'intelligence artificielle, le connexionnisme, voire (si notre modèle neuronal de la détection d'incompatibilité y apparaît comme plausible) les neurosciences ?

Dans l'introduction de ce document, j'ai indiqué quelques enjeux liés à un début de compréhension des phénomènes supérieurs du langage : les enjeux scientifiques, avec l'espoir de comprendre le comportement qui caractérise notre espèce, et les enjeux technologiques, pour que la communication entre humains et machines puisse approcher en complexité et en acceptabilité la communication inter-humaine. J'espère que ces perspectives amèneront les chercheurs concernés par ces questions à considérer les thèses exposées dans ce document. Il ne leur est pas demandé d'y adhérer directement. Comme tout chercheur, j'attends bien sûr de mes collègues qu'ils critiquent mon travail, qu'ils s'efforcent de le réfuter, qu'ils confrontent avec moi leurs idées sur le comportement conversationnel. Mais surtout, j'attends d'eux qu'ils mesurent comme moi les enjeux de cette méta-conversation.



**Références Bibliographiques 263**

**Index 271**

**Table des matières 273**



## Introduction : appliquer les règles de la conversation à l'apprentissage

### Liste des articles et ouvrages référencés dans le texte

- Ackley David H., Hinton Geoffrey E., Sejnowski T.J. (1985). A Learning Algorithm for Boltzman Machines. *Cognitive Science* 9, 1985, pp. 147-169
- Airenti G., Bara B.G., Colombetti M. (1989). Knowledge for Communication. In Taylor M.M., Néel F., Bouwhuis D.G. , *The Structure of Multimodal Dialogue*, North Holland, Amsterdam 1989, pp. 145-158
- Anderson John R. (1983). *The Architecture of Cognition*. Harvard University Press, Cambridge, USA 1983
- Anderson John R., Conrad Frederick G, Corbett A.T. (1989). Skill Acquisition and the LISP Tutor. *Cognitive Science* 13, 1989, pp. 467-506
- Andler Daniel (1987). L'apprentissage dans les Sciences Cognitives: approches théoriques. In , *Intellectica* n°2/3, A.R.C., Paris 1987, pp. 213-234
- Anscombe Jean-Claude, Ducrot Oswald (1988). *L'argumentation dans la langue*. Ed. Pierre Margada, Bruxelles 1988
- Ayel Marc, Rousset M-Christine (1990). *La cohérence dans les bases de connaissances*. Editions CEPADUES, Toulouse 1990
- Baker Michael (1991). An Analysis of Cooperation and Conflict in students' collaborative explanations for Phenomena in Mechanics. in Tiberghien A., Mandl H.: *Knowledge Acquisition in Physics & Learning Environments*, Springer Ver., 1991
- Baker Michael (1991). Modelling Negotiation in Intelligent Teaching Dialogues. In Elsom-Cook M., Moysse R.: *Knowledge Negotiation*, Paul Chapman Publ., Londres 1991
- Baker Michael (1992). Le rôle de la collaboration dans la construction d'explications. In *Actes des deuxièmes journées Explication du PRC-GDR-IA du CNRS*, INRIA, Sophia-Antip. 1992, pp. 25-40
- Barthélémy Jean-Pierre (1990). *Similitude, arbres et typicalité*, ENST 1990
- Bateson Gregory, Jackson Don D., Haley, Weakland (1956). Toward a Theory of Schizophrenia. *Behavioral Science*, vol. I, N° 4, 1956
- Bersini Hugues (1990). Before and After Breakdowns. In , *4ème colloque de l'A.R.C.*, ARC / INRIA, Paris 1990
- Berwick Robert C. (1991). From Rules to Principles in Language Acquisition: A View from the Bridge, in Powers d., Reeker L.: *Proceedings of MLNLO'91 (Machine Learning of Natural Language and Ontology)*, Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern 1991
- Bolt Richard H. (1968). Computer-Assisted Socratic Instruction. In Orr W.D., *Conversational computer*, J. Wiley, New York 1968
- Bonnardel Nathalie (1989). *L'évaluation de solutions dans la résolution de problèmes de conception*. INRIA - rapport de recherche n°1072, Paris 1989

- Bonnardel Nathalie (1990). L'effet du modèle de l'interlocuteur dans les dialogues de transfert de projet. In , *Congrès APPLICA 90*, Lille 1990
- Bransford John D., Stein B.S., et al. (1982). Differences in Approaches to Learning: an Overview. *Journal of experimental psychology: General*, vol.111,n° 4, 1982, pp. 390-398
- Bunt Harry (1991). Dynamic Interpretation and Dialogue Performance. In , *Second Venaco Workshop: The Structure of Multimodal Dialogue*, Pre-Proceedings, DRET, 1991
- Carbonell J., Michalski R., Mitchell T. (1983). An Overview of Machine Learning. in R.S. Michalski, J.G. Carbonell, T.M. Mitchell: *Machine Learning*, Palo Alto: Tioga, , pp. 3-23
- Cauzinille-Marmèche E., Mathieu J. (1988). Adapter les interventions tutorielles au modèle cognitif de l'étudiant. In Caverni Jean-Paul, Bastien Claude, Mendelsohn P., *Psychologie cognitive - Modèles et méthodes*, Guy Tiberghien - Presses Universitaires de Grenoble, Grenoble 1988, pp. 175-190
- Cerbah Farid (1992). Generating Causal Explanations: from Qualitative Models to Natural Language Texts. In Neuman Bernd, *Proceedings of ECAI92*, John Wiley, Vienne 1992, pp. 490-494
- Chalmers David J. (1990). Why Fodor and Pylyshyn were Wrong: the Simplest Refutation. *Twelfth Annual Conference of the Cognitive Science Society*, Cambridge 1990, pp. 340-347
- Chalmers David J., French Robert M., Hofstadter D.R. (1991). *High-Level Perception, Representation, and Analogy: A Critique of Artificial Intelligence Methodology*. CRCC Technical Report 49, Center for Research on Concepts and Cognition, Indiana University, 1991
- Changeux Jean-Pierre, Dehaene Stanislas (1989). Neuronal models of cognitive functions. *Cognition*, 33, 1989, pp. 63-109
- Chomsky Noam (1975). *Réflexions sur le langage*. Flammarion, Paris 1981
- Choukair Lamia (1993). Méthodologie statistique pour l'analyse de séries spatio-temporelles. *Thèse de Doctorat*, TELECOM-Paris 93-E-045
- Clancey William J. (1981). *The Epistemology of A Rule-Based Expert System: A Framework for Explanation*. STAN-CS-81-896, Stanford University, 1981
- Clancey William J. (1987). Methodology for Building an Intelligent Tutoring System. In Kearsley Greg P., *Artificial Intelligence & Instruction - Applications and Methods*, Addison-Wesley Publishing Company, Menlo Park, USA 1987, pp. 193-228
- Collins Allan (1976). Processes in Acquiring Knowledge. In Anderson, Spiro and Montague (eds.), *Schooling and Acquisition of Knowledge*, Erlbaum Assoc., Hillsdale NJ 1976
- Cooper L., Shepard R. (1985). Le retournement mental des objets. *Pour la Science* n°88, 1985
- Cosmides Leda (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* n° 31, 1989, pp. 187-276
- Coulthard Malcom (1977). *An Introduction to Discourse Analysis*. Longman, London 1985
- Crain Stephen (1991). Language Acquisition in the Absence of Experience. *Behavioral and Brain Sciences* 14, 1991, pp. 597-650
- Crystal David, Davy Derek (1975). *Advanced Conversational English*. Longman, Londres 1975
- Dawkins Richard (1982). *The Extended Phenotype - The Gene as the Unit of Selection*. W.H. Freeman & Co, Oxford 1982
- de Amorim S.G., Barthélemy J-P., Ribeiro C.C. (1990). *Clustering and Clique Partitioning: Simulated Annealing and Tabu Search Approaches*. Les cahiers du GERAD - G-90-35, 1990
- Dessalles jean louis (1984a). Présentation non linéaire de l'information didactique grâce au vidéotex. Congrès 'Les Grandes Ecoles et la Télématique', ENST Bretagne, Brest 1984
- Dessalles jean louis (1984b). SAVANT: l'enseignement assisté par télématique dans la formation des ingénieurs de l'ENST. *L'Echo des Recherches* n°117, 1984, pp. 67-76
- Dessalles jean louis (1984c). *Une Troisième voie pour l'assistance à l'enseignement supérieur: SAVANT*. Forum EAO-84, Ecole Sup. de Comm. de Lyon, 1984, pp. 35
- Dessalles jean louis (1985). Stratégies naturelles d'acquisition des concepts et applications E.A.O.. In *Actes du colloque COGNITIVA 85*, CESTA, Paris 1985, pp. 713-719
- Dessalles jean louis (1986). SAVANT 2, ou l'audiovisuel interactif à Sup'Telecom. *Les Cahiers du CEFI* n°15, 1986, pp. 46-49
- Dessalles jean louis (1990). Computer Assisted Concept Learning. In Norrie D.H., Six H.-W., *Lecture Notes in Computer Science 438 - Computer Assisted Learning*, Springer-Verlag, Berlin 1990, pp. 175-183

- Dessalles Jean-Louis (1991a): *Conversation Assisted Learning: The SAVANT3 Dialog Module*, in Forte Eddy N.: *Proceedings of Calisce'91*, Presses Polytechniques et Universitaires Romandes, Lausanne 1991, pp. 159-165
- Dessalles Jean-Louis (1991b). The simulation of conversations. In Kohonen Teuvo, Fogelman-Soulié Françoise, *COGNITIVA 90 - Proceedings of the Third Cognitiva Symposium (Madrid)*, North Holland, Amsterdam 1991, pp. 483-492
- Dessalles Jean-Louis (1992a). Biomimetic Use of Genetic Algorithms. In Männer R., Manderick B., *Parallel Problem Solving from Nature, 2*, North-Holland, Amsterdam 1992, pp. 127-135
- Dessalles Jean-Louis (1992b). L'incidence logique de l'interaction dans la communication d'informations. *Technologies Idéologies Pratiques Vol. X n° 2-4*, 1992
- Dessalles Jean-Louis (1992c). Les aspects cognitifs de l'émergence. In Amy B., Ducret J-J, Grumbach A., *Emergence dans les modèles de la cognition - Actes des journées*, TELECOM-Paris 92-S-003, Paris 1992, pp. 47-59
- Dessalles Jean-Louis (1992d). *Les contraintes logiques des conversations spontanées*. Rapport technique TELECOM-Paris 92-d-011, Paris
- Dessalles Jean-Louis (1992e). Model-Based Surprise and Explanation: a way to negotiate concepts. In Brezillon Patrick, *Proceedings of the ECAI-92 Workshop on Improving the Use of KBS with explanations*, Rapp. LAFORIA 92/21 Univ. Paris VI, Paris 1992, pp. 107-113
- Dessalles Jean-Louis (1992f). SAVANT3: un système d'EIAO fondé sur l'explication conversationnelle. In *Actes des deuxièmes journées Explication du PRC-GDR-IA du CNRS*, INRIA, Sophia-Antipolis, pp. 77-86
- Dessalles Jean-Louis (1992g). SAVANT: How to Help Engineers to Learn New Concepts. *European Journal of Engineering Education Vol.17 n°2*, 1992, pp. 189-194
- Dessalles Jean-Louis (1993a). Ancrage sans extraction de régularités: le mécanisme d'appariement symbolique. In *Formation des symboles dans les modèles de la cognition - IMAG/LIFIA & TELECOM-Paris 1993*
- Dessalles Jean-Louis (1993b). Détection collective. *Actes des journées de Rochebrune 93*, 1993
- Dessalles Jean-Louis (1993c). From I.T.S. to I.C.S. : Learning with an Intelligent Critic, not with a Tutor. *Conference on Computer Assisted Engineering Education CAEE'93*, Polytechnica University, Bucharest 1993
- Dessalles Jean-Louis (1993d). Usage naturel du langage: modèle, simulation et application à l'apprentissage. In Boussaid Omar, Broussaud Marcel, et al. : *Pluridisciplinarité dans les Sciences Cognitives*, Hermès, Paris 1993
- Dessalles Jean-Louis, Rajman Martin (1992). Concepts and Procedures in Engineering Education: Designing Specific Teaching aids. *International Conference on Trans-European Cooperation in Engineering Education*, Prague 1992
- Dessalles Jean-Louis, Meyers Pascal (1993). Exemple d'une simulation argumentée pour l'apprentissage de Prolog. *Actes des Troisièmes Journées EIAO de Cachan*, Eyrolles, Paris 1993
- Drew Paul, Holt Elisabeth (1990). Complaignable Matters: the use of idiomatic expressions in making complaints. In Conein Bernard, De Fornel Michel, Quéré L., *Les formes de la conversation*, CNET, Paris 1990, pp. 108-143 (T.1)
- Dubois Danièle (1991). *Sémantique et cognition, Concepts, catégories et typicalité*. Editions du CNRS, Paris 1991
- Ducrot Oswald (1969). Préface. In Searle John R., *Les actes de langage - Essai de philosophie du langage*, Hermann (Speech Acts, Cambridge University Press), Paris 1972
- Eibl-Eibesfeldt Irenäus (1967). *Ethologie - Biologie du comportement*. Naturalia et Biologia ed. scientifiques, Paris 1977
- Falzon Pierre (1991). Cooperative Dialogues. In *Distributed Decision Making: Cognitive Models for Cooperative Work*, J.Rasmussen & al. ed., J.Wiley, London 1991, pp. 145-189
- Favaron O., Maheo M., Fournier I. (1990). *Combinatoire et Algorithmique*. Orsay Plus, Univ. Paris Sud, Paris 1990
- Fischer Gerhard, Lemke Andreas C., et al. (1991). The Role of Critiquing in Cooperative Problem Solving. *ACM Transactions on Information Systems, Vol.9, n°3*, 1991, pp. 123-151

- Fisher Douglas H., Pazzani Michael J., Langley Pat (1991). *Concept Formation - Knowledge and Experience in Unsupervised Learning*. Morgan Kaufmann Publ., San Mateo 1991
- Flowers Margot, McGuire Rod, Birnbaum L. (1982). Adversary Arguments and the Logic of Personal Attacks. In Lehnert Wendy G., Ringle Martin H., *Strategies for Natural Language Processing*, L.E.A., Hillsdale N.J. 1982, pp. 275-294
- Fodor Jerry A. (1975). *The language of thought*. Havard Univ. Press, Oxford 1975
- Fodor Jerry A. (1983). *La modularité de l'esprit*. ed. de Minuit, Paris 1986
- Fodor Jerry A., Pylyshyn Zenon W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 1988, pp. 3-71
- Fox Barbara A. (1987). Interactional Reconstruction in Real-Time Language Processing. *Cognitive Science* 11, 1987, pp. 365-387
- Freeman Walter (1991). La physiologie de la perception. *Pour la Science n° 162*, 1991, pp. 70-78
- Ganascia Jean-Gabr. (1987). AGAPE: De l'appariement structurel à l'apprentissage. In *Intellectica n°2/3*, A.R.C., Paris 1987, pp. 6-27
- Garitte Catherine, Dessalles Jean Louis (1993). (en préparation). 1993
- GENE (1993). Explication et négociation entre un système à base de connaissances et son utilisateur. *Journées Explication et Coopération Homme-Machine - CNAM*, Paris 1993
- Goldberg David E. (1989). *Genetic Algorithms in Search Optimization & Machine Learning*. Addison Wesley Publishing Company, USA 1989
- Goodwin C., Heritage J. (1990). Conversation analysis. *Annual Review of Anthropology* 19, 1990, pp. 283-307
- Grice H.Paul (1975). Logic and Conversation. In Cole Peter, Morgan Jerry L., *Syntax and Semantics volume 3: Speech Acts*, Academic Press, New York 1975, pp. 41-58
- Grize Jean-Blaise (1967). Historique, logique des classes et des propositions, logique des prédicats, logiques modales. In Piaget Jean, et al. , *Logique et connaissance scientifique*, Gallimard (La Pléiade), Paris 1976, pp. 135-289
- Grumbach Alain (1989). *Modèle symbolique de l'apprentissage de l'utilisation d'un dispositif*. , ENST 1989
- Grumbach Alain (1990). Modèle symboliconnexionniste de l'apprentissage. In Amy Bernard, Ducret Jean-J., Grumbach A., *Modèles symboliconnexionnistes - work-shop*, Actes du workshop, TELECOM-Paris 92-S-001, Paris 1990
- Grumbach Alain (1990). Un modèle d'apprentissage des besoins aux connaissances. *Cahiers pédagogiques n° 287 - Octobre*, 1990, pp. 55-57
- Grumbach Alain (1993). *Genèse du symbole artificiel*. A paraître, 1993
- Guez Stéphane (1990). A Computational Model for Arguments Understanding. *13th International Conference on Computational Linguistics (Coling 90)*, Vol 3, 1990, pp. 132-137
- Harnad Stevan (1990). The Symbol Grounding Problem. *Physica D* 42, North-Holland, 1990, pp. 335-346
- Haton Jean-Paul, et al. (1991). *Le raisonnement en Intelligence Artificielle*. InterEditions, Paris 1991
- Heritage John (1990). Interactional Accountability: a Conversation Analytic Perspective. In Conein Bernard, De Fornel Michel, Quéré L., *Les formes de la conversation*, CNET, Paris 1990, pp. 23-49 (T1)
- Hofstadter Douglas R. (1985). *Metamagical Themas: Questing for the Essence of Mind and Pattern*. Basic Books, New York 1985
- Hofstadter Douglas R., Mitchell Melanie (1988). Concepts, Analogies, and Creativity. *The 1988 Canadian Artificial Intelligence Conf.*, 1988
- Hofstadter Douglas R., McGraw Gary (1993). *Letter Spirit: An Emergent Model of the Perception and Creation of Alphabetic Style*. CRCC Technical Report number 68, Indiana Univ. 1993
- Hollan James D., Hutchins Edwin, Weitzman L (1987). STEAMER: An Interactive, Inspectable, Simulation-Based Training System. In Kearsley Greg P., *Artificial Intelligence & Instruction - Applications and Methods*, Addison-Wesley Publishing Company, Menlo Park, USA 1987, pp. 113-134
- Holland John, Holyoak Keith J., Nisbett, Thagard (1986). *Induction - Processes of Inference, Learning, and Discovery*. The MIT Press, Cambridge MA 1986
- Houdé O. (1990). Six-year-olds have taxonomic knowledge but fail to solve logical categorization problems: Context and versatility. *Archives de Psychologie*, 58, 1990, pp. 283-309
- Hubel D., Wiesel T. (1979). Les mécanismes cérébraux de la vision. *Pour la Science n° 25*, 1979

- Hutchins Edwin (1980). *Culture and Inference - A Tobriand Case Study*. Harvard University Press, Cambridge 1980
- Inhelder Bärbel, Piaget Jean (1979). Procédures et structures. *Archives de psychologie*, XLVII, 181, 1979, pp. 165-176
- Joab Michelle (1990). *Modélisation d'un dialogue pédagogique en langage naturel*. Thèse de Doctorat Paris VI n° 90-7, Paris 1990
- Johnson W. Lewis, Soloway Elliot (1987). PROUST: An Automatic Debugger for Pascal Programs. In Kearsley Greg P., *Artificial Intelligence & Instruction - Applications and Methods*, Addison-Wesley Publishing Company, Menlo Park, USA 1987, pp. 49-67
- Johnson-Laird P.N. (1983). *Mental Models*. Cambridge University Press, London 1983
- Johnson-Laird P.N., Byrne R.M.J. (1991). *Deduction*. Lawrence Erlbaum Associates, 1991
- Kodratoff Yves (1987). Is AI a Sub-Field of Computer Science - or is AI the Science of Explanations?. In Bratko Ivan, Lavrac Nada, *Progress in Machine Learning*, Sigma Press, Wilmslow, England 1987
- Kohonen Teuvo (1984). *Self-Organization and Associative Memory*. Springer Verlag, Berlin 1988
- Kuhn Thomas S. (1962). *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago 1962
- Lakoff George (1988). A suggestion for a Linguistics with Connectionist Foundations. *Proceedings of the 1988 Connectionist models Summer School - CMU*, 1988
- Lalumia Joseph (1974). De la science à la métaphysique et à la philosophie. *Diogène n°88*, Gallimard, Paris 1974
- Lee M.C. (1990). Designing an Intelligent Prolog Tutor. In Norrie D.H., Six H.-W., *Lecture Notes in Computer Science 438 - Computer Assisted Learning*, Springer-Verlag, Berlin 1990, pp. 420-431
- Lemaire Benoît (1992). Aspects constructifs de la production d'une explication: l'architecture Esmeralda. In , *Actes des deuxièmes journées Explication du PRC-GDR-IA du CNRS, INRIA, Sophia-Antip. 1992*, pp. 165-178
- Leyton Michael (1993). *Symmetry, Causality, Mind*. MIT Press, Cambridge 1992
- Locke John (1690). *An Essay Concerning Human Understanding*. Dent & Sons, London 1961
- Lorenz Konrad (1974). *Evolution et modification du comportement - l'inné et l'acquis*. Bibliothèque Payot, Paris 1974
- Lorenz Konrad (1978). *Vergleichende Verhaltensforschung - Grundlagen der Ethologie*. Springer Verlag, Wien 1978
- Lumer Erik D., Huberman Bernardo A. (1992). Binding Hierarchies: A Basis for Dynamic Perceptual Grouping. *Neural Computation* 4, 1992, pp. 341-355
- Mager R. (1977). *Comment définir des Objectifs Pédagogiques*. Bordas, Paris 1977
- Mann W., Thompson S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *TEXT* 8(3), 1988, pp. 243-281
- Markman Ellen M. (1990). Constraints Children Place on Word Meanings. *Cognitive Science* 14, 1990, pp. 57-77
- Maynard Douglas W. (1990). Perspective Display sequences in conversation. In Conein Bernard, De Fornel Michel, Quéré L., *Les formes de la conversation*, CNET, Paris 1990, pp. 181-212 (T.2)
- McTear Michael F. (1985). *Children's Conversation*. Basil Blackwell, Oxford 1985
- Medin Douglas L., Wattenmaker William D., Michalski R.S. (1987). Constraints and Preferences in Inductive Learning: An Experimental Study of Human and Machine Performance. *Cognitive Science* 11, 1987, pp. 299-339
- Medin Douglas L. (1989). Concepts and Conceptual Structure. *American Psychologist* 44, 12, 1989, pp. 1469-1481
- Mehler Jacques (1974). A propos du développement cognitif. In Morin E., Piattelli-Palmarini M.: *L'unité de l'Homme - 2 - Le cerveau humain*, Seuil, Paris 1974, pp. 38-49
- Michalski Ryszard S. (1983). Learning from Observation: Conceptual Clustering. In Michalski Ryszard S., Carbonell Jaime G., Mitchell T.M., *Machine Learning - An Artificial Approach*, Tioga Publ. Company, Palo Alto 1983, pp. 331-363
- Mitchell Melanie, Hofstadter Douglas R. (1990). The Emergence of Understanding in a Computer Model of Concepts and Analogy-Making. *Physica D* 42, North-Holland, 1990, pp. 322-334

- Moeschler Jacques (1985). *Argumentation et conversation - Eléments pour une analyse pragmatique du discours*. Hatier, Paris 1985
- Moeschler Jacques (1990). Conversation: cohérence et pertinence. In Conein Bernard, De Fornel Michel, Quéré L., *Les formes de la conversation*, CNET, Paris 1990, pp. 79-104
- Monod Jacques (1970). *Le hasard et la nécessité*. Seuil, Paris 1970
- Mooney Raymond J. (1991). Explanation-Based Learning as Concept Formation. In Fisher Douglas H., Pazzani Michael J., Langley Pat, *Concept Formation - Knowledge and Experience in Unsupervised Learning*, Morgan Kaufmann Publ., San Mateo 1991, pp. 179-205
- Morel Mary-Annick (1983). Vers une rhétorique de la conversation. *DRLAV n°29*, 1983, pp. 29-68
- Moyse R. (1989). Knowledge Negotiation Implies Multiple Viewpoints. In Bierman Dick, Breuker Joost, Sandberg J. , *Artificial Intelligence and Education*, I.O.S., Amsterdam 1989, pp. 140-149
- Naville Pierre (1942). *La psychologie du comportement*. Gallimard, Paris 1963
- Nicaud J-François, Nguyen Xuan Anh, Saidi M. et al. (1990). APLUSIX: un environnement d'apprentissage pour l'acquisition de connaissances stratégiques en résolution de problèmes. In Kohonen Teuvo, Fogelman-Soulié Françoise, *COGNITIVA 90 - Proceedings of the Third Cognitiva Symposium (Madrid)*, North Holland, Amsterdam 1991, pp. 357-364
- Nicolle Anne (1992). Compère-Robot. In Bourguine Paul, *Journées de Rochebrune: apprentissage, évolution, adaptation*, AFCET, 1992
- Norman M.A., Thomas P.J. (1991). Informing HCI design through conversation analysis. *International Journal of Man-Machine Studies* 35, 1991, pp. 235-250
- Nowé Ann, Jonckers V. (1991). The Use of Very High Level Clichés in PATAT: a Program Analysis Tool Using Algorithm Transformation. In Forte Eddy N., *Proceedings of Calisce'91*, Presses Polytechniques et Universitaires Romandes, Lausanne 1991, pp. 209-216
- Ohlsson Stellan (1991). Interview, by J.Sandberg and Y.Barbard. *AICOM vol 4, n° 4*, 1991, pp. 137-144
- Papert Seymour (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, New York 1980
- Paris Cécile L. (1992). Systèmes Experts explicatifs. In , *Actes des deuxièmes journées Explication du PRC-GDR-IA du CNRS*, INRIA, Sophia-Antip. 1992, pp. 3-23
- Petrie-Brown Alison (1989). Intelligent Tutoring Dialogs: The Structure of an Interaction. In Bierman Dick, Breuker Joost, Sandberg J. , *Artificial Intelligence and Education*, I.O.S., Amsterdam 1989, pp. 195-201
- Piaget Jean (1967). *Biologie et connaissance*. Gallimard, Paris 1967
- Piaget Jean (1932). *Le jugement moral chez l'enfant*. P.U.F., Paris 1969
- Piaget Jean, Inhelder Bärbel (1947). *La représentation de l'espace chez l'enfant*. P.U.F., Paris 1972
- Piaget Jean (1979). La psychogénèse des connaissances, et sa signification épistémologique. In Piattelli-Massimo, Palmarini , *Théories du langage - Théories de l'apprentissage*, Seuil, Paris 1979, pp. 53-64
- Piatelli- Massimo, Palmarini (1988). Evolution, selection and cognition: From 'learning' to parameter setting in biology and in the study of language. *Cognition* 31, 1988, pp. 1-44
- Plunkett Kim, Marchman Virginia (1990). *From Rote Learning to System Building*. CRL Technical Report 9020, Univ. of California, San Diego 1990
- Poitrenaud Sébastien, Richard Jean-Fran., et al. (1990). La description des procédures: leur décomposition hiérarchique et leur rôle dans la catégorisation des objets. In *4ème colloque de l'A.R.C.*, ARC / INRIA, Paris 1990, pp. 331-348
- Polity Y., Francony J.M., et al. (1990). Recueil de dialogues homme-machine en langue naturelle écrite. *Les Cahiers du CRISS 17*, 1990
- Popper Karl R. (1959). *La logique de la découverte scientifique*. Payot, Paris 1982
- Quinlan J.Ross (1983). Learning Efficient Classification Procedures and their Application to Chess and Games. In Michalski Ryszard S., Carbonell Jaime G., Mitchell T.M., *Machine Learning - An Artificial Approach*, Tioga Publ. Company, Palo Alto 1983, pp. 463-482
- Racah Pierre-Yves (1990). Modelling argumentation and modelling with argumentation. *Argumentation*, 4, Bruxelles 1990
- Racah Pierre-Yves (1991). Inférence et langue. In *Langues et langage: carrefour des sciences de la cognition*. Documents de la Troisième Ecole d'été de L'ARC, 1991

- Rätz Thomas, Lusti Markus (1992). Explanation Strategies: realization in a tutor for database normalization. In Brezillon Patrick, *Proceedings of the ECAI-92 Workshop on Improving the Use of KBS with explanations*, Rapp. LAFORIA 92/21 Univ. Paris VI, Paris 1992, pp. 47-56
- Reichman Rachel (1985). *Getting Computers to talk like you and me*. MIT Press, Harvard 1985
- Reichman Rachel (1989). Convention versus Intention. In Taylor M.M., Néel F., Bouwhuis D.G. , *The Structure of Multimodal Dialogue*, North Holland, Amsterdam 1989, pp. 121-134
- Richard J-François (1983). *Logique du fonctionnement et logique d'utilisation*. INRIA - rapport de recherche n° 202, Rocquencourt 1983
- Richard J-François, Poitrenaud S., Tijus C. (1990). Modèle de description des procédures pour la représentation des connaissances des utilisateurs. In , *Congrès APPLICA 90*, Lille 1990
- Rock Irvin, Palmer Stephen (1991). L'héritage du gestaltisme. *Pour la Science n° 160*, 1991
- Rosch E. (1978). *Principles of Categorization*. in Rosch & Lloyd: *Cognition and Categorization* - Hillsdale, Erlbaum, 1978, pp. 27-48
- Sabah Gérard (1988). *L'intelligence artificielle et le langage*. Hermès, Paris 1988
- Safar Brigitte, Berthault Pascale, Sylvestre J. (1992). Place des explications dans la conception d'une interface intelligente entre une base de données et un usager. *12èmes Journées Internationales Avignon'92*, Avignon 1992
- Savage L.J. (1972). *The Foundations of Statistics*. Dover Publications Inc., New York 1972
- Schank Roger C., Lehnert Wendy (1979). The Conceptual Content of Conversation. *International Joint Conference on Artificial Intelligence*, Tokyo 1979, pp. 769-771
- Schank Roger C. (1980). Failure-Driven Memory. *Cognition and Brain Theory 4*, (1), 1980, pp. 41-60
- Schank Roger C. (1984). *The Cognitive Computer - On language, Learning and Artificial Intelligence*. Addison Wesley publishing Cy,, 1984
- Schank Roger C., Edelson Daniel J. (1989). Discovery Systems. In Bierman Dick, Breuker Joost, Sandberg J. , *Artificial Intelligence and Education*, I.O.S., Amsterdam 1989, pp. 236-237
- Searle John R. (1969). *Les actes de langage - Essai de philosophie du langage*. Hermann (Speech Acts, Cambridge University Press), Paris 1972
- Shannon Claude E. (1948). Mathematical Theory of Communication. *Bell Systems Technical Journal 27*, 1948, pp. 379-423,623-656
- Shatz Carla (1992). La maturation du cerveau. In , *Pour La Science n° 181*, , 1992, pp. 50-59
- Skinner Burrhus F. (1969). *La révolution scientifique de l'enseignement*. Dessart, Bruxelles 1969
- Sleeman Derek (1982). Assessing aspects of competence in basic algebra. In Sleeman Derek, Brown J.S., *Intelligent Tutoring Systems*, Academic Press, Londres 1982, pp. 185-199
- Sleeman Derek, Kelly A.E., Martinak R. (1989). Studies of Diagnosis and Remediation with High School Algebra Students. R.D. Ward, J.L. Moore - *Cognitive Science 13*, 1989, pp. 551-568
- Smolensky Paul (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences 11*, 1988, pp. 1-74
- Sperber Dan, Wilson Deirdre (1986). *La pertinence*. Les Editions de Minuit, Paris 1989
- Stepankova Olga (1992). An Introduction to Qualitative Reasoning. In Marik V., Stepankova O., Trapp R., *Advanced Topics in Artificial Intelligence*, Lecture Notes in Artificial Intelligence 617, Springer Verlag, 1992, pp. 404-418
- Stevens Albert, Collins Allan, Goldin Sarah E. (1979). Misconceptions in student's understanding. *Int. J. Man-Machine Studies 11*, 1979, pp. 145-156
- Stubbs Michael (1983). *Discourse Analysis*. Basil Blackwell, Oxford 1983
- Swartout William R. (1983). XPLAIN: a System for Creating and Explaining Expert Consulting Programs. *Artificial Intelligence 21*, 1983, pp. 285-325

- Tannen Deborah (1984). *Conversational Style - Analyzing Talk Among Friends*. Ablex Publishing Corporation, Norwood 1984
- Taylor I. (1989). Analyzing Conversation (in Three Languages). In Taylor M.M., Néel F., Bouwhuis D.G. , *The Structure of Multimodal Dialogue*, North Holland, Amsterdam 1989, pp. 273-286
- Thompson Patrick W (1987). Mathematical Microworlds and Intelligent Computer-Assisted Instruction. In Kearsley Greg P., *Artificial Intelligence & Instruction - Applications and Methods*, Addison-Wesley Publishing Company, Menlo Park, USA 1987, pp. 83-110
- Thompson Kevin, Langley Pat (1991). Concept Formation in Structured Domains. In Fisher Douglas H., Pazzani Michael J., Langley Pat, *Concept Formation - Knowledge and Experience in Unsupervised Learning*, Morgan Kaufmann Publ., San Mateo 1991, pp. 127-161
- Trognon Alain, Brassac Christian (1988). Actes de langage et conversation. In *Intellectica n°6*, A.R.C., Paris 1988, pp. 211-232
- Trognon Alain (1988). *Problème d'élaboration de la cohérence conversationnelle*. conf. du 15-03-88 au LIMSI, Orsay, Paris 1988
- Trognon Alain (1990). Relations intersubjectives dans les débats. dans A. Berrendonner, H. Parret: *L'interaction Communicative*, Ed. Peter Lang, Paris 1990, pp. 195-213
- VanLehn Kurt (1990). *Mind Bugs: The Origins of Procedural Misconceptions*. MIT Press, Cambridge, USA 1990
- Veale Tony, Keane Mark T. (1992). Conceptual Scaffolding: Using Metaphors to Build Knowledge Structures. In Neuman Bernd, *Proceedings of ECAI92*, John Wiley, Vienne 1992, pp. 174-178
- Visalberghi Elisabeth (1989). Les singes et l'outil: une question d'intelligence?. *La Recherche* 214, 1989, pp. 1277-1280
- Visalberghi E., Frigaszy D. (1990). . In Parker S., Gibson K., *Language and Intelligence in Monkeys and Apes: developmental perspectives*, Cambridge University Press, 1990
- Visalberghi E., Frigaszy D. (1990). . In Parker S., Gibson K., *Language and Intelligence in Monkeys and Apes: developmental perspectives*, Cambridge University Press, 1990
- Vuchinich Samuel (1980). Logical Relations and Comprehension in Conversation. *Journal of Psycholinguistic Research*, Vol.9, N°5, 1980
- Walbridge Charles T. (1989). Genetic Algorithms: What Computers Can Learn from Darwin. *Technology Review* - January, 1989, pp. 47-53
- Watson John B. (1913). Psychology as the Behaviorist views it. *Psychological Review*, vol XX, 1913
- Watzlawick P., Helmick Beavin J., Jackson D.D. (1967). *Une logique de la communication*. Seuil, Paris 1972
- Wenger Etienne (1987). *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann Publishers, INC., Los Altos, Cal.,USA 1987
- Zeki Semir (1992). Les images visuelles. In , *Pour La Science n° 181*, , 1992, pp. 60-68

# Index

- accord, 44
- algorithmes génétiques, 89
- ancrage des concepts, 148
- antagoniste (réaction), 40
- appareil conceptuel, 147
- appariement, 97; 156; 160; 173
- apprentissage
  - appariement, 97
  - compilation, 94
  - concepts et procédures, 191
  - convergence, 119
  - difficulté, 111
  - empirisme, 87; 142
  - extraction, 101
  - harmonie, 116
  - imitation, 97
  - indifférent, 116
  - innéisme, 95
  - maturation, 96
  - optimisation, 88
  - par extraction, 89
  - par la conversation, 199
  - relations logiques, 122
  - stabilité, 118
  - théorie des fusibles, 98
- arguments
  - analyse, 75
  - calcul, 70
  - synthèse, 75
- atol, 222
  
- banalisation, 34
  - dans le mode (in)désirable, 39
  - dans le mode improbable, 34
- Behaviorisme, 90; 186
- bidétermination, 151; 153; 158; 160; 162; 163; 164
  
- cassure, 14; 19; 20; 27
- clarification de contexte, 45
- clause de référence, 30; 73; 207
  
- clause négative, 22; 70
- clause saturée, 30
- co-réactions, 43
  - co-étonnement, 43
  - co-lamentation, 43
  - co-réjouissance, 43
- compositionnalité, 177
- concepts
  - abstraction, 141
  - ancrage, 148
  - appareil conceptuel, 147; 171
  - compréhension, 200
  - concepts et incohérence, 138
  - définition, 139
  - encodage implicite, 155
  - formation par extraction analogique, 153
  - intégration perceptive, 143
  - intension, 140
  - les concepts comme des détecteurs, 149
  - mémorisation, 201
  - négation, 144; 150; 177
  - sémantique, 150
  - vs procédure, 191
- connaissances
  - définition, 139
  - conceptuelles vs procédurales, 190
- connexionnisme, 92; 108; 167
- constitueance, 175
- constructivisme, 100; 120
- contexte logique, 19
- contraintes
  - cognitives, 238; 245
  - conversationnelles, 13; 14; 27; 30; 52; 245
- conversation, 13
- coopération, 34; 42; 59
- corrélation
  - seuil de corrélation, 101
- couche C, 171

- critique, 227; 235
- détecteur  
collectif, 169  
primitif, 174
- mémorisé, 174
- détection, 97; 149; 168  
de proximité, 153  
des incohérences, 171; 257
- EBL, 76; 162
- empreinte, 97
- épistémologie, 252
- explication, 33; 144; 162; 225
- explicitation, 230
- extraction, 101; 143; 152
- fusibles  
théorie des fusibles, 98; 122
- génération de textes, 250
- Gestaltisme, 122
- grammaire conversationnelle, 47; 48
- Grice, 15; 42; 59
- harmonie, 114; 123
- Hofstadter, 141
- imitation, 97
- implicite, 155
- incohérence, 138; 171; 257
- indifférence, 102; 143; 153; 157  
définition, 107
- Information Conversationnelle, 26
- intension, 140
- invalidation  
dans le mode (in)désirable, 38  
dans le mode improbable, 33  
dans le mode paradoxal, 31  
directe, 31  
explication, 33  
indirecte, 32; 57; 229
- isotropie, 107
- logique conversationnelle, 21
- mode  
(in)désirable, 25  
improbable, 23  
paradoxal, 24
- moteur, 213
- négation, 144; 150; 177
- paradise, 72
- pertinence, 15; 35; 65  
espace de pertinence, 28  
logique, 64  
maxime gricéenne, 60  
optimale, 61
- pertinence logique, 27; 40; 45; 52
- Piaget, 100; 120; 179
- pragmatique, 13; 19; 33; 57; 249  
conversationnelle, 64
- problématique, 14; 27; 29; 45; 63; 165;
- 200
- procédures, 191
- réaction antagoniste, 40
- redondance, 222
- relativité, 107
- robustesse, 113
- savant<sup>1</sup>, 188; 217
- savant<sup>2</sup>, 189; 217
- savant<sup>3</sup>, 194; 203; 235
- SBL, 159
- Sperber et Wilson, 13; 20; 61
- sphère égocentrée, 35
- styles conversationnels, 249
- suggestion, 38
- sujet de la conversation, 30
- supervision, 101
- symétrie, 102
- synchronisation neuronale, 169
- systematicité, 145; 146; 167; 174
- tournoi narratif, 37
- troubles conversationnels, 248
- unicité matérielle, 145
- why, 193; 194

# Table des matières

Introduction : appliquer les règles de la conversation à l'apprentissage.....	3
1. Comprendre les aspects supérieurs du langage .....	3
2. Etudier la communication spontanée.....	4
3. Apprendre les connaissances qui peuvent être communiquées.....	4
4. Enseigner les concepts par une conversation avec un ordinateur.....	5
Introduction à la partie I .....	9
Modèle de la communication spontanée.....	9
I.1. Les contraintes logiques des conversations spontanées.....	11
1. Introduction.....	11
2. L'usage naturel du langage naturel .....	11
3. Objectif de l'étude des conversations : découvrir des contraintes .....	13
4. Contraintes et conversation spontanée.....	14
5. Les conversations spontanées sont-elles logiques ? .....	15
6. Décrire les conversations au moyen de la logique.....	18
7. Introduire un nouveau sujet dans la conversation.....	23
8. La deuxième intervention .....	31
9. Le déroulement des conversations.....	47
10. Implications cognitives.....	52
11. Conclusion .....	54
Annexe : quelques règles de réécriture logiques .....	55
I.2. Aspects de la pragmatique conversationnelle .....	57
1. Introduction.....	57
2. Le problème de la communication en situation.....	57
3. Grice : la communication vue comme une coopération .....	59
4. Sperber et Wilson : le principe de pertinence optimale.....	61
5. Comparaison et critique des théories de Grice et de Sperber et Wilson .....	62
6. Vers une pragmatique conversationnelle.....	64
7. Conclusion .....	65
I.3. Un modèle computationnel de l'argumentation.....	67
1. Introduction.....	67
2. Analyse et synthèse des arguments conversationnels.....	67
3. Le problème de la reconstruction d'une conversation.....	68
4. De la connaissance statique à l'argumentation.....	70
5. Fonctionnement de la reconstruction.....	74
6. Les stratégies de paradise .....	78
7. Limitations actuelles de paradise.....	84
8. paradise en tant que modèle cognitif .....	85
9. Perspectives .....	85
Annexe : exemples de reconstruction de dialogues par paradise.....	87

Introduction à la partie II .....	85
Apprendre pour converser .....	85
II.1. Modèles de l'apprentissage .....	87
1. Introduction.....	87
2. Empirisme ou Instruction.....	87
3. Apprentissage par optimisation .....	88
4. Apprentissage par extraction de régularités.....	89
5. Innéisme et maturation.....	94
6. Apprentissage par appariement.....	95
7. Constructivisme de Jean Piaget .....	98
8. Analogie, instanciation, généralisation.....	99
9. Classification des modes d'apprentissage .....	99
10. Conclusion .....	102
II.2. Limites de l'apprentissage indifférent .....	105
1. Introduction.....	105
2. Indifférence d'un mécanisme d'apprentissage.....	106
3. Difficulté de l'apprentissage indifférent.....	110
4. Convergence, indifférence et théories de l'apprentissage.....	119
5. Indifférence et apprentissage de relations logiques.....	123
6. Conclusion .....	128
Annexe 1 : Les isométries de l'hypercube .....	131
Annexe 2 : taille minimum d'un jeu d'exemples non ordonné.....	133
Annexe 3 : Apprentissage statistique.....	135
II.3. L'apprentissage des connaissances conceptuelles dirigé par l'expérience .....	137
1. Introduction : de la conversation à la cognition .....	137
2. La nature des connaissances conversationnelles .....	138
3. Les concepts sont-ils des perceptions intégrées ?.....	141
4. Représentation cognitive des concepts logiques.....	143
5. La nécessité d'ancrer les concepts logiques.....	148
6. La formation de concepts ancrés .....	151
7. Exemples d'appariement guidé pour la formation des concepts.....	158
8. Apprentissage de concepts pertinents.....	164
9. Conclusion .....	165
II.4. Plausibilité cognitive d'un appareil conceptuel autonome .....	167
1. Introduction.....	167
2. Traitement symbolique vs. traitement connexionniste .....	167
3. Implémentation neuronale de la détection .....	168
4. Modélisation neuronale de la détection d'incohérence .....	171
5. Points forts du modèle .....	174
6. Limites du modèle .....	179
7. Conclusion : "ramener la logique au neurone" .....	180
Introduction à la partie III.....	183
Expliquer par l'incohérence .....	183
III.1. Les systèmes d'aide à l'apprentissage des connaissances .....	185
1. Introduction.....	185
2. La problématique de l'EAO .....	185
3. Les approches "classiques" de l'EAO .....	186
4. Connaissances conceptuelles et connaissances procédurales.....	190
5. Des systèmes qui enseignent des concepts .....	193

6. Des systèmes qui enseignent des savoir-faire procéduraux simples .....	195
7. Des systèmes qui enseignent des savoir-faire procéduraux complexes.....	197
8. Conclusion .....	197
III.2. SAVANT 3 : enseigner par la conversation .....	199
1. Introduction.....	199
2. L'importance pédagogique d'un contexte argumentatif.....	199
3. Principe de fonctionnement de savant3 .....	203
4. L'acceptabilité d'une conversation avec savant3 .....	207
5. Conclusion .....	208
Annexe 1 : Exemples d'interaction avec savant3.....	209
Annexe 2 : Le moteur de clauses paradoxales.....	213
III.3. Un système auteur conversationnel pour SAVANT3.....	217
1. Introduction.....	217
2. Un cours basé sur savant3.....	217
3. L'environnement auteur de savant3 .....	220
4. Un système conversationnel pour l'auteur .....	220
5. Conclusion .....	224
III.4. Explication et critique argumentatives dans les systèmes à base de connaissances	225
1. Introduction.....	225
2. La problématique de l'explication dans les SBC .....	226
3. Anatomie de l'explication conversationnelle .....	228
4. L'importance de la pertinence des explications pour les SBC.....	231
5. L'explication comme une critique argumentée .....	235
6. Conclusion : le respect des contraintes cognitives de l'utilisateur.....	238
Introduction à la partie IV .....	243
Une nouvelle approche du traitement du langage et de la modélisation cognitive	243
IV.1. Des perspectives nouvelles.....	245
1. Introduction.....	245
2. Contraintes conversationnelles et contraintes cognitives : un lien étroit	245
3. Analyse et traitement des aspects supérieurs du langage .....	249
4. Comportement conversationnel et comportement scientifique .....	252
5. Le jeu conversationnel.....	255
IV.2. Conclusion.....	257
1. Une étude cognitive de la détection d'incohérence.....	257
2. Histoire d'un modèle .....	257
3. L'avenir d'un modèle.....	259
Références bibliographiques.....	263
Liste des articles et ouvrages référencés dans le texte.....	263
Index .....	271
Table des matières .....	273



## Résumé de thèse

# *Modèle cognitif de la communication spontanée, appliqué à l'apprentissage des concepts*

**jean louis Dessalles**

TELECOM Paris - Dépt. Informatique - 46 rue Barrault, 75013 Paris  
Tel.: (33 1) 45 81 75 29 - Fax: (33 1) 45 81 31 19 - E-mail: dessalles@enst.fr

La conception de systèmes intelligents acceptables par les humains, notamment dans le domaine de l'enseignement, passe par une meilleure compréhension des processus cognitifs liés au langage. Malheureusement, le comportement langagier, si particulier à notre espèce, reste encore très mal compris.

Cette étude présente une *théorie logique de la conversation*, prolongée par des hypothèses minimales sur le *fonctionnement cognitif* sous-jacent. Cette théorie trouve une application dans un système d'*Enseignement Assisté par Ordinateur* capable de soutenir une argumentation conversationnelle avec l'étudiant.

## **Théorie problématique de la conversation**

L'étude d'un corpus de conversations naturelles nous a permis de mettre en évidence un phénomène assez spectaculaire : toute conversation est organisée autour d'une relation "problématique". Une telle relation peut être une incohérence (un phénomène bizarre), une improbabilité (un fait "incroyable"), une (in)désirabilité (un fait contrariant ou au contraire particulièrement positif). Lorsque la relation problématique n'apparaît pas clairement (par ex. en cas de malentendu), on observe des *cassures* du type "Et alors", "Pourquoi dis-tu ça". Toutes les répliques de la conversation doivent avoir un lien *logique* défini avec cette relation problématique.

Nous observons donc que les individus imposent des contraintes très fortes à leurs interlocuteurs, au point qu'il est possible dans certains cas de prédire le contenu des répliques. C'est ce que nous avons montré en reconstituant des conversations réelles avec notre programme PARADISE.

## **Théorie problématique de l'apprentissage**

Tout au long de notre vie, nous accumulons des connaissances qui nous permettent par exemple de juger de la cohérence de toute situation évoquée dans la conversation. Comment s'effectue cet apprentissage ?

Première solution : les individus réalisent une *extraction* de l'ensemble des situations possibles pour un contexte donné à partir de l'expérience des situations similaires vécues (par ex. par un mécanisme de type statistique). Nous réfutons un tel mécanisme "extractif" en montrant qu'il est beaucoup trop inefficace.

Deuxième solution : les individus manipulent réellement des connaissances de type logique et acquièrent de nouvelles connaissances *à l'occasion de la perception d'une relation problématique*. Nous suggérons que les relations problématiques ont une représentation mentale. Nous proposons une structure connexionniste originale, particulièrement simple et biologiquement plausible, capable de détecter les situations incohérentes.

## **Théorie problématique de l'enseignement**

Notre thèse est donc que les processus conversationnels sont à la base de l'apprentissage des connaissances conceptuelles. Pour explorer les conséquences de cette hypothèse sur le plan pratique, nous avons développé un système d'EIAO, SAVANT3, qui est capable d'argumenter avec l'étudiant sur des sujets

techniques. Sa stratégie consiste à placer l'étudiant dans une situation problématique, en lui montrant que ses déclarations sont contradictoires. A elle seule, cette stratégie permet au système de conduire automatiquement un dialogue pertinent à partir d'une connaissance totalement déclarative très limitée (une quinzaine de règles par dialogue).

L'objectif de notre étude est ainsi de montrer à quel point il est important de bien mesurer les contraintes qui limitent notre comportement conversationnel. L'enjeu porte à la fois sur la compréhension du comportement linguistique, sur la compréhension des processus cognitifs spécifiquement humains, et sur la conception de principes d'interaction humain/machine qui ouvrent la voie à de nouvelles applications.

