II.3. L'apprentissage des connaissances conceptuelles dirigé par l'expérience

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[1]. 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 rare), mais cela est autre

[1] 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.

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[2]. 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 explicitables, 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 founit 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 conférera 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] 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.

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 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[3] servent à *évoquer* (*déclencher*) des concepts déjà forgés, ou *déclenchent* la formation de nouveaux concepts[4]. 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[5].

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[6]. La sensibilité à une incompatibilité se traduit par un étonnement durable, qui ne peut se résoudre que par une invalidation. Il nous semble donc

---

[3] 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); jaloux(A,C)*).

[4] 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.

[5] 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 !

[6] 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 !

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 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 qualitativelent 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(7) :

> "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(8).

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.

---

7 Nous avons donné une citation plus complète de ce passage au début du chapitre II.1.

8 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 consituer 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) :

$$\begin{array}{lccc} & \textit{perceptions} & & \textit{concepts} \\ \mathcal{M}_p : & \{P_1, P_2, \ldots P_n\} & \rightarrow & C \end{array}$$

$$\begin{array}{lccc} & \textit{concepts} & & \\ \mathcal{M}_c : & \{C_1, C_2, \ldots C_m\} & \rightarrow & C' \end{array}$$

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 provennaient 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_α 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[9].

Ce qui importe ici est que *non C* a une représentation mentale, et que mécanisme $C \rightarrow$ *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 :

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

alors *non* p_i 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[10] ; Lille ne serait pas une ville très importante ; Lille serait beaucoup plus éloignée que 60 km ; Lille serait fléchée[11]). Ce phénomène est systématique : *non* p_i sera éventuellement "combattu", mais son énoncé ne provoquera

---

[9] 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 $C \rightarrow$ *non C*.

[10] Comme St Quentin en Yvelynes.

[11] C'est cette dernière explication qui s'est révélée être la bonne : Lille était fléchée en Flamand, sous le nom de Rijsel !

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 : $\mathcal{R}([q, p_1, p_2 \ldots p_n]) = 1$ ssi *non q* est une invalidation directe de $[p_1 \,\&\, p_2 \,\&\, \ldots\, p_n] \Rightarrow \mathbf{F}$. $\mathcal{R}$ vaut systématiquement *1* sur $\mathcal{V} = \{[q|L] \text{ t.q. } q \in L\}$ [12].

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 systématicité 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 systématicité 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[13]. 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 systématicité, 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 aissi leur *prise en compte* des concepts, comme indiqué schématiquement sur la figure 1.

---

12 Nous utilisons la notation Prolog : $[q|L]$ est la liste résultant de l'adjonction de l'élément q en tête de la liste L.

13 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 systématicité dans l'application d'un mécanisme logique comme la recevabilité d'une explication (cf. le "bon sens" de Descartes).

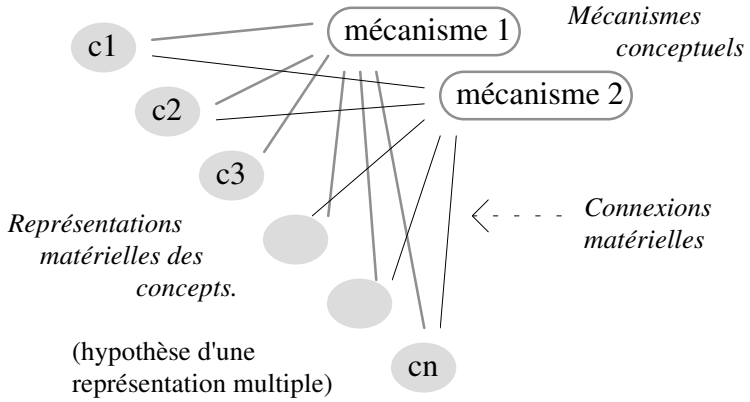


figure 1 : Représentation schématique de la problématique de l'unicité matérielle. Certains mécanismes conceptuels ne peuvent être que matériellement uniques. Mais si les représentations matérielles des concepts étaient, elles, physiquement distinctes, comme indiqué sur la figure, alors les liens qui 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 avions 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 $[p_1 \,\&\, p_2 \,\&\, \ldots\, p_n] \Rightarrow \boldsymbol{F}$ est une incompatibilité ressentie, $\mathcal{R}([q, p_1, p_2 \ldots p_n])$ a une valeur booléenne (recevable / non recevable) *pour tout concept q* connu du sujet. La figure 2 résume la situation.

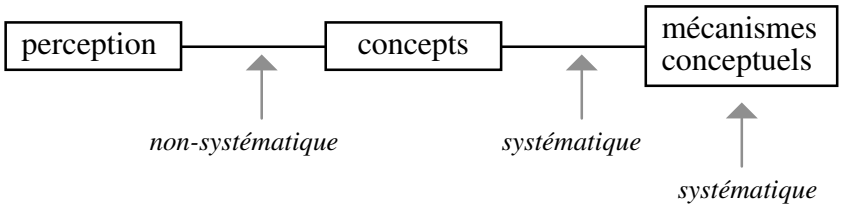


figure 2: représentation schématique des aspects systématiques dans le traitement cognitif des concepts. La systématicité se situe au niveau de certains mécanismes conceptuels, ainsi qu'au niveau de la prise en compte des concepts par ces mécanismes. En revanche, l'évocation des concepts par les perceptions 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 mécanismes conceptuel systématique ? Nous ne voyons qu'une seule solution :

Tous les concepts logiques sont hébergés par un seul système et partagent la meme 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 systématicité. On ne peut imaginer que la *prise en compte* d'un concept, dans un test comme celui que 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).

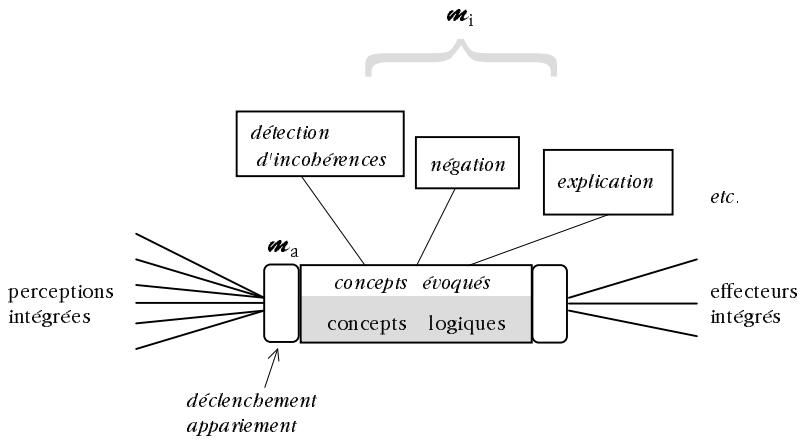


figure 3: schéma de l'appareil conceptuel censé abriter les représentations matérielles de tous les concepts. Il est relié par des connexions multiples (non systématiques) aux appareils perceptifs (perceptions intégrées) et aux appareils effecteurs intégrés (pour le déclenchement des actions). Les mécanismes 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 systématicité 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 systématicité 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[14].

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 avions 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].

---

[14] 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.

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 leur contenu. Par exemple une explication *non q* dans le contexte de l'incompatibilité $[p_1 \,\&\, p_2 \,\&\, \ldots\ p_n] \Rightarrow \boldsymbol{F}$ 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 systématicité, 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(15) :

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* (non C)
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(16) 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(17) 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.

---

15 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.

16 (voir note 3 p. 140)

17 Par exemple, un canari est plus typique (et plus vite reconnu) comme un oiseau que ne l'est un pinguoin. 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 systématicité 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 : $\{s_1, s_2, \ldots s_n\} \rightarrow s$.

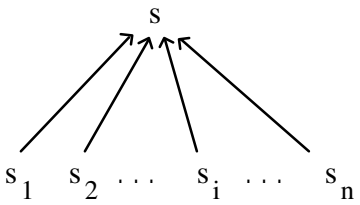


figure 4 : illustration schématique d'une extraction de type analogique. La représentation *s* est extraite à partir d'un ensemble de représentations s_i auxquelles elle ressemble.

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*[18], 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

$$\forall \{s_1, \ldots s_K\} \in E^K \quad \frac{1}{K}\sum_{i=1}^{K} d_o(s, s_i) \ll N \quad \text{avec} \quad s = g(s_1, \ldots s_K)$$

$d_O(\,)$ 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 :

---

[18] 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 ξ un seuil caractéristique du détecteur. $d(\ )$ et ξ peuvent dépendre des $\{s_i\}$.

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 $s = g(s_1, \ldots s_K)$ 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 :

$$\forall \{s_1, \ldots s_K\} \in E^K \quad \forall \rho \text{ isométrie dans E} \quad g(\rho(s_1), \ldots \rho(s_K)) = \rho(g(s_1, \ldots s_K))$$

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 $\{s_1, \ldots s_K\} \rightarrow D_{g(s1 \ldots sK)}$ est un mécanisme d'apprentissage indifférent si :

$$\text{pour toute isométrie } \rho, \quad D_{g(\rho(s_1), \ldots \rho(s_K))} \circ \rho = D_{g(s_1, \ldots s_K)}$$

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 $g(s, \ldots s) = s$, 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 rédhibitoires.

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(19), 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 $\{s_i\}$ 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 systématicité 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_i 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_i, 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(20) 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 :

---

19 Les concepts ne consituent certainement pas des "bonnes formes", au sens de la psychologie gestaltiste (cf. discussion chap. II.1).

20 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 $\propto$ est encodé implicitement dans une représentation *s* si il existe une fonction *h* telle que $\propto = h(s)$. Nous dirons aussi que $\propto$ est encodé implicitement dans le détecteur de proximité D_S s'il est encodé implicitement dans *s*.

Paramètre strict : Un paramètre $\propto = h(s)$, encodé implicitement dans une représentation *s*, est un paramètre *strict* si il existe une projection τ de E dans E, différente de l'identité, telle que $h \circ \tau = h$. Cela signifie que $\propto$ 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 $\propto = h(s)$ un paramètre strict encodé implicitement dans *s*. Alors il existe s^o et une isométrie ρ_o de *E* dans *E* telle que

$$\rho_o(s^o) \neq s^o \quad \text{et} \quad \forall s \; h(\rho_o(s)) = h(s).$$

En effet, il existe par hypothèse, $\propto$ étant un paramètre strict, une projection τ 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* $\rho_o(x) = x \oplus (x_o \oplus \tau(x_o))$, où $\oplus$ désigne le "ou" exclusif. On a bien :

$$\rho_o(s^o) = \tau(x_o) \oplus x_o \oplus \tau(x_o) = x_o \neq s^o$$

car $\tau(x_o) \oplus \tau(x_o) = 0$, et

$$h(\rho_o(s)) = h(\tau(\rho_o(s))) = h(\tau(s \oplus x_o \oplus \tau(x_o))) = h(\tau(s) \oplus \tau(x_o) \oplus \tau(x_o))$$

car τ 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 $\propto = h(s)$ un paramètre strict encodé implicitement dans *s*. D'après le lemme, il existe s^o et une isométrie ρ_o tels que

$$\rho_o(s^o) \neq s^o \quad \text{et} \quad \forall x \; h(\rho_o(x)) = h(x)$$

Nous allons montrer par l'absurde que D_S ne peut pas détecter $\propto$. 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 Δ_s du paramètre $\propto$ codé implicitement dans *x*.

Comme $D_{s^o}(x)$ est issu d'un mécanisme analogique indifférent, nous avons la propriété suivante :

$$\forall \rho \text{ isométrie } \forall x \quad D_{s^o}(x) = D_{\rho(s^o)}(\rho(x))$$

En particulier pour ρ_o :

$$\forall x \quad D_{s^o}(x) \quad = D_{\rho_o(s^o)}(\rho_o(x)) = \Delta_{\rho_o(s^o)}(h(\rho_o(x)))$$

$$= \Delta_{\rho o(s^o)}(h((x))$$

$$= D_{\rho o(s^o)}(x)$$

Or cette conclusion $\forall x \quad D_{s^o}(x) = D_{\rho o(s^o)}(x)$ est absurde dans la mesure où D_s est un détecteur de proximité et que $s^o \neq \rho_o(s^o)$. 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[21].

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 :

[21] 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.

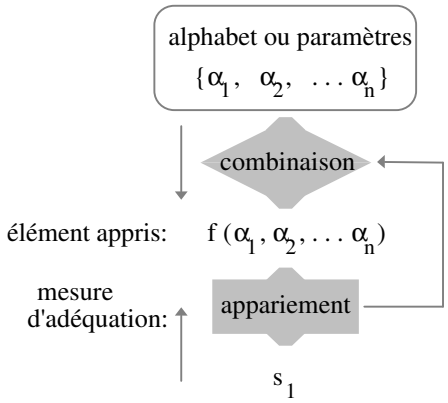


figure 5 : illustration schématique d'un apprentissage par appariement. Les objets α_i sont donnés au départ. Le système crée des combinaisons $f(\alpha_1, \alpha_2, \ldots \alpha_n)$ de ces objets pour qu'elles soient en *adéquation* avec la situation s_1 à REprésenter. Le résultat de l'apprentissage est une combinaison adéquate.

Le mécanisme d'appariement : $s_1 \rightarrow f(\alpha_1, \alpha_2, \ldots \alpha_n)$ obéit à un principe fondamentalement différent de celui de l'extraction analogique. L'élément appris provient d'un "alphabet" limité d'éléments préexistants α_i, généralement à la suite d'une combinaison $f(\ )$. Cet élément appris $f(\alpha_1, \alpha_2, \ldots \alpha_n)$ 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[22]. 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 α_i indépendants de toute expérience. Ces α_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

[22] En communications numériques, les éléments α_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.

détecteur conceptuel est alors en place. Si le deuxième chausse-pied que nous voyons est rouge alors que 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 bidétermiantion (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 :

SBL	
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 [23].

> 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.

[23] 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 :

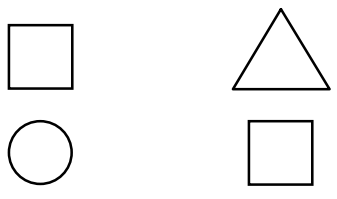
Ces objets sont structurés, c.à.d. qu'il ont des composants qui sont eux-mêmes des objets. L'apprentissage des concepts correspondant à de tels objets suppose la prise en compte des relations entre les composants d'un même objet. Par exemple la description d'un concept peut être : "il y a trois objets, *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. *a posteriori* 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 appariable, 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) :

EBL	
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 :
repose_sur(X,Y) & plat(Y) ⇒ stable(X)
peut_être_saisi(X) & léger(X) ⇒ portable(X)
concave(Y) & partie_de(Y,X) & vers_le_haut(Y) ⇒ récipient(X)
partie_de(Y,X) & forme(Y, demi-tore) & adapté_aux_doigts(Y) & rigide(Y) ⇒ peut_être_saisi(X)
isolant(Y) & fait_de(X,Y) & petit(X) ⇒ peut_être_saisi(X)
isolant(céramique)
isolant(porcelaine)
petit(X) ⇒ léger(X)
diamètre(X) > 2cm & diamètre(X) < 12cm ⇒ adapté_aux_doigts(X)
stable(X) & portable(X) & récipient(X) ⇒ tasse(X)

description de l'exemple :
appartient_à(obj1, Fred) & léger(obj1) & couleur(obj1, rouge) & partie_de(h1, obj1) & rigide(h1) & diamètre(h1) = 3cm & forme(h1, demi_tore) & repose_sur(obj1, b1) & plat(b1) & concave(c1) & partie_de(c1, obj1) & vers_le_haut(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 :

léger(X) & partie_de(H, X) & rigide(H) & diamètre(H) > 2cm & diamètre(H) < 12cm & forme(H, demi_tore) & repose_sur(X, B) & plat(B) & concave(C) & partie_de(C, X) & vers_le_haut(C) ⇒ tasse(X)

Un autre exemple aurait pu conduire à :

petit(X) & fait_de(X,Y) & isolant(Y) & repose_sur(X, B) & plat(B) & concave(C) & partie_de(C, X) & vers_le_haut(C) ⇒ 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") :

G003(H, X) ⇔ partie_de(H, X) & rigide(H) & diamètre(H) > 2cm & diamètre(H) < 12cm & forme(H, demi_tore)

Nous avons maintenant :

G001(X) ⇔ léger(X) & G003(H, X) & repose_sur(X, B) & plat(B) & concave(C) & partie_de(C, X) & vers_le_haut(C)

Ces nouveaux concepts G001, G002, G003 peuvent être utilisés directement pour analyser des objets et les apparier 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) ⇔ léger(X) & partie_de(H, X) & rigide(H) & diamètre(H) > 2cm & diamètre(H) < 12cm & forme(H, 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) ⇔ léger(X) & peut_être_saisi(X) & repose_sur(X, B) & plat(B) & concave(C) & partie_de(C, X) & 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_i, un objet sera détecté par *C* dans un contexte donné dès qu'il est détecté par chaque C_i 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*[24].

La formation et la mémorisation de nouveaux concepts exigent un contexte problématique.

[24] 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 systématicité 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.