

Chapitre 4 :
Rôle inférentiel des concepts

Introduction

Pour de nombreux auteurs, les représentations conceptuelles sont avant tout le support des inférences. Par rapport aux questions abordées dans le chapitre précédent, il s'agit d'un changement de problématique. Dans une perspective behavioriste, un concept n'a d'autre caractérisation que les comportements observables qui lui sont associés : catégoriser ou raisonner d'une certaine manière dans un certain contexte. Dans une perspective représentationnaliste, la question est de savoir de quelles manières les représentations conceptuelles peuvent avoir un rôle épistémique, comme celui de provoquer des inférences. Dans quelle mesure les concepts, de par leur forme, sont-ils responsables de ces inférences ?

Les pensées évoquées par la perception ou la communication langagière peuvent causer l'évocation d'autres pensées, de manière plus ou moins systématique. Cet enchaînement entre pensées opère sur un médium qui est supposé être celui des concepts. Certaines configurations conceptuelles, par exemple l'idée I_1 que "le chien a mordu le garçon", entraînent d'autres configurations conceptuelles, par exemple l'idée I_2 que "le garçon a mal". Ces enchaînements, le plus souvent, ne doivent rien à la forme des expressions langagières qui évoquent ces pensées. Or, ils ne sont pas dus au hasard. L'idée I_1 a plus de chances d'entraîner l'idée I_2 dans un cerveau humain que, par exemple, une idée I_3 selon laquelle "le nombre 17637 n'est pas premier". Il est donc naturel de rechercher des mécanismes permettant d'expliquer le fait que les inférences se fassent dans certaines directions privilégiées. Il est également naturel de supposer que ce sont, en dernier ressort, les concepts qui, de par leurs propriétés causales, sont responsables des inférences que l'on produit à partir d'eux. Si l'on inverse la perspective, l'étude des inférences que les êtres humains produisent devient un bon moyen de caractériser les concepts.

La question que nous voulons explorer dans ce chapitre concerne l'étendue des tâches inférentielles que le système conceptuel prend en charge. S'il existe un système inférentiel indépendant, doté d'un ensemble de représentations et de mécanismes permettant la production du raisonnement, alors le système conceptuel en tant que tel peut être minimal. Prenons l'exemple d'une inférence classiquement qualifiée de déductive. On peut imaginer qu'une connaissance de la forme *tout A est B* serait disponible dans un module de raisonnement, tandis que le rôle du système conceptuel serait limité à des tâches comme la reconnaissance d'occurrences, par exemple *x est A*. Dans une telle conception, de nouvelles connaissances, comme la connaissance déduite *x est B*, pourraient être produites par des mécanismes propres au module de raisonnement. À l'opposé, on peut concevoir que le système conceptuel prenne en charge l'ensemble des opérations de raisonnement. Ainsi, la déduction de *x est B* à partir de *x est A* pourrait être entièrement due au fait que le concept *B* et le concept *A* entretiennent un rapport structurel. Le rôle inférentiel des concepts devient alors prépondérant. Les concepts, dans un tel schéma, ne servent pas seulement de base aux raisonnements, ils en constituent la cause. Ainsi, un jugement de catégorisation comme *x est B*, qui n'est pas issu d'une reconnaissance directe, serait dû au rôle inférentiel des concepts eux-mêmes.

La motivation principale pour conférer un rôle inférentiel aux concepts est liée à l'idée qu'il existe des inférences nécessaires, dont la validité serait due à la structure des concepts que de telles inférences mettent en jeu. Dans une première section, nous examinons cette position, ainsi que certaines des critiques auxquelles elle a donné lieu, notamment en ce qui concerne la possibilité de prédire cette partie du raisonnement par le seul emploi de définitions. Dans une deuxième section, nous abordons une critique plus fondamentale, concernant l'existence même d'une frontière entre inférences nécessaires et inférences

contingentes. Une telle critique débouche sur une vision globale du système conceptuel dans laquelle le rôle inférentiel d'un concept donné ne vient pas de sa seule structure, mais des liens qu'il entretient avec les autres concepts. Dans cette approche, le système conceptuel est conçu comme assez riche pour héberger l'ensemble des mécanismes de raisonnement. Nous évoquerons la critique principale adressée à un tel système, liée à son caractère holistique. Dans la troisième section du chapitre, nous abordons une manière radicalement différente d'expliquer le pouvoir inférentiel des concepts à l'aide de structures. L'idée n'est plus de doter les concepts d'une structure pour reproduire le raisonnement dans ce qu'il a de nécessaire, mais dans ce qu'il a de typique. La cause des inférences que l'on produit à partir d'un concept n'est pas entièrement contenue dans la structure de ce concept. Elle n'est pas non plus reportée dans la structure de l'ensemble du réseau conceptuel. Elle est due au caractère régulier ou exemplaire de ces inférences, représentées dans la typicité des traits qui caractérisent le concept.

4.1. Concepts et nécessité

Certaines inférences semblent emporter l'adhésion de tout être doté du bon sens cartésien. Ainsi, du constat exprimé par la phrase *X a convaincu Y que A*, un être rationnel qui possède le concept *CONVAINCRE* semble obligé de conclure avec certitude la phrase *Y croit que A*. Une manière d'expliquer le caractère nécessaire de cette conclusion consiste à dire qu'elle existe, en germe, dans le sens de la prémisse. Par contre l'inférence qui conduit à conclure une phrase comme *X y a consacré beaucoup de temps* présente un aspect contingent et dépend de diverses connaissances que l'on possède sur les individus *X* et *Y*. Comment distinguer ces deux types d'inférences et expliquer la nécessité apparente de la première ?

La nécessité de certains énoncés du langage semble être imposée par la seule compréhension du sens des mots qui les constituent. Le problème du sens, dans ce cas, se retrouve intimement lié à la question du jugement épistémique¹. Considérons les trois énoncés suivants et le jugement épistémique qu'on leur associe spontanément.

1. Un individu non marié n'est pas marié.
2. Un célibataire n'est pas marié.
3. Un adolescent n'est pas marié.

On perçoit intuitivement une démarcation entre les deux premières phrases et la troisième. Les deux premières présentent un caractère nécessaire qui manque à la troisième. La manière dont on justifierait ces phrases, dans la plupart des contextes, diffère. Si quelqu'un prétend ne pas accepter les énoncés 1 et 2, on commence par vérifier que cet interlocuteur a bien compris le sens de ces énoncés. Cette vérification, dans le cas du premier énoncé, se borne à tester la bonne compréhension des deux formes de négation. Dans le cas du deuxième énoncé, la seule vérification possible consiste à tester que l'interlocuteur connaît le sens du mot *célibataire* et son rapport avec le sens du mot *marié*. Dans les deux cas, le jugement épistémique semble lié, de manière mécanique, au résultat de la production et de la compréhension sémantique. Ce phénomène a donné lieu à des analyses philosophiques et a suscité la conception de formalismes logiques appropriés.

¹ Il est courant, chez de nombreux auteurs, d'employer les termes "vrai" et "vérité" pour désigner ce jugement épistémique, comme si ce jugement était évalué à l'aune d'une référence omnisciente et indiscutable. N'ayant pas besoin d'une telle hypothèse, nous essaierons d'éviter ces termes dans notre propre discours.

'L-true' is meant as an explicatum for what Leibnitz called necessary truth and Kant analytic truth. [...] A sentence \mathcal{F}_i is *L-true* in a semantical system S if and only if \mathcal{F}_i is true in S in such a way that its truth can be established on the basis of the semantical rules of the system S alone, without any reference to (extra-linguistic) facts. (CARNAP 1947 [9] p. 8)

Les deux premiers énoncés de notre exemple font partie de ces jugements analytiques. Leur acceptabilité ne dépendrait que du traitement sémantique du langage et serait du ressort de ce que nous appelons le système conceptuel. Le jugement épistémique concernant le troisième énoncé de l'exemple, en revanche, fait appel à des faits considérés comme conventionnels ou typiques², dépendant de la culture ou de notre expérience. Généralement, ce type de connaissance est considéré comme ne faisant pas partie du système conceptuel. L'emploi correct du concept ADOLESCENT ne requiert pas la connaissance selon laquelle un adolescent, typiquement ou conventionnellement, n'est pas marié. Les deux premiers énoncés de notre exemple sont remarquables car le jugement épistémique que leur compréhension déclenche n'a pas recours à ce type de connaissances considérées comme externes. Pour le premier, la seule capacité requise est la maîtrise du mécanisme de la négation. Pour le deuxième énoncé, le modèle sémantique doit expliquer le lien obligatoire qui existe entre le mot célibataire et l'expression non marié. Pour ce faire, il est naturel d'imaginer qu'il s'agit d'un lien structurel, par exemple en disant que le concept CÉLIBATAIRE, en tant que représentation, contient comme constituant le concept associé à l'expression non marié, par exemple en postulant que le concept CÉLIBATAIRE équivaut strictement à la définition "individu non marié". La particularité des jugements analytiques s'expliquerait alors par le fait que certains concepts contiennent d'autres concepts en tant que composants.

Suppose we understand some predicators in a given language; that is to say, we know which properties they express. Suppose, further, that we have experienced each of these properties; that is to say, we have, for each of them, found some things which, according to our observations, have that property. We can form compound predicators out of the given predicators with the help of logical particles. Then we understand a compound predicator because its meaning is determined by the meanings of the component predicators and the logical structure of the compound expression. It is important to notice that our understanding of a compound predicator is no longer dependent upon observations of any things to which it applies, that is, any things which have the complex property expressed by it. (CARNAP 1947 [9] p. 20)

De cette manière, le rôle inférentiel du concept lui serait conféré par sa définition. Dans le cas des inférences analytiques, la conclusion serait valide parce qu'elle se trouve en germe dans la structure du concept de départ. Il est tentant de rendre analytiques la plupart des inférences en augmentant les définitions, de manière à enfermer dans le concept lui-même, par exemple sous forme d'une conjonction, toutes les inférences qu'il peut produire.

² Si le lecteur considère que le sens du mot adolescent contient, de la même façon que le mot célibataire, l'idée qu'il s'agit d'un individu non marié, il suffit de le remplacer par le mot étudiant.

For example, the most common sense of the English noun “chair” can be decomposed into a set of concepts which might be represented by the [following] semantic markers [...]:

(object), (Physical), (Non-Living), (Artifact), (Furniture), (Portable),
 (Something with legs), (Something with a back), (Something with a seat), (Seat for one).

It is obvious that this analysis leaves out a considerable information. [...] [Another] way of thinking about semantic markers is as symbols that mark the components of senses of expressions on which inferences from sentences containing the expressions depend. They are, in this case, thought of as marking aspects of the logical form of the sentences in whose semantic representation they occur. The [first sentence] entails [the rest]:

There is a chair in the room.
 There is a physical object in the room.
 There is something nonliving in the room.
 There is an artifact in the room.
 There is a piece of furniture in the room.
 There is something portable in the room.
 There is something having legs in the room.
 There is something with a back in the room.
 There is a seat for one person in the room.

The semantic markers [listed above], which comprise a reading for “chair”, mark the elements in the sense of this word on which [these] inferences depend.

(KATZ 1972 [58] p. 40)

Dans cet exemple, le concept CHAISE se trouve doté d'une structure comportant un certain nombre de traits, chacun de ces traits pouvant donner lieu à des inférences dont la validité semble être aussi indiscutable que l'appartenance du trait au concept. On peut, bien entendu, imaginer que ces traits sont à leur tour composés d'autres traits, ce qui augmente la complexité de la structure. Les inférences considérées comme analytiques ne se résument toutefois pas à ce schéma d'extraction de traits. Elles peuvent concerner les arguments des prédicats, par exemple dans le cas des verbes transitifs.

A selection of causatives with circumstantial functions appears [below]:

Sue forced/pressured/tricked/talked Jim into singing.
 Sue got/forced/caused/coerced Jim to sing.
 [CAUSE ([SUE], [GO_{Circ}([JIM]_i, [TO_{Circ}([i SING])])])]
 Sue kept/restrained/prevented Jim from singing.
 [CAUSE ([SUE], [STAY_{Circ}([JIM]_i, [NOT AT_{Circ}([i SING])])])]
 Sue allowed/permitted Jim to sing.
 [LET([SUE], [GO_{Circ}([JIM]_i, [TO_{Circ}([i SING])])])]
 Sue released Jim from singing.
 [LET([SUE], [GO_{Circ}([JIM]_i, [FROM_{Circ}([i SING])])])]
 Sue exempted Jim from singing.
 [LET([SUE], [STAY_{Circ}([JIM]_i, [NOT AT_{Circ}([i SING])])])]

[...] The familiar inference patterns for CAUSE, LET, GO, and STAY appear here as usual. In [the first example], Sue's action results in Jim's coming to sing [...]. In [the second], Sue's action results in Jim's continuing *not* to sing. In [the third], Sue could have prevented Jim from singing, but she didn't, so Jim probably sang. In [the fourth], Sue was forcing Jim to sing, and now allowed him to stop; in [the fifth], Sue could have forced Jim to sing, but chose not to. In these last two, Jim ends up probably not singing.

(JACKENDOFF 1983 [49] p. 200)

Ces exemples illustrent le type de décomposition que l'on peut proposer pour le sens de verbes comme *forcer* ou *permettre*. Il s'agit dans chaque cas d'une "influence" exercée sur le cours d'un événement. L'événement est présenté comme le "passage" d'un individu vers une action, à l'aide de fonctions conceptuelles comme *ALLER* ou *RESTER*. L'influence est présentée comme le résultat de fonctions conceptuelles telles que *CAUSER* ou *LAISSER*. Les structures fonctionnelles des verbes sont appariées avec celles de leurs arguments, ce qui produit des structures plus complètes sur lesquelles peuvent agir des règles d'inférence. On peut ainsi inférer le fait que l'individu a fini par accomplir l'action en question ou non. Les inférences produites de cette manière présentent un caractère nécessaire, car elles sont, formellement, analytiques³.

L'hypothèse de l'existence de structures définitionnelles, internes au système conceptuel, ne saurait épuiser l'explication de l'ensemble des inférences possibles. En général, les auteurs acceptent qu'un certain nombre d'inférences fassent intervenir des mécanismes externes, et ne soient pas strictement liées à la structure des concepts.

It was mentioned earlier that determination of truth presupposes knowledge of meaning (in addition to knowledge of facts); now, cognitive meaning may be roughly characterized as that meaning component which is relevant for the determination of truth. The non-cognitive meaning components, although irrelevant for questions of truth and logic, may still be very important for psychological effects of a sentence on a listener, e.g., by emphasis, emotional associations, motivational effects. (CARNAP 1947 [9] p. 237)

Le corollaire implicite de ce type de concession est que, si on limite l'ambition à la reproduction du seul rôle épistémique des concepts, alors la démarche qui consiste à établir une équivalence entre chaque concept et une définition répond à la question. Le reste des phénomènes inférentiels, les effets d'emphase, les associations, *et cætera*, reste hors du système conceptuel proprement dit.

Le fait de définir chaque concept, de considérer que chaque concept est équivalent, du point de vue des inférences, à sa définition, constitue-t-il une solution acceptable au problème du jugement épistémique ? Il faut noter d'emblée qu'une partie essentielle du jugement épistémique est laissée de côté. Le système conceptuel, grâce aux structures qu'il héberge, serait responsable de tous les jugements épistémiques nécessaires et de ceux-là seulement. Cette explication par les structures n'épuise pas toutefois le jugement épistémique, qui puise le plus souvent ses justifications dans des connaissances factuelles ou des relations contingentes. Ainsi, lorsque nous jugeons que tel individu, étant étudiant, n'est pas marié, nous exploitons une régularité perçue dans notre environnement. Nous ne tirons pas ce jugement de la définition du concept *ÉTUDIANT*. Le caractère contingent de tels jugements épistémiques serait précisément dû au fait qu'il ne dépendrait pas des connaissances inhérentes à la signification. Il faut donc reconnaître l'existence de deux sources de connaissances : d'une part les connaissances proprement conceptuelles, enfermées dans les définitions, et d'autre part les connaissances contingentes, contenues dans une sorte de base de connaissances. Ces deux sources de connaissances seraient exploitées par le même mécanisme logique pour produire le jugement épistémique. Cette limitation du pouvoir des définitions, qui entraîne une complication du modèle, se double d'une difficulté concernant l'existence même de ces définitions censées reproduire le pouvoir inférentiel des concepts.

Nous pouvons schématiser ce qui précède par un ensemble à deux composants : (1) Un système conceptuel, consistant en un ensemble de structures ; (2) Un système de

³ Les deux types de structures définitionnelles cités précédemment sont donnés à titre d'exemples. Nous reviendrons longuement sur les types de structures attribués aux concepts et les mécanismes généraux utilisés pour expliquer les inférences que l'on peut produire à partir d'eux (CF. CHAPITRE 7).

raisonnement, comportant des mécanismes logiques capables d'exploiter les structures conceptuelles du premier système et hébergeant des connaissances contingentes, c'est-à-dire des liens contingents établis entre les concepts grâce à l'expérience. Ensemble, ces deux systèmes sont supposés expliquer le jugement épistémique.

Le point fort de l'approche réside dans l'explication élégante des inférences nécessaires. Grâce au remplacement des concepts par leur définition, ces inférences nécessaires sont rendues triviales. Si la signification du mot célibataire équivaut à la définition "individu et non marié", alors l'inférence Pierre est célibataire \models Pierre est non marié devient trivialement valide, puisqu'elle équivaut à l'inférence Pierre est un individu \wedge Pierre est non marié \models Pierre est non marié. Un mécanisme inférentiel logique standard, comportant notamment le schéma inférentiel $P \wedge Q \models Q$, suffit à expliquer la validité de cette inférence.

Malheureusement, le caractère esthétique de ce schéma n'est qu'apparent. Un premier problème concerne l'exigence de l'existence de définitions adéquates pour n'importe quel concept. Quelle définition substituer au concept ROUGE ? Assurément, le concept COLORÉ doit être inclus dans cette définition, puisque l'inférence X est rouge \models X est coloré est nécessairement valide. Cependant, cet élément de définition est insuffisant. Lorsqu'il s'agit par exemple de traduire l'inférence contingente si le feu est rouge, alors s'arrêter, on s'aperçoit que les concepts ROUGE et COLORÉ ont des comportements inférentiels différents. Une substitution par cette définition partielle, qui correspondrait à l'énoncé si le feu est coloré, alors s'arrêter, risque de produire des inférences erronées. Le problème vient de ce que l'on peut distinguer des concepts comme ROUGE, ORANGE ou VERT par leur comportement épistémique. Or, si ces concepts n'existent, du point de vue inférentiel, que sous la forme de définitions, et que leur définition ne comporte que le seul concept COLORÉ, leur différence, en ce qui concerne leur comportement épistémique, demeure inexplicée. Cette difficulté, évidemment, est liée au caractère incomplet de la définition retenue. Est-on capable de compléter la définition du concept ROUGE jusqu'à la lui rendre substituable ? La réponse est non (FODOR & GARRETT & WALKER & PARKES 1980 [38]). Quelle que soit la précision de la définition proposée⁴, il sera possible de trouver un contre-exemple, comme nous l'avons fait en mentionnant les concepts ORANGE et VERT. On pourra discriminer conceptuellement les concepts ROUGE et ROUGE₁, ce dernier étant le concept couvert par la définition. Dans la mesure où la discrimination est possible au niveau conceptuel, on s'expose à des erreurs si l'on substitue ROUGE₁ à ROUGE dans le calcul inférentiel.

Quite generally, if an informally valid argument turns on a definition, then there will be some clause that we can conjoin to the consequent which will make the corresponding bi-conditional true. Any informally valid argument which does not meet this condition can't be a definitional argument. [...] Even given the apparatus of definitional analysis, it looks as though some informally valid arguments can't be captured within the inferential apparatus of standard logic.

(FODOR & GARRETT & WALKER & PARKES 1980 [38] p. 272)

Si une définition peut être trouvée pour chaque concept, elle doit inclure les conséquents de toute inférence analytique non triviale (*informally valid argument*) effectuée à

⁴ L'exemple a été choisi par les auteurs cités précédemment pour qu'il soit d'emblée difficile de compléter la définition. Noter qu'une définition du concept ROUGE par une plage de longueurs d'onde serait grossièrement inadéquate. Elle est inadéquate d'un point de vue perceptuel puisque la perception de la couleur dépend significativement du contexte visuel, *id est* éclairement et zones voisines. Elle est inadéquate d'un point de vue conceptuel, puisque les délimitations conceptuelles ne sont pas asservies aux discriminations perceptives. Elle est inadéquate, enfin, d'un point de vue cognitif, puisque la valeur de longueur d'onde n'est pas supposée être disponible en tant qu'élément explicite au niveau du système conceptuel des individus.

partir du concept. L'impossibilité de terminer cette procédure, au moins pour certains concepts, fait que certains concepts, comme le concept ROUGE, restent sans définition. En conséquence, le projet de transformer toute inférence analytique non triviale en une inférence triviale en substituant la définition au concept est un échec. Il en résulte une contradiction : la logique standard ne peut expliquer les inférences analytiques non triviales, alors que leur analyticit     tait cens  e reposer sur un m  canisme logique. Dans ces conditions, sur quoi faire reposer l'analyticit   d'inf  rences comme Pierre est c  libataire \models Pierre est non mari   ou la chaise est rouge \models la chaise est color  e ?

L'id  e que l'on pourrait expliquer les inf  rences n  cessaires    l'aide de d  finitions est encore plus probl  matique en ce qui concerne les structures fonctionnelles propos  es pour d  finir les verbes. Selon cette approche, la d  composition des verbes jusqu'au niveau de composants simples comme CAUSE,   TAT,   V  NEMENT ou FORCE est cens  e expliquer la r  alisation de certaines inf  rences en   vitant de recourir    des r  gles contingentes. Dans un tel mod  le, l'inf  rence qui,    partir de la phrase A convainc B que X, conclut la phrase B croit X est valide car, premi  rement, l'assertion de d  part   quivaut, au niveau conceptuel,    la phrase A pousse B    croire X ; et deuxi  mement, il existe un m  canisme g  n  ral de raisonnement qui lie la repr  sentation conceptuelle de la phrase A pousse B    croire X    la repr  sentation conceptuelle de la phrase B croit X. Cette deuxi  me   tape est cens  e   tre une op  ration formelle sur les repr  sentations conceptuelles. Or, on voit bien que cette op  ration d  pend de la pr  sence, dans la repr  sentation conceptuelle de la pr  misse, du concept POUSSER. Elle pr  sente donc un caract  re contingent. Il ne s'agit pas d'une r  gle d'inf  rence g  n  rale qui, comme une r  gle formelle de la logique du type $P \wedge Q \models Q$, est susceptible de s'appliquer    n'importe quelle proposition. Il faut donc accepter l'existence, dans le syst  me de raisonnement, d'un certain nombre de r  gles contingentes comme celle qui prescrit que la r  alisation de l'  v  nement A pousse B    E entra  ne l'existence de l'  tat E. Malheureusement, cette exigence du projet d  finitionnel entra  ne sa ruine, puisque l'objectif poursuivi en tentant de rendre les inf  rences analytiques   tait de supprimer leur caract  re contingent.

[...] the idea that informally valid arguments will prove to be formally valid when couched at the semantic level is equivalent to the idea that only their logical form is relevant to determining the validity of semantic level arguments; and, as things now stand, there is simply no reason to believe that this is true. If it is not, then a reconstruction of informal validity may require an enrichment of the inferential apparatus of standard logic (*e.g.*, the incorporation of rules which govern the behavior of formulae containing words like "cause") even if it *also* assumes the existence of definitions.

(FODOR & GARRETT & WALKER & PARKES 1980 [38] p. 271)

La seule mani  re de sauver le projet consiste    consid  rer que le nombre de r  gles d'inf  rences qu'il faut ajouter au syst  me logique standard reste limit  . Cela revient    dire qu'il faut d  terminer un petit nombre de primitives, comme CAUSE,   TAT,   V  NEMENT, FORCE, susceptibles d'entrer dans des sch  mas inf  rentiels, au m  me titre que les connecteurs logiques comme ET, OU, IMPLIQUE⁵. C'est la t  che que se sont assign  e un certain nombre de s  manticiens (CF. CHAPITRE 5). Nous aurons l'occasion de mentionner de nombreuses autres critiques d'ordre technique    l'encontre de cette approche (CF. CHAPITRE 7). Nous consid  rons dans l'imm  diat une critique fondamentale, portant sur l'existence m  me de jugements analytiques non triviaux.

⁵ Noter que les connecteurs de la logique standard sont redondants. Il est par exemple possible d'exprimer l'ensemble des expressions de la logique des propositions    l'aide de la conjonction et de la n  gation. Il est m  me possible de les exprimer    l'aide du seul op  rateur d'incompatibilit  . L'enrichissement envisag   ici n'est donc pas anodin.

4.2. Concepts et connaissance

Le fait que les concepts puissent être considérés comme équivalents à des définitions semble un fait incontournable dès que l'on prend en compte le caractère nécessaire des jugements analytiques. Ainsi, on conclut qu'un célibataire est nécessairement non marié en raison de l'équivalence entre le concept CÉLIBATAIRE et sa définition "individu non marié". Or, cette équivalence définitionnelle ne va pas de soi.

There are those who find it soothing to say that the analytic statements of the second class reduce to those of the first class, the logical truths, by *definition*; 'bachelor', for example, is *defined* as 'unmarried man'. But how do we find that 'bachelor' is defined as 'unmarried man'? Who defined it thus, and when?
(QUINE 1953 [89] p. 24)

Le problème soulevé ci-dessus vient de ce que la notion d'analyticité repose sur la notion de synonymie, et que cette dernière demeure non définie. Pour imposer l'absolue nécessité d'un jugement analytique, il faut considérer que la définition est substituable au concept, *salva veritate*. Autrement dit, la définition est censée être un exact synonyme du concept. Or, comment établir une telle synonymie ? Un test empirique, quelles que soient son étendue et sa précision, ne saurait faire l'affaire.

For most purposes extensional agreement is the nearest approximation to synonymy we need care about. But the fact remains that extensional agreement falls far short of cognitive synonymy of the type required for explaining analyticity [...]. The type of cognitive synonymy required there is such as to equate the synonymy of 'bachelor' and 'unmarried man' with the analyticity of [the statement 'All and only bachelors are unmarried men'], not merely with [its] truth.
(QUINE 1953 [89] p. 31)

Il ne s'agit pas d'établir l'exactitude d'un jugement considéré comme analytique, mais bien d'établir sa nécessité. Or, aucune vérification empirique ne peut produire une telle nécessité. Si l'analyticité des jugements épistémiques reste la seule source de leur nécessité, on se retrouve confronté à une pétition de principe. Nous avons besoin de l'analyticité pour établir la synonymie.

[...] to say that 'bachelor' and 'unmarried man' are cognitively synonymous is to say no more nor less than the statement: 'All and only bachelors are unmarried men' is analytic.
(QUINE 1953 [89] p. 29)

Comme la synonymie était censée fonder l'analyticité, il en résulte une forte suspicion quant à la validité théorique de ces notions. Les seuls énoncés analytiques qui subsistent sont les tautologies stériles du type un homme non marié est non marié. Tous les autres énoncés, si l'on suit cet argument, doivent être considérés comme synthétiques. Autrement dit, une connaissance factuelle est impliquée dans tout jugement qui n'est pas syntaxiquement tautologique. En conséquence, un concept donné ne peut être responsable, du fait de sa seule structure interne, des inférences que l'on peut produire à partir de lui.

La remise en cause de la distinction analytique - synthétique est parfaitement acceptable si l'on trouve le moyen de doter le système conceptuel d'une structure lui permettant de guider l'ensemble des inférences, quelle que soit leur nécessité. Une telle approche consiste à ramener dans le concept toute connaissance acquise à son propos. Ainsi, le rôle inférentiel des concepts devient complet, puisque leur structure suffit à expliquer toute inférence que l'on produit à partir d'eux. Par exemple, on stockera sous forme de liens entre concepts tout ce qui

était considéré comme des inférences nécessaires dans l'approche précédente. On introduira de même les relations contingentes sous forme de liens. Le système conceptuel, ainsi conçu, prend la forme d'un grand graphe censé résumer l'ensemble des connaissances de l'individu.⁶

Dans ce genre d'approche du système conceptuel, un concept se caractérise par l'ensemble de la théorie dans laquelle il intervient. Ce changement de perspective a d'importantes conséquences. Le rôle de l'expérience dans la formation des connaissances est reconnu. On s'attend toutefois à ce qu'il reste confiné à une part marginale, car l'importance des liens que les connaissances entretiennent entre elles devient prépondérante. Par exemple, bon nombre de connaissances, inscrites dans le réseau des concepts, seraient dictées par l'exigence de simplicité et de cohérence (QUINE 1953 [89]). À ce titre, les connaissances du sens commun fonctionnent comme celles de la science. Dans les deux cas, les individus postulent l'existence d'ensembles d'entités et de relations de manière à projeter une structure efficace sur le flux de l'expérience. L'édifice, comme c'est le cas des mathématiques, est sous-déterminé par l'expérience et a pour principale raison d'être la simplicité et la cohérence.

Cette vision du système conceptuel correspond à celle qui est adoptée par un courant important de la psychologie développementale. Pour expliquer les transitions dans les systèmes de croyance des enfants, il semble utile de présenter les connaissances des plus jeunes, non comme des versions partielles et dégradées de celles de leur aînés, mais comme des théories à part entière. Par exemple, les enfants de cinq ans ont une notion de vivant qui ne les induit pas à grouper les plantes et les animaux en une seule classe. La plupart refusent qu'une plante puisse être malade ou puisse mourir de faim (CAREY 1985 [8] p.168). Ce n'est qu'avec l'accès à une théorie élémentaire de la biologie, vers l'âge de neuf ans, que les enfants regroupent spontanément les animaux et les plantes dans la catégorie des êtres vivants (CAREY 1985 [8] p.182). Pourtant, les plus jeunes enfants observent de nombreuses similarités entre animaux et plantes, par exemple le fait que les deux grandissent. Ce qui retient les enfants de cinq ans de former un concept d'entité vivante incluant les plantes est qu'un tel concept ne jouerait pas de rôle dans leurs théories prébiologiques. De fait, leur concept VIVANT trouve sa place dans des théories enfantines où il s'oppose au concept CASSÉ, *id est* non fonctionnel, ou au concept IMAGINAIRE. Ce type d'observation conduit de nombreux auteurs à l'idée que les concepts n'existent que par leur rôle inférentiel.

Concepts must be identified by the roles they play in theories. (CAREY 1985 [8] p. 198)

All representations will be theory-laden. (GOPNIK & MELTZOFF 1997 [43] p. 44)

Selon cette position, la raison d'être des représentations conceptuelles que nous développons se situe au sein de théories élaborées pour rendre compte du monde qui nous entoure. Par exemple, l'enfant âgé de neuf mois semble disposer d'une théorie du genre "un objet réapparaît généralement au même endroit". Lorsqu'un objet disparaît sous un foulard, l'enfant va le chercher sous un autre foulard où il a précédemment eu l'occasion de trouver l'objet. Son aîné, âgé de dix-huit mois, est capable de visualiser des trajectoires qui se situent hors de son champ visuel. Cela lui permet de former une autre théorie, du type "l'objet se trouve sur l'une des trajectoires invisibles possibles". Grâce à cette théorie plus puissante, il peut former, de manière productive, des hypothèses raisonnables sur la localisation d'un objet qui a disparu (GOPNIK & MELTZOFF 1997 [43] p.101). Le développement ontogénétique du système conceptuel humain apparaît ainsi comme strictement analogue à l'avancement de la connaissance scientifique au cours de l'histoire. Certains changements conceptuels, comme celui qui a conduit Einstein à abandonner la notion de temps absolu, s'accompagnent de

⁶ Nous analyserons des exemples de formalismes de ce type plus tard (CF. CHAPITRE 7).

changements en profondeur et entraînent le remplacement de théories entières, par exemple celui de la théorie classique de Newton au profit de la théorie de la relativité. Il en irait de même du développement des individus (GOPNIK & MELTZOFF 1997 [43] p. 16).

L'idée selon laquelle le système conceptuel des individus fonctionnerait suivant les mêmes principes que la connaissance scientifique, accumulée au cours des siècles, ne va pas sans poser de problème. L'analogie nous laisse dépourvus lorsqu'il s'agit d'énoncer les principes d'évolution du système conceptuel sous l'effet de l'apprentissage. Les changements de paradigmes scientifiques ont été décrits comme dus à des causes exogènes, par exemple sociales. Ces mécanismes, dont la valeur prédictive est faible en ce qui concerne le contenu des théories scientifiques, ne peuvent pas être transposés dans le domaine cognitif. Nous reviendrons sur ce point (CF. CHAPITRE 7).

Le fait de décrire le système conceptuel comme une collection de théories crée également des difficultés lorsqu'il s'agit d'identifier les représentations qui sont évoquées par les mots du langage. Comment isoler ce qui se rapporte à un concept donné, si toute la connaissance se présente sous la forme d'un gigantesque graphe où tout se retrouve, directement ou indirectement, relié ? L'idée fondamentale est que, dans une théorie, les concepts n'existent que les uns par rapport aux autres.

No individual concept can be understood without some understanding of how it relates to other concepts.
(KEIL 1989 [60] p. 1)

Le risque théorique est de parvenir à un holisme conceptuel dans lequel la moindre inférence met en jeu potentiellement l'ensemble du réseau. Les théories dont dispose l'individu sont censées lui permettre de repérer des régularités et de produire des explications causales. Si ces théories sont toutes connectées entre elles, comment procéder, en pratique, pour prédire les inférences que l'individu produira en entendant un énoncé du langage ? Ces inférences peuvent en principe dépendre de l'état de n'importe quel endroit du réseau de connaissances, car tout point du réseau est relié de manière plus ou moins directe à la principale théorie mise en jeu dans l'énoncé. De ce fait, un même individu pourra produire des inférences radicalement différentes au cours du temps, même si, localement, sa théorie est restée inchangée. Comme l'effet cognitif d'un énoncé sur un individu est impossible à prédire, il en sera de même pour la communication. Une solution de ce problème de holisme peut consister à limiter l'interconnexion entre théories.

One of the dangers of proposing that concepts are intrinsically relational is the implication that they really are just the same as encyclopedic knowledge about the world. Everything is related to everything else in a vast network of roughly equal density. I will try to block this implication by arguing that beliefs tend to cluster in highly structured bundles with special properties that distinguish these clusters from the general interconnectedness of knowledge. Thus, lexical concepts are not viewed simply as the more or less arbitrary association of words to nodes in some vast network; instead, as a subset of all concepts, they tend to be associated with only particular configurations that are isolatable from other aspects of knowledge.
(KEIL 1989 [60] p. 4)

Circonscrire chacune des théories prêtées à l'individu est une nécessité dans ce cadre. Sans cela, on ne peut espérer prédire ni la production d'inférences, ni les changements de théorie au cours du développement, ni la possibilité même de communication puisque deux individus n'ont aucune chance de disposer du même réseau de théories. Le problème est que l'on ne dispose d'aucun critère pour circonscrire les théories. Une théorie sur les voitures doit inclure des connaissances sur leur fonction, donc sur la locomotion ; elle doit donc être connectée à une théorie sur la locomotion animale, à une théorie sur le mouvement, *et cætera*.

De même, elle doit être liée à une théorie chimique sur la combustion des hydrocarbures dont on récupère l'énergie dans le moteur. Elle doit être également liée à une théorie économique sur la rentabilité d'un véhicule. Qui peut dire que la théorie relative aux voitures s'arrête à tel point précis, qu'ensuite il s'agit d'une théorie chimique ou économique ? La théorie du cylindre et des bougies relève-t-elle de la voiture, de la chimie ou de la mécanique ? Celle des bougies et de l'allumeur ne serait-elle pas mieux placée au sein d'une théorie électrique ? En l'absence de critère de délimitation, ce type de question reste sans réponse, et la théorie des théories tombe dans un holisme sans pouvoir de prédiction.

Le problème qui vient d'être soulevé n'est pas seulement technique. Il ne s'agit pas de trouver un critère commode permettant de donner une délimitation raisonnable des théories au sein du graphe des connaissances conceptuelles. Il s'agit d'isoler, en suivant une règle générale, le sous-ensemble des connaissances relatives à un concept qui lui confère l'intégralité de son rôle inférentiel. Or, ce rôle inférentiel, que la théorie des théories prétend expliquer de manière exhaustive, s'exerce lorsque le concept intervient dans le sens d'un énoncé. Les connaissances jugées propres au concept doivent donc être délimitées de manière à déterminer la participation du concept dans l'élaboration du sens de l'énoncé. Un simple "rayon d'influence" dans le graphe des connaissances ne saurait répondre à cette exigence.

Dans l'approche qui consiste à conférer des structures définitionnelles aux concepts, ce problème ne se pose pas. Le principe de compositionnalité⁷ permet de caractériser entièrement la structure conceptuelle d'une expression linguistique complexe à partir de celle de ses composants, à l'aide d'un calcul approprié. Le principe de compositionnalité est facile à respecter avec un système de définitions, car il n'y a pas de différence de nature entre la structure conceptuelle des énoncés et celle des concepts eux-mêmes. On peut ainsi, aisément, déduire le comportement inférentiel du syntagme chat noir, si l'on suppose que sa représentation conceptuelle a une structure qui dérive de celle du concept CHAT qui dérive à son tour de celle du concept ANIMAL. En particulier, on peut reproduire le caractère nécessaire de l'inférence allant de la phrase il y a un chat noir sur le toit à la phrase il y a un animal noir sur le toit, en utilisant, par exemple, une simple déduction logique sur les définitions.

À partir du moment où l'on renonce aux définitions et que le concept CHAT est caractérisé par sa place dans un réseau de liens conceptuels, la caractérisation conceptuelle du syntagme chat noir devient, pour le moins, délicate. Si l'on tente de respecter le principe de compositionnalité en restant dans le cadre des théories, on risque de réintroduire le problème de l'analyticité que le système des relations devait permettre de contourner.

For, on the one hand, an inference is compositional iff its validity is determined by the *inferential roles* that its constituents contribute; and, on the other hand, an inference is analytic iff its validity is determined by the *semantic values* of its constituents; but, on the third hand (as it were), [inferential role semantics] says that the meaning of a constituent is the same thing as the inferential role it contributes to the expressions that contain it. So to repeat, the cost of representing an inference as compositional is that you then have to represent it as analytic. (FODOR & LEPORE 1992 [39] p. 181)

Si l'on suppose que les concepts sont censés, par leur seule structure, expliquer l'intégralité du rôle inférentiel d'un composé, on ne peut engendrer que des jugements analytiques. Or, nous avons commenté le caractère indéfendable d'une telle position.

Peut-on renoncer à la compositionnalité des théories ? Dans ce cas, comment définir la théorie qui caractérise un composé ? La théorie des chats noirs est-elle un sous-ensemble de la théorie des chats, ou un sous-ensemble de la théorie des couleurs ? À partir de la phrase il y a un chat noir sur le toit, la conclusion il y a un animal noir sur le toit a-t-elle le même statut qu'une

⁷ Nous reviendrons longuement sur ce sujet (CF. CHAPITRE 6).

conclusion contingente comme il y a un chat méchant sur le toit ? N'a-t-on pas besoin, pour réaliser cette deuxième inférence, de supposer l'existence d'une théorie spécifique aux chats noirs ? Ces difficultés, et d'autres que nous mentionnerons plus tard, conduisent à penser que le système des théories est incapable de rendre compte de la production des inférences adéquates à partir des compositions conceptuelles (CF. CHAPITRE 7).

La théorie des théories était motivée par le rejet des définitions et le souci d'effacer la différence entre jugement analytique et jugement synthétique. Pour ce faire, le système conceptuel était doté de toutes sortes de liens inter-conceptuels, nécessaires ou contingents, de manière à être en charge des jugements épistémiques. On peut imaginer, à l'inverse, de renoncer à l'idée que le système conceptuel possède un quelconque pouvoir inférentiel.

Une manière radicale de remettre en cause la distinction entre inférences analytiques et inférences synthétiques est de considérer que les inférences que l'on peut produire à partir d'un concept n'ont rien à voir avec la façon dont on le caractérise. Dans ce cas, la caractérisation des concepts, avec leurs propriétés intentionnelles et causales, se fait selon des considérations autres, par exemple le lien du système conceptuel avec le langage ou avec la perception.

[...] if Quine is right [...] then what you mean can't be reduced to what inferences you are prepared to accept. [...] Because, Quine argues, what inferences you are prepared to accept [...] depends not only on what you intend your words to mean, but also how you take the (non-linguistic) world. [...] Quine's rejection of analyticity, insofar as it's actually argued for (and insofar as we are proposing to concede it), is a rejection only of the possibility of an *epistemic* criterion for "true in virtue of meaning." In principle, at least, everything else [about semantics] remains wide open. (FODOR & LEPORE 1992 [39] p. 57)

Ainsi, il ne peut pas y avoir adéquation entre la signification d'un mot et les connaissances relatives au concept associé qui nous permettent de produire des inférences. Toutes les connaissances sont en partie *a posteriori*, elles ne sauraient résulter du seul fait de posséder le concept. En conséquence, il faut distinguer la signification d'un mot et son rôle épistémique, ce dernier dépendant de facteurs contingents.

Le constat selon lequel il n'existerait pas de jugement analytique rend illusoire tout espoir de rendre "locale" la confirmation des raisonnements (QUINE 1953 [89]). Une mise en adéquation de connaissances avec l'expérience peut impliquer n'importe quel élément du réseau de connaissances. Pour autant, on ne saurait automatiquement conclure de ce holisme de confirmation que la signification elle-même ne peut être locale (FODOR & LEPORE 1992 [39]). À partir du moment où un concept n'est pas assimilé à son pouvoir inférentiel, il est possible de refuser le holisme sémantique tout en acceptant le holisme de confirmation. Cela suppose d'abandonner la vision "pragmatiste" qui réduit un concept à ses effets sur le jugement épistémique. Dans une perspective cognitive, la signification est relative à la représentation conceptuelle et à son ancrage (CF. CHAPITRE 3). Elle peut être, en principe, divorcée du rôle inférentiel. À la limite, on peut considérer que rien, dans la représentation conceptuelle elle-même, ne contraint les inférences dans lesquelles elle est impliquée. Ces inférences seraient du ressort d'un mécanisme de raisonnement, alimenté par une base de connaissances. Ainsi, les concepts n'auraient aucun pouvoir inférentiel propre. Ce pouvoir leur serait conféré de l'extérieur. Le holisme, dans ce cas, serait celui de la base de connaissances, non celui des significations. Cette position sera discutée et évaluée plus tard (CF. CHAPITRE 8).

Le passage aux théories était une tentative pour remédier à l'incapacité des définitions à reproduire la nécessité des inférences auxquelles un concept peut conduire. Il existe cependant un moyen d'augmenter les structures conceptuelles pour expliquer les inférences, même si

l'on observe que la plupart de ces inférences n'ont aucun caractère nécessaire. En particulier, comme nous allons le voir, il est possible d'expliquer les inférences présentant un aspect régulier ou exemplaire à l'aide de structures appropriées.

4.3. Concepts et typicité

L'une des caractéristiques manifestes de nombreuses inférences que nous faisons spontanément est leur aspect incertain. Lorsque nous supposons qu'un objet, dont nous savons qu'il s'agit d'un oiseau, vole, notre inférence est souvent valide. Ce jugement est cependant incertain, car nous nous exposons à l'erreur. Lorsque nous décidons qu'une boîte de peinture doit être classée parmi les jouets, nous tolérons encore une incertitude, liée cette fois au flou du jugement. L'ambition de l'approche que nous allons présenter dans cette section est de doter les concepts d'une structure permettant la production de ces jugements incertains.

Une façon de comprendre cette incertitude liée aux inférences consiste à la faire dériver des propriétés de la capacité humaine de catégoriser les objets perçus.

To have a concept of X is to know something about the properties of entities that belong to the class of X, and such properties can be used to *categorize* novel objects. Conversely, if you know nothing about a novel object but are told it is an instance of X, you can *infer* that the object has all or many of X's properties; that is, you can run the "categorization device in reverse."
(SMITH & MEDIN 1981 [96] p. 8)

L'opération de catégorisation est, dans certaines approches, à la base de la plupart des jugements inférentiels. Or, la manière dont les sujets catégorisent est incorrectement prédite par les modèles qui dotent les concepts de propriétés nécessaires. De nombreuses expériences ont montré que l'appartenance d'une série d'objets à une catégorie donnée ne montrait pas, le plus souvent, de transition discontinue, contrairement à ce que l'on attendrait si la catégorie était définie par des traits binaires (ROSCH 1975 [92]). Les tableaux suivants indiquent les moyennes des jugements d'appartenance, entre 1 et 7, donnés par un ensemble de 209 sujets (ROSCH 1975 [92]).

<i>Furniture</i>		<i>Vehicle</i>	
Chair	1.04	Automobile	1.02
Table	1.10	Truck	1.17
Dresser	1.37	Jeep	1.35
Desk	1.54	□an	1.95
Bed	1.58	Train	2.15
Divan	1.70	Bicycle	2.51
Buffet	2.89	Yacht	3.76
Telephone	6.68	Elevator	5.90

Pour expliquer ce phénomène, il est naturel de penser que le jugement des sujets se fonde, non sur une série de tests binaires, mais sur une mesure de similarité. La représentation d'un concept est censée permettre cette mesure de similarité.

In this case, cognitive representations of categories clearly contained more of the information needed to respond to category members which had been rated good examples of the category [...] than to respond to category members which had been rated bad examples. In other words, cognitive representations of categories appeared to be more similar to the good examples than the poor examples. (ROSCHE 1975 [92] p. 225)

Un concept pourrait être identifié à une moyenne des propriétés des instances de la catégorie correspondante, un prototype.

We can only come up with the following two assumptions [of the probabilistic view]: (1) the representation of a concept is a summary description of an entire class; and (2) the representation of a concept can *not* be restricted to a set of necessary and sufficient conditions; rather, it is some sort of measure of central tendency of the instances' properties or patterns. (SMITH & MEDIN 1981 [96] p. 61)

Les concepts ressemblent ainsi à des représentations moyennes. Certaines inférences concernant les objets perçus peuvent être prédites par la distance entre l'objet perçu et cette représentation moyenne. Cette manière de décrire les concepts résout certaines difficultés du jugement épistémique. L'inférence selon laquelle un oiseau vole n'est pas nécessaire, pas plus que celle qui conduit à penser qu'il chante. Ces deux informations seraient exclues du concept OISEAU dans un système à base de définitions nécessaires. Or, le caractère contingent des deux inférences est relatif : la première est généralement considérée comme une caractéristique presque universelle des oiseaux, la seconde comme une propriété fortuite de certains d'entre eux. Le système des prototypes permet de reproduire cette gradation dans la sûreté du jugement.

Il existe une variété de moyen de doter les concepts d'une structure de manière à reproduire les jugements incertains. Le plus naturel est sans doute d'adapter les définitions à base de traits en attachant une probabilité à chacun de ces traits. On ne retient dans la structure des concepts que les traits qui possèdent une descriptivité (*cue validity*) suffisante⁸.

The second assumption of the featural approach is the critical one: the features that represent a concept are salient ones that have a substantial probability of occurring in instances of the concept. More precisely, if F_i is a feature and X_j a concept, F_i will be a feature of X_j to the extent that (1) F_i is salient (either perceptually or conceptually), and (2) the probability of F_i given X_j , $P(F_i | X_j)$, is high (for example, F_i tends to be true of an instance labeled as an X_j). (SMITH & MEDIN 1981 [96] p. 62)

La descriptivité d'une catégorie se déduit de celle des traits qui la composent. Elle mesure la capacité de ces traits à prédire la catégorie, et par là-même la manière dont la catégorie se différencie des autres catégories.

Il est également possible d'incorporer la gradualité dans les traits eux-mêmes, et non seulement dans leur probabilité d'occurrence. Cette deuxième variante caractérise un concept comme par une structure de traits dotés d'une plage de valeurs typique.

⁸ Cette notion correspond à la probabilité conditionnelle $P(F|C)$ d'observer le trait F dans la catégorie C .

The second assumption [of the dimensional approach] has two parts: (1) any dimension used to represent a concept must be a salient one, some of whose values have a substantial probability of occurring in instances of the concept; and (2) the value of a dimension represented in a concept is the (subjective) average of the values of the concept's subsets or instances on this dimension. [...] Part 1 is the dimensional equivalent of the featural assumption [...]. Knowing only that a dimension is represented in a concept does not tell us what particular value of that dimension is represented, so part 2 is needed to specify the unknown value. (SMITH & MEDIN 1981 [96] p. 102)

La validité des inférences que l'on produit à propos d'instances de tels concepts va dépendre de la typicité des traits présents dans la structure du concept, affectée de la probabilité du trait. Cette grandeur peut être comparée à des seuils, ce qui permet de considérer l'inférence tantôt comme valide, tantôt comme invalide, tantôt comme douteuse. Un tel système permet donc d'expliquer une réponse du type "je ne sais pas" lorsque l'on demande si un tapis est une pièce d'ameublement.

Certaines variantes de l'approche fondée sur la typicité confèrent aux représentations conceptuelles une structure plus concrète que celle d'assemblages de traits moyens et fréquents. Ainsi, un concept serait représenté par un ensemble réduit d'exemplaires qui peuvent être des instances ou des sous-catégories (SMITH & MEDIN 1981 [96] p. 145). Ceci permet d'expliquer que certains jugements impliquant des concepts soient sensibles au contexte. Ainsi, les jugements concernant un oiseau domestique et ceux qui portent sur un oiseau rapace seraient produits par des exemplaires différents du concept OISEAU, exemplaires qui font partie intégrante de cette représentation.

Le caractère concret de la représentation conceptuelle peut rejoindre la vision empiriste selon laquelle les concepts seraient représentés par des percepts plus ou moins généralisés. Les concepts pourraient correspondre à des patrons (*templates*) inanalysables en termes de traits, et le jugement de catégorisation se réduirait à une mesure globale de la ressemblance entre l'objet perçu et le patron (SMITH & MEDIN 1981 [96] p. 131). Cette ressemblance serait suffisante pour conférer un rôle inférentiel aux concepts : nous déduirions qu'un pigeon est un oiseau parce que le patron PIGEON peut être superposé, de manière globale, au patron OISEAU.

Cette contiguïté possible de l'approche fondée sur la typicité et d'une théorie empiriste de l'ancrage (CF. CHAPITRE 3) permet de voir la première comme un prolongement de la seconde. De même qu'un concept empiriste est défini comme une moyenne statistique de paramètres perceptuels, un concept, dans la théorie des prototypes, comporte des traits conceptuels dont les valeurs et la fréquence sont le fruit d'une moyennage. Ce rapprochement permet d'expliquer d'une part l'ancrage des concepts, et d'autre part leur capacité à engendrer des inférences incertaines, par le même mécanisme de représentation moyenne.

La nature statistique des catégories typiques ouvre la voie à des mécanismes d'apprentissage particulièrement simples. Les systèmes de regroupement (*clustering*) comme il en existe en statistique (par exemple l'algorithme des nuées dynamiques) ou en intelligence artificielle (par exemple les techniques SBL) permettent de produire des catégories telles que la distance moyenne entre objets soit minimale au sein des catégories et maximale entre elles (CF. ANNEXE).

La théorie des prototypes présente l'intérêt indéniable de rendre compte de toute la gamme des inférences, en incluant les jugements incertains. Toutefois, plusieurs critiques lui ont été adressées. L'une d'elle concerne le lien qui est postulé entre la ressemblance d'un objet perçu avec un prototype et l'appartenance de cet objet perçu à la catégorie correspondante. Par exemple, l'appartenance à la catégorie "nombre pair" n'est pas graduelle, et pourtant certains nombres pairs comme deux ou quatre apparaissent comme plus typiques que les autres (ARMSTRONG & GLEITMAN & GLEITMAN 1983 [1]). Le problème ne se pose

pas seulement pour les concepts scientifiques. Chaque individu a un jugement personnel sur les traits typiques du concept GRAND-MÈRE, qui peut être lié à l'image de sa propre grand-mère. Pourtant, ce même individu statuera sur le fait qu'une personne est grand-mère non d'après son apparence, mais d'après le constat qu'elle a des descendants de deuxième degré. Ce divorce entre prototypicité et jugement d'appartenance rompt le lien théorique qui fondait le caractère graduel de ce jugement d'appartenance. Du même coup, on perd le fondement théorique de bon nombre d'inférences que l'on est censé produire à partir de la possession des concepts. Si les catégories ne peuvent plus être définies par une relation de ressemblance avec un élément moyen ou avec un nombre restreint d'exemplaires, il faut imaginer deux gradients plus ou moins indépendants, un gradient de typicité et un gradient d'appartenance (KAMP & PARTEE 1995 [55]). Cette complication affaiblit cependant la théorie qui doit expliquer la genèse de ces deux gradients.

Une autre limitation de la théorie des prototypes est liée à son caractère statistique. La théorie prévoit que les capacités inférentielles des individus sont sensibles à leur expérience. La première fois qu'un individu observe un oiseau dépourvu de plumes, il est surpris. Si de tels oiseaux deviennent fréquents dans son environnement, il finira par s'habituer et ne sera plus en situation d'étonnement. La théorie standard des prototypes ne permet pas de prédire un quelconque étonnement. Certes, l'utilisation de seuils permettrait à un système à base de catégories typiques de détecter une exception : un objet perçu qui serait atypique par l'un de ses traits seulement. Un tel objet pourrait ainsi être classé comme atypique au sein de sa catégorie. En cas de répétition d'expériences similaires, soit la catégorie se modifie, soit il se forme une nouvelle catégorie. Dans les deux cas, la théorie prévoit correctement la fin de l'étonnement avec l'habitude. Ce que ce genre de théorie ne prévoit pas, cependant, c'est que l'individu peut se trouver en situation d'étonnement permanent. Par exemple, si notre sujet observe des oiseaux dépourvus de bec, il en inférera que ces animaux ne peuvent pas s'alimenter. Cet étonnement épistémique n'est pas sensible à la répétition du stimulus : quel que soit le nombre de ces oiseaux mutants qu'il lui sera donné d'observer, sa déduction restera valide et son étonnement également. Contrairement à l'exemple des plumes, la modification de la catégorie ou la formation d'une nouvelle catégorie ne changent rien. L'étonnement épistémique ne peut cesser que par une explication, par exemple que ces oiseaux se nourrissent par une opération de succion. L'histoire des sciences est féconde en exemples d'étonnements durables. Avant que le mécanisme d'écholocalisation soit avancé au milieu du siècle dernier, la propriété des chauves-souris de se mouvoir dans l'obscurité totale restait la source d'un étonnement permanent. La théorie des prototypes est incapable de reproduire le type de jugement conduisant à l'étonnement épistémique. Dans un système cognitif où tous les jugements peuvent être faits et défaits par le fait de l'habitude, il ne peut y avoir de contradictions logiques. L'inférence x est un oiseau \rightarrow x possède un bec ne peut, dans la théorie des prototypes, avoir d'autre source que les régularités de l'environnement. Il n'y a pas de place pour les explications causales et pour le trouble épistémique causé par l'observation d'une mise en défaut de la causalité. En ce sens, il est permis d'affirmer que la théorie des prototypes, telle qu'elle est décrite en moyenne, ne suffit pas à rendre compte des capacités inférentielles humaines. Nous reviendrons sur ce point lorsqu'il s'agira de proposer des mécanismes capables d'expliquer les jugements incertains tout en ménageant la possibilité d'un jugement épistémique (CF. CHAPITRE 9).

Une troisième difficulté que l'on rencontre en cherchant à appliquer la théorie des prototypes concerne les composés linguistiques. La prédiction des capacités inférentielles d'un composé linguistique comme pomme rayée ne pose aucun problème dans un système de définitions. Pour obtenir la définition du composé, il suffit d'utiliser les opérations habituelles de la logique classique. Ainsi, le concept associé au syntagme pomme rayée pourra être défini comme la conjonction logique des définitions des concepts POMME et de RAYÉ. Si l'on

abandonne les définitions binaires en faveur de coefficients numériques d'appartenance ou de typicalité, alors il faut donner un nouveau sens à la conjonction qui apparaît dans la définition d'une pomme rayée. Ceci peut être fait dans le cadre de la logique floue, en représentant le jugement de typicalité par une fonction attribuant une valeur numérique entre 0 et 1 aux différents objets. Par exemple, une chaise sera une pièce d'ameublement à 99%, une lampe à 70% et un téléphone à 40%. Dans un tel cadre, on peut dire que si un objet est pomme à 70% et rayé à 60%, alors il est pomme rayée au plus à 60%, selon la règle du minimum⁹. Malheureusement, il est très difficile d'accorder ce type de calcul avec les résultats suggérés par l'intuition. En particulier, pour un objet qui est pomme à 50%, le coefficient alloué à la conjonction "pomme et non pomme" sera 0.5, alors que l'on attend 0 en raison de la contradiction¹⁰ (OSHERSON & SMITH 1981 [79]). De plus, ces calculs rendent impossible qu'un objet soit plus typique pour la propriété combinée que pour la propriété de base (OSHERSON & SMITH 1981 [79]). Or, un perroquet est plus typique en tant qu'oiseau domestique qu'en tant qu'oiseau. Différents moyens ont été proposés pour remédier à ces difficultés. Par exemple, la théorie des super-valuations consiste à conserver des valeurs binaires lorsque l'appartenance est hors de doute, et à ne pas allouer de valeur d'appartenance lorsque ce n'est pas possible. Ainsi, pour un objet J qui serait classé comme pomme à 50%, l'expression $\text{pomme}(J)$ ne reçoit pas de valeur de vérité. En revanche, la conjonction $\text{pomme}(J)$ et $\text{non pomme}(J)$ peut recevoir la valeur 0, car toutes les manières d'instancier la super-valuation fournissent ce résultat (KAMP & PARTEE 1995 [55]). Pour assigner une valeur à un composé comme pomme rayée, il est nécessaire de tenir compte de la dissymétrie entre le nom et l'adjectif. Le coefficient associé à l'expression $\text{rayé}(J)$ doit être "re-calibré" dans le contexte fourni par l'expression $\text{pomme}(J)$ avant d'entrer dans le calcul de la combinaison (KAMP & PARTEE 1995 [55]).

Noter que ce genre de mécanisme échoue lorsqu'il s'agit d'expliquer des combinaisons de type lion en pierre. La représentation moyenne que l'on peut former à propos d'une expression comme lion en pierre ne semble pas déductible de la représentation moyenne de lion et de la représentation moyenne de pierre. Les inférences que l'expression lion en pierre peut occasionner ne sont donc pas déductibles, au sein de la théorie des prototypes, des inférences que les composants lion et pierre peuvent susciter. De plus, la représentation moyenne ne semble même pas nécessaire pour la formation du jugement d'appartenance des concepts composés.

Thus the extension of *male nurse* is determined in the simplest way possible, namely as the standard logical conjunction of the constituent concepts *male* and *nurse*. [...] This establishes that no prototype is needed to fix its extension. But it is not just that no prototype is required for this; it is hard to see how the prototype for *male nurse* could play any significant role in determining membership. (KAMP & PARTEE 1995 [55] p. 168)

La théorie des super-valuations ne suffit pas non plus à expliquer de tels composés. Aucune re-calibration du prédicat pierre ne donne une intersection non vide avec le prédicat lion. Cependant, on peut supposer que c'est précisément cet échec qui provoque l'activation

⁹ Plusieurs règles ont été proposées pour remplacer les opérations binaires de conjonction et de disjonction. Par exemple, le coefficient attaché à la conjonction peut être représenté par le produit des coefficients associés aux opérandes. Le choix le plus courant consiste à prendre le minimum pour la conjonction et le maximum pour la disjonction. Ces définitions redonnent les résultats de la logique classique lorsque les coefficients sont des valeurs binaires 0 et 1.

¹⁰ De même, la disjonction "pomme ou non pomme" se verra accorder un coefficient de 0.5 au lieu de 1. Une version multiplicative de la conjonction donnera 0.25 à celle de "pomme et non pomme", ainsi qu'à celle de "pomme et pomme" !

d'un sens dérivé pour le nom modifié. Le mécanisme de modification peut s'appliquer au sens dérivé, par exemple la forme du lion dans notre exemple. Le choix de ce sens dérivé se révèle cependant difficile à modéliser dans le cadre des théories actuelles (KAMP & PARTEE 1995 [55]).

Ces difficultés concernant la déduction des propriétés des concepts associés aux expressions composées ne concernent pas uniquement la théorie des prototypes. Dans le chapitre suivant, nous allons étudier ce problème en détail en abordant le caractère compositionnel des concepts.

Conclusion

Ce chapitre nous a permis d'étudier l'interface entre le système conceptuel et le raisonnement. La question première, lorsqu'on s'intéresse à cette interface, porte sur le partage, entre le système conceptuel et le système de raisonnement, des connaissances qu'un mot de la langue peut convoquer. Une manière d'effectuer le partage consiste à ne retenir dans le système conceptuel que les connaissances nécessaires, en les incluant dans des définitions. Une autre acception du système conceptuel y inclut l'ensemble des connaissances typiques relatives aux concepts. Les concepts sont caractérisés, dans cette approche, par des prototypes ou des exemplaires. Le système conceptuel héberge ainsi non seulement les connaissances nécessaires, mais également les régularités et les connaissances incertaines associées aux concepts. Une version plus radicale, où le système conceptuel héberge tout ce que l'on peut savoir sur les concepts, conduit à des modèles comme la théorie des théories dans laquelle un concept est caractérisé par l'ensemble des connaissances dans lesquelles il est susceptible d'intervenir. La version opposée est une vision du système conceptuel dans laquelle les connaissances ne jouent aucun rôle dans la caractérisation des concepts.