

Partie 2 :
**Questions concernant la nature
des concepts**

Introduction

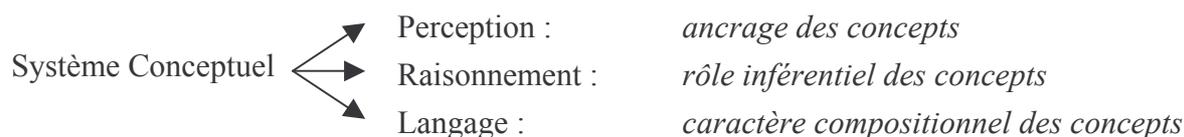
Dans la littérature des sciences cognitives, la notion de “concept” renvoie, de manière non exclusive, à l’une ou l’autre des trois acceptions suivantes.

Représentation : Il s’agit d’une entité mentale qui réfère à, ou désigne, une réalité dans le monde ou une donnée de la perception. Cette acception de la notion de concept, suggérée aussi par la notion d’idée, est sous-jacente aux théories classiques de l’entendement, de la connaissance, et de la signification. Dans ce sens, les concepts garantissent l’interaction du fonctionnement cognitif avec l’environnement, et ils se doivent disposer d’un ancrage dans cette interaction.

Prédicat : Une capacité cognitive largement étudiée par la psychologie est la faculté de raisonnement. Face à un objet perçu particulier, l’être humain ne retient que certaines de ses caractéristiques pour le considérer comme étant un élément d’une classe d’objets. Il peut alors utiliser le prédicat attribué à cette classe pour intégrer sa perception dans un processus de raisonnement. Dans ce cadre, la possession d’un concept garantit la convocation des connaissances associées et permet d’en tirer certaines inférences. Les concepts se voient donc dotés d’un rôle inférentiel.

Mot : La construction du sens exprimé par les langues humaines est un problème central de toute théorie linguistique. Le rôle que jouent les mots dans le processus se révèle incontournable. Dans une analyse cognitive des mécanismes sémantiques, les concepts constituent la projection mentale des mots. Leur manipulation dans des processus de composition serait dictée par la grammaire de la langue concernée. Les concepts acquièrent ainsi un caractère compositionnel.

Ces différentes acceptions sont rarement distinguées dans les textes et sont souvent suggérées simultanément lors de l’emploi du mot “concept”. La confusion qui en résulte parfois peut être évitée si l’on se place dans l’hypothèse d’un système conceptuel autonome. Les trois acceptions ci-dessus correspondent à trois facettes des entités internes au système conceptuel.



Sachant que le système conceptuel possède trois interfaces, avec la perception, avec le raisonnement et avec le langage, il n’est pas étonnant que les concepts soient dotés d’un ancrage, d’un rôle inférentiel et d’un caractère compositionnel. Étant donné que les concepts ne sont pas directement observables, les différents auteurs ont utilisé l’une ou l’autre de ces trois interfaces pour tenter de caractériser les concepts. Nous organiserons donc notre revue sélective des textes selon ces trois interfaces. Nous ne chercherons pas à être exhaustive. Nous retiendrons les théories des concepts susceptibles d’étayer la réflexion que nous mènerons dans les parties suivantes sur le problème de la construction du sens.

Chapitre 3 :
Ancrage des concepts

Introduction

Les concepts ne sont pas des entités suspendues dans un éther cognitif. En tant que supports mentaux de la signification, ils constituent un intermédiaire entre un mot et l'objet ou la situation que ce mot permet de désigner dans le monde perçu. De ce fait, les concepts doivent être ancrés dans nos expériences. En d'autres termes, le système conceptuel, qui a en charge la construction du sens et héberge les significations, doit posséder une interface avec la perception pour que les significations puissent porter sur les données de nos expériences.

3.1. Les données du problème

À partir du moment où les concepts sont considérés comme des représentations mentales, le problème de leur ancrage se pose. Or, la manière dont cet ancrage peut être réalisé fait l'objet de débats opposant des positions bien distinctes. Ces débats touchent à des questions fondamentales comme la nature du contenu et de la forme des représentations conceptuelles, et sont indissociables du type de modélisation auquel on souscrit. Dans ce chapitre, notre objectif n'est pas d'argumenter et de prendre parti, mais plutôt de présenter certaines positions présentes dans la littérature concernant cette question centrale de l'ancrage. Nous ne chercherons pas à être exhaustive, ne retenant ici que les théories qui interfèrent avec notre problématique. Nous commençons, dans cette section, par clarifier certains points de vocabulaire.

Représentations, modélisation et ancrage

La notion de représentation est avant tout un outil théorique pour l'explication du fonctionnement cognitif. La variété des comportements humains ne saurait être expliquée par une simple correspondance avec des configurations de stimuli. La majorité des modèles de la cognition humaine utilisent la notion de représentation pour médiatiser le lien complexe entre l'expérience du sujet et le comportement observable qu'il produit. Toutefois, ces modèles se divisent en deux groupes, selon le type de calcul qu'ils postulent, si bien que la notion de représentation est utilisée dans deux sens radicalement différents.

Standard explanations of how systems come to exhibit sophisticated cognitive performances advert to internal representations. Computationalists take representations to be static configurations of symbol tokens. Dynamicists conceive representations very differently. They find their representations among the kind of entities that figure in [dynamic system theory] [...] Unlike digital computers, dynamical systems are not inherently representational. A small but influential contingent of dynamicists have found the notion of representation to be dispensable or even a hindrance for their particular purposes. (□AN GELDER 1998 [105] p. 622)

Un modèle computationnel de la cognition repose sur l'application séquentielle de règles explicites à des représentations symboliques. Les représentations jouent le rôle de *tokens*, de simples "jetons" manipulés par le système d'après leur forme. Les règles, dans la mesure où elles sont explicites, sont elles-mêmes constituées de *tokens*. Ces *tokens* sont supposés avoir une existence physique identifiable dans le système (NEWELL 1980 [78]). Cette nécessité de considérer l'implantation matérielle des *tokens* mise à part, un modèle

symbolique constitue une description qui se situe à un niveau relativement abstrait, celui des manipulations symboliques formelles telles que celles qu'effectue une machine de Turing. Cependant, du fait de l'existence matérielle des symboles manipulés, un tel modèle est bien davantage qu'un simple outil théorique permettant de prédire l'évolution du système. Il existe un isomorphisme supposé entre les opérations décrites dans le modèle et les mécanismes physiques qui se déroulent dans le système. Dans un tel schéma, la notion de représentation est donc incontournable. Non seulement les *tokens* constituent les ingrédients fondamentaux du modèle, mais le modèle prédit même la possibilité de caractériser ces *tokens* de manière indépendante, par exemple par une technique de neurophysiologie.

Cette manière de concevoir le traitement cognitif introduit automatiquement une distinction de type syntaxe *versus* sémantique. La manipulation des représentations est une mécanique formelle, une syntaxe, qui dépend exclusivement de la structure de ces représentations. Pour que cette mécanique ne tourne pas à vide, il faut bien que ces représentations soient ancrées, soit dans les mécanismes perceptifs, soit dans les mécanismes effecteurs. Elles possèdent, de ce fait, une sémantique : elles sont interprétables dans le domaine des perceptions ou dans celui des actions.

We need the syntactic or the symbolic level because we must preserve certain interpretations over mental operations [...] This we can do only if we have a semantic function whose definition has access to the generative structure of the symbolic expressions. [...] To count as a computation rather than simply any functionally described physical system, [a machine] must contain symbols that are interpreted. [...] This quality of symbols and of computational states, whereby they can consistently be given a semantic interpretation, is not the only thing that makes useful computation possible; but it is one of the most important characteristics shared by computation and cognition.

(PYLYSHYN 1984 [88] p. 62)

Les propriétés sémantiques doivent être compatibles avec la syntaxe : à partir d'un premier ensemble de représentations directement ancrées, la mécanique syntaxique engendre de nouvelles représentations qui peuvent être, à leur tour, sémantiquement interprétées. Cette contrainte impose que les représentations intermédiaires, celles qui interviennent dans la chaîne de traitement symbolique, soient également interprétables, même si elles ne sont pas directement associées, par le biais de l'expérience, à la perception ou aux actions¹ (HARNAD 1990 [46]). Comme elles possèdent une structure formelle interne, elles peuvent hériter leur ancrage de leurs constituants et du calcul qui les a produites.

What is the representation of a zebra? [In our example] it is just the symbol string "horse & stripes". But because "horse" and "stripes" are grounded in their respective iconic and categorical representations, "zebra" inherits the grounding, through its grounded *symbolic* representation. In principle, someone who had never seen a zebra (but had seen and learned to identify horses and stripes) could identify a zebra on first acquaintance armed with this symbolic representation alone (plus the nonsymbolic representations of horses and stripes that ground it). Once one has the grounded set of elementary symbols provided by a taxonomy of names (and the iconic and categorical representations that give content to the names and allow them to pick out the objects they identify) the rest of the symbol strings of a natural language can be generated by symbol composition alone, and they will all inherit the intrinsic grounded of the elementary set.

(HARNAD 1990 [46] p. 343)

¹ Cela n'empêche pas l'introduction de tokens non interprétables en tant que tels. Par exemple, dans un système logique classique, les symboles de prédicats ou les constantes reçoivent une interprétation directe, au contraire des symboles logiques dont la fonction se limite à diriger l'interprétation des expressions complexes.

La situation est tout autre si l'on adopte une théorie dynamique de la cognition. Les théories dynamiques considèrent le modèle formel comme un outil de prédiction qui simule le système physique, sans postuler l'existence d'un isomorphisme. Ainsi, les équations de Kepler ne sont pas censées posséder la moindre contrepartie matérielle dans le système des planètes. La description formelle à l'aide de symboles désignant la position d'une planète, sa vitesse ou l'aire du rayon vecteur est propre au modèle, et ne s'applique que métaphoriquement au système matériel. La notion de représentation, contrairement aux modèles symboliques, est ici facultative. Une partie des défenseurs des systèmes dynamiques sont même anti-représentationnalistes, considérant que rien, dans le système, n'est interprétable dans les termes d'un domaine extérieur au système.

L'idée fondamentale est donc que les facultés cognitives sont inextricablement liées à l'historique de ce qui est vécu, de la même manière qu'un sentier au préalable inexistant apparaît en marchant. L'image de la cognition qui s'ensuit n'est pas la résolution de problèmes au moyen de représentations, mais plutôt le faire-émerger créateur d'un monde, avec la seule condition d'être opérationnel : elle doit assurer la pérennité du système en jeu.

(□ARELA 1988 [106] p. 111)

Cependant, une telle position ne représente pas un consensus parmi les défenseurs de l'approche dynamique. Notamment, certains modèles connexionnistes s'en éloignent en introduisant la notion de représentation distribuée.

While [behaviorist accounts of behavior] do involve simple mechanisms of learning, there is a crucial difference between our models and the radical behaviorism [...]. In our models, we are explicitly concerned with the problem of internal representation and mental processing, whereas the radical behaviorist explicitly denies the scientific utility and even the validity of the consideration of these constructs. [Our connectionist models] all concern internal mechanisms for activating and acquiring the ability to activate appropriate internal representations.

(RUMELHART & MCCLELLAND 1986 [93] p. 121)

Certaines caractéristiques des modèles dynamiques méritent l'étiquette de représentation. On peut appeler représentation l'état global du système, celui qui résume à lui seul tout son passé. Ainsi, dans un système de planètes considéré du point de vue de la mécanique newtonienne, la position du centre de gravité des planètes et leur vitesse suffit à résumer l'état du système et à permettre la prédiction de ses états futurs. Dans un modèle de réseau de neurones, il en est de même de l'ensemble de valeurs constitué par les poids synaptiques et les activités des neurones. Il y a dans ce cas une seule représentation globale, dont les sous-ensembles peuvent être également considérés comme des représentations. Ainsi, les données relatives à une planète constituent la représentation associée à cette planète. De même, on peut appeler représentation l'état d'activité d'un sous-ensemble de neurones.

Should we identify one's conceptual framework with the configuration of synaptic *weights* in one's brain? Or with the *partitions* they effect across the activation vector space of the assembled neurons to which they connect? Or perhaps with the overall input-output *function* that the network comes to instantiate? The weights uniquely dictate both the partitions and the function, but despite the functional primacy of the weights, there are good reasons for identifying the partitions, and the function they serve, as reflecting most directly the antecedent notion of a "conceptual framework". [...] At least for now, therefore, let us adopt the partitions and the functions they serve as the closest available neural analogue of what the philosophical tradition conceives as our "conceptual framework".

(CHURCHLAND 1989 [19] p. 232)

Dans un réseau de neurones sujet à des phénomènes de synchronisation, le choix du support matériel des représentations laisse moins de doute. Tout ensemble synchrone de neurones peut être considéré, de manière fort naturelle comme une représentation, et ceci d'autant plus facilement que la synchronisation est parfois invoquée comme corrélat de la conscience (CRICK & KOCH 1990 [23]).

Les systèmes dynamiques comme les réseaux de neurones présentent la propriété de posséder des attracteurs. Ainsi, un réseau de neurones placé dans certaines conditions initiales, va évoluer spontanément vers un état d'équilibre. Dans la mesure où ils sont façonnables par apprentissage, ces attracteurs sont interprétables par un observateur : ils sont ancrés dans la configuration d'entrée du système qui a permis leur stabilisation par apprentissage (HOPFIELD 1982 [48]). Il est donc intéressant de considérer que les attracteurs d'un tel système dynamique constituent autant de représentations.

Ce type d'ancrage, propre aux systèmes dynamiques, a cependant ses limites. C'est ainsi que pour certains auteurs, la sémantique des attracteurs a peu de choses à voir avec les entrées du système. Le rôle de ces entrées est ramené à celui de simples perturbations. Le système possède un certain nombre d'attracteurs, mais ces attracteurs ne dépendent que de manière fortuite de l'histoire des interactions du système avec l'extérieur. D'où l'idée selon laquelle le système crée des significations pour lui-même (□ARELA 1988 [106]). Certains attracteurs peuvent sembler posséder une signification pour la raison qu'ils sont corrélés à des configurations de stimuli, mais ce genre de signification n'existerait que dans l'œil du modélisateur (SKARDA & FREEMAN 1987 [95]). Selon cette manière de concevoir les systèmes cognitifs, il ne saurait donc y avoir d'ancrage proprement dit.

Lorsque l'on s'intéresse au langage, il est difficile de souscrire d'emblée à la position qui vient d'être invoquée. Il semble que les mots perçus dans le message de l'interlocuteur aient un effet moins aléatoire que celui de simples perturbations. Puisque l'interaction langagière est possible, il est permis de supposer qu'elle ne résulte pas de la confrontation entre deux systèmes "autistes" ne connaissant que leurs propres significations internes. Nous considérons donc, dans ce document, que l'hypothèse représentationnaliste mérite d'être explorée, que ce soit sous sa forme computationnelle ou dans sa réalisation dans les systèmes dynamiques. Nous nous intéresserons donc à des systèmes dont certains états peuvent être qualifiés de représentationnels. Nous nous demanderons comment ces états peuvent entretenir un rapport d'ancrage avec d'autres états du système ou avec certaines configurations des entrées et des sorties. Nous constaterons, dans ce chapitre, que cette question de l'ancrage est particulièrement problématique.

Représentations : le contenu

La notion de représentation constitue un outil commode, peut-être incontournable, pour élaborer et exprimer un modèle de la sémantique du langage. Cependant, l'enjeu dépasse la question de la modélisation du langage. La question de la nature des représentations est au centre d'un débat philosophique concernant leur propriété intentionnelle, propriété attribuée à certains phénomènes cognitifs. Certains états mentaux semblent posséder la propriété remarquable d'être obligatoirement à propos de quelque chose. Ainsi, une grande partie des états conscients, par exemple l'état correspondant au fait de voir orange ou d'avoir mal, sont à propos de quelque chose : ils ont un contenu. Ainsi, l'état conscient associé à la couleur orange sera lié à la vision d'un livre ou de la surface d'un mur ; une douleur donnée sera ressentie comme si elle se trouvait dans le pied ou dans une dent. D'autres types d'états mentaux possèdent manifestement cette propriété d'avoir un contenu. Par exemple, une croyance ou un désir n'existe cognitivement, la plupart du temps, que si elles sont à propos d'un état de choses, par exemple le fait de réussir un examen. La question de l'intentionnalité

se pose donc pour plusieurs types d'états mentaux, parmi lesquels les représentations mentales associées aux expressions langagières occupent une place de premier plan. Ainsi, la représentation mentale associée au mot *arbre* semble être à propos de quelque chose : l'image d'un arbre qui se trouve dans la cour ou cet arbre lui-même. Dans le cas des représentations mentales impliquées dans le traitement sémantique du langage, la propriété d'intentionnalité, le fait de posséder un contenu, permet d'expliquer la possibilité même de la communication : si les expressions langagières renvoient à des représentations mentales qui ont un contenu, les expressions langagières servent à communiquer à propos de ce contenu. Comme nous allons le constater, les choses ne sont pas si simples.

Il est important de noter que la notion technique de représentation n'est pas obligatoirement liée à celle de contenu. Il est par exemple courant, dans les modèles de la syntaxe, de postuler l'existence de représentations dont la raison d'être est interne au système et qui n'ont pas de contenu. C'est ainsi que le "pronom" PRO est invoqué pour expliquer certains phénomènes syntaxiques, notamment l'existence des impersonnels. Dans la phrase *il pleut*, l'impersonnel *il* donne une réalité phonologique au pronom PRO, ce qui n'est pas le cas dans les langues dites *pro-drop* comme le persan. Le pronom PRO est une représentation, un *token* qui entre dans le calcul formel proposé par le modèle et dont la présence semble incontournable pour expliquer certains phénomènes syntaxiques. En tant que représentation, le pronom PRO est supposé avoir une réalité cognitive. Pourtant, rien ne permet, dans le modèle, de lui allouer un contenu.

À l'autre extrême, on pourrait penser que certaines représentations ont une intentionnalité tellement affirmée que leur contenu, ce sur quoi elles portent, se retrouve dans leur forme. Selon cette idée, la représentation serait une re-présentation, une sorte de copie d'une autre chose qui lui préexisterait. Cette idée n'est pas entièrement sans fondement. On sait par exemple que certaines propriétés topologiques sont conservées entre la rétine, les relais thalamiques et les différentes cartes corticales du système visuel, si bien que l'image d'une pomme qui se forme sur la rétine est, par certains aspects, copiée dans certaines aires du cortex. Si l'on conçoit le fonctionnement cognitif comme un traitement de type purement analogique, ces images corticales peuvent constituer d'authentiques re-présentations qui peuvent déclencher d'autres traitements, par exemple une réponse motrice consistant à cueillir la pomme. Cependant, de nombreux autres types de traitements, notamment les traitements de type computationnel, supposent des niveaux de représentation dans lesquels les différentes parties des images corticales sont intégrées et ne sont pas conservées. Dans ce cas, la représentation n'entretient plus aucun rapport de ressemblance avec l'image rétinienne. Dans la plupart des modèles, l'intentionnalité des représentations ne suppose pas que la forme de la représentation re-présente son contenu. On doit plutôt concevoir le lien intentionnel comme un renvoi : la représentation renvoie à son contenu, il est possible d'obtenir l'un à partir de l'autre, sans pour autant que les deux se ressemblent. Dans ce cas, l'intentionnalité rejoint d'emblée la notion d'ancrage. Il est important d'insister sur le fait qu'il est parfaitement possible de postuler l'existence, dans le système cognitif, de représentations dotées d'une intentionnalité sans que celles-ci soient des re-présentations.

La question que nous essayons de traiter dans ce travail concerne la nature des représentations mentales associées aux expressions langagières. S'agit-il de simples pions manipulés selon une mécanique aveugle ? Dans ce cas, le caractère intentionnel d'une expression langagière pourrait n'être qu'une illusion. Il serait produit *post hoc* par une opération d'interprétation. La phrase *donne-moi le crayon qui est sur la table* exprime un souhait à propos de la possession du crayon en question. Ce souhait pourrait n'être qu'une interprétation produite par l'interlocuteur et n'avoir aucune réalité cognitive interne à l'esprit du locuteur, pas plus que le souhait exprimé par un ordinateur qui affiche la phrase *j'ai faim* sur son écran. Le caractère intentionnel d'une expression langagière serait ainsi dérivé de l'acte

interprétatif de l'observateur et n'exprimerait aucune propriété intrinsèque au processus cognitif qui a produit cette expression (DAVIDSON 1984 [25]). Sur de telles bases, une théorie du langage se contentera de chercher le caractère intentionnel de la communication langagière dans l'acte de traduction dans une autre langue, dans la prédiction des comportements ou dans l'attribution des valeurs de vérités aux énoncés par l'auditeur.

Ce type d'approche peut convenir pour une théorie behavioriste de la communication langagière. Malheureusement, le pouvoir prédictif que l'on peut attendre d'une telle théorie est limité, si on le compare à celui des théories représentationnalistes. Lorsqu'un locuteur énonce la phrase donne-moi le crayon qui est sur la table, sa phrase ne saurait être le produit d'un simple processus associatif. Il est plausible de supposer qu'elle résulte d'un processus impliquant des représentations. Par exemple, le locuteur désire écrire une lettre, se retrouve dans l'embarras de ne pas avoir de crayon, aperçoit le crayon qui est sur la table, et forme le plan qui consiste à demander à l'interlocuteur un petit service. Les éléments de ce processus, tel qu'on peut scientifiquement le reconstituer, se situent au même niveau que la description qui vient d'être esquissée. Il est, de plus, plausible de postuler qu'il s'agit de représentations qui renvoient aux objets perçus comme lettre, crayon, table. Les règles qui dirigent ce calcul peuvent s'exprimer au même niveau. Par exemple, une telle règle stipule de résoudre toute incompatibilité détectée entre attitudes propositionnelles. Ainsi, le locuteur résout une incompatibilité entre le désir d'écrire la lettre et le constat qu'il n'en a pas les moyens matériels. Si l'interlocuteur réagit en prononçant la phrase ce n'est pas un crayon, mais un porte-mine, c'est qu'il conçoit une incompatibilité entre ce qu'on lui dit et ce qu'il croit pour l'avoir perçu. De même, l'interlocuteur qui répond par la phrase il vaut mieux écrire ta lettre au stylo résout l'incompatibilité entre la croyance que la lettre va être écrite au crayon et le désir de conférer à la lettre un caractère plus officiel. De tels calculs opèrent sur des représentations dotées d'une propriété intentionnelle. Renoncer à l'existence cognitive de telles représentations conduit à renoncer au pouvoir prédictif de ce type de calcul.

À partir du moment où l'on recherche un modèle de la génération et de la compréhension des expressions langagières qui met en jeu plusieurs niveaux de représentations, on peut s'interroger sur l'intentionnalité propre à chaque niveau. Comme nous venons de le voir, le niveau où les phrases sont interprétées suppose une forme d'intentionnalité, celle des attitudes propositionnelles. On pourrait envisager d'en rester là, en supposant que toutes les représentations qui se situent en-dessous, les représentations associées aux syntagmes, aux mots, aux syllabes, *et cætera*, sont dépourvues d'intentionnalité. Dans ce texte, nous allons considérer en détail l'hypothèse selon laquelle il existe des représentations mentales associées aux entités lexicales, les concepts, possédant un caractère intentionnel propre, qui peut être qualifié de sémantique². Les entités sub-lexicales, comme la syllabe, peuvent recevoir d'autres types de contenu, par exemple une forme acoustique. Le point de vue que nous adopterons dans la majeure partie de ce texte est que les concepts sont les représentations mentales minimales dotées d'une sémantique, c'est-à-dire qu'ils s'expriment dans le monde perçu.

On all hands, then, concepts serve both as the domains over which the most elementary mental processes are defined, and as the most primitive bearers of semantic properties. Hence their centrality in representational theories of mind. (FODOR 1994 [36] p. 96)

Une position fréquemment adoptée consiste à considérer que le contenu des concepts se situe dans le "monde", considéré comme objectivement accessible. Cette position semble raisonnable, dans la mesure où l'on utilise les concepts pour communiquer à propos des

² Nous reviendrons sur ce point en introduisant la notion plus stricte de concept lexical (CF. CHAPITRE 6).

entités et des situations du monde. Les concepts, en tant que représentations mentales, auraient ainsi cette propriété de renvoyer à des entités indépendantes de l'esprit humain. Le postulat de l'existence de ce type d'entité est qualifié de réaliste. Les candidats pour ce type d'entité sont des objets concrets, des objets abstraits, des classes d'objets, des attributs d'objets, des propriétés, des faits, des états de faits, *et cætera*. La question de l'existence indépendante de ces entités et de leurs qualités est une question métaphysique qui divise les penseurs en deux groupes, selon qu'ils sont réalistes ou non par rapport aux catégories universelles dans le monde. La contrepartie sémantique de la question revient à se demander si l'on établit un lien entre les concepts et de telles entités externes, ce qui constitue le point de vue externaliste, ou si le contenu des concepts est lui-même une entité mentale, ce qui constitue le point de vue internaliste.

Nous n'entrerons pas dans le débat métaphysique du réalisme, car son issue n'a pas de conséquence sur notre thèse. Nous choisissons, dans ce texte, d'éviter l'hypothèse de l'existence d'un monde structuré indépendant de l'esprit. En revanche, nous adopterons une position, empruntée à la tradition kantienne, qualifiée de réalisme empirique. Le réalisme empirique affirme l'existence des entités mentales qui consistent en notre expérience. Nous parlons ici des expériences au sens large, ce qui inclut la perception mais aussi les sensations et les attitudes. En ce qui concerne le monde extérieur, la seule conviction que nous pouvons avoir quant à son existence est liée au sentiment intuitif que quelque chose nous résiste. L'intuition d'une résistance qui contrarie nos souhaits ou notre volonté nous conduit à faire l'hypothèse d'une réalité extérieure, sans que nous ne puissions dire quoi que ce soit sur la cause précise de cette résistance. Les seules informations auxquelles nous avons accès, en tant que sujet connaissant, nous sont fournies par l'état d'activité de nos capteurs sensoriels. C'est à partir de ces informations et à propos de ces informations que nous formons des expressions langagières.

En marge de cette question métaphysique se pose le problème de choisir entre les versions internaliste et externaliste de l'explication de l'intentionnalité des concepts. Les concepts, en tant qu'entités mentales intentionnelles, portent-ils sur d'autres entités mentales ou portent-ils sur des entités qui sont dans le monde ? Il ne s'agit pas de répondre à cette question dans l'absolu. Lorsque l'on aborde cette question sous l'angle de la modélisation, la tâche est d'expliquer l'établissement du lien intentionnel pour un sujet qui est le produit exclusif d'une phylogenèse et d'une ontogenèse, en renonçant à toute autre source d'information. Lorsque l'on parle de l'objet "ballon" ou de la propriété "rouge" ou de la situation "jeu", il faut lier les concepts correspondants, *id est* les représentations mentales associées à ces mots, à des phénomènes vécus par le sujet. Là encore, l'hypothèse d'un monde structuré peut être évitée, puisque le sujet doit l'intentionnalité de ses représentations mentales à ses dispositions biologiques et aux seules données contenues dans ses expériences³. Nous sommes donc contrainte de renoncer, pour des raisons épistémologiques, aux facilités que procure l'externalisme pour adopter une position que nous appelons internalisme méthodologique : les concepts acquièrent leur contenu dans le domaine des entités mentales associées aux expériences du sujet⁴. Cette position ne présuppose aucun jugement concernant la correspondance entre les concepts et le monde. Elle consiste à proposer d'expliquer la mise en place des concepts par le seul fait des dispositions mentales et des expériences du sujet. C'est dans cette perspective que la notion d'intentionnalité rejoint celle d'ancrage. Le caractère intentionnel des concepts résulte ainsi du fait qu'ils sont, en tant que représentations mentales, ancrés dans les données de l'expérience.

³ Nous discutons plus loin la possibilité d'une intentionnalité innée qui serait le résultat de la phylogenèse, pour conclure que cette possibilité ne change pas fondamentalement le point de vue que nous adoptons ici.

⁴ Nous démarquons plus loin notre position de celle qui est habituellement appelée internaliste.

Dans la perspective de l'internalisme méthodologique, l'intentionnalité des concepts se présente comme un phénomène d'évocation. Les concepts sont intentionnels car il existe des expériences qui les évoquent et qu'ils évoquent. La question de l'intentionnalité des concepts se ramène donc, du point de vue de la modélisation, au problème d'expliquer l'interface entre le système conceptuel et d'autres systèmes de la cognition, notamment la perception, qui produisent les entités mentales associées aux expériences. Cette manière de procéder peut sembler en contradiction avec la position externaliste. C'est en partie le cas. Ceux qui adoptent le point de vue réaliste considéreront qu'il existe trois domaines : les choses du monde, les expériences et les concepts. Ils peuvent imaginer deux interfaces, entre le monde et les expériences d'une part, entre les expériences et le système conceptuel d'autre part. Le projet externaliste consiste notamment à établir un lien intentionnel objectif entre les concepts et le monde en composant ces deux interfaces. Ainsi, là où l'internalisme méthodologique explore des mécanismes d'évocation des concepts dans les données de l'expérience, le point de vue externaliste propose des conditions d'application des concepts aux entités du monde extérieur. Cependant, la définition de la première interface, entre le monde et les expériences, suppose que l'on dispose d'un "oracle", un regard externe qui embrasse les deux domaines. Dans une perspective matérialiste, cet oracle est habituellement constitué par la science. Ceci signifie que les catégories attribuées au monde sont fixées par nos connaissances scientifiques actuelles ou potentielles. Il ne reste ensuite qu'à établir un lien entre ces catégories "réifiées" et la cognition. Une telle méthodologie est toutefois problématique, si l'on admet que la science n'est qu'un produit de la cognition humaine. Il y a donc pétition de principe, puisque c'est l'analyse des capacités cognitives qui doit nous amener aux lois qui gouvernent l'apparition de ces catégories du savoir. On ne saurait ancrer ce qui constitue une partie fondamentale de la cognition humaine, le système conceptuel, dans les catégories provisoires produites par l'activité de la communauté scientifique. Lorsque l'on cherche à élaborer un modèle plausible de la compétence sémantique des humains, on ne peut pas accepter que les nouvelles théories que des physiciens des particules peuvent produire dans leur laboratoire puissent avoir une influence quelconque sur cette compétence. Toutefois, les théories externalistes nous intéresseront dans la mesure où, en dédoublant l'interface concept - monde, nous pouvons leur donner une interprétation internaliste, en particulier en ce qui concerne l'acquisition des concepts. Là où une théorie externaliste postule l'existence d'entités indépendantes de la cognition pour établir un lien intentionnel avec les concepts nouvellement acquis, nous ne nous intéresserons qu'au seul lien d'ancrage entre les concepts et les expériences que ces entités indépendantes sont supposées susciter.

On pourrait reprocher à l'internalisme méthodologique de supprimer la notion de référence. Dans un cadre externaliste, un mot du lexique renvoie à une entité du monde qui constitue la référence de ce mot. L'existence des références externes pour les mots est une hypothèse forte, motivée entre autres par le souci d'expliquer que la communication au moyen du langage soit possible. L'efficacité du langage, dans ce cadre, s'explique de la manière la plus simple qui soit : en prononçant un mot, le locuteur désigne une entité du monde à son interlocuteur. Ainsi, le projet externaliste peut se focaliser, d'une part, sur l'établissement du lien référentiel des mots aux entités du monde, et d'autre part, sur la propagation de ce lien à toutes les expressions langagières. En supprimant toute référence directe au monde, l'internalisme méthodologique semble renoncer à expliquer la communication. Si les mots n'évoquent que des représentations mentales, si l'on ne peut pas se référer à des entités du monde, il semble impossible d'établir une corrélation entre les états mentaux des interlocuteurs. Nous considérons cependant que cette conclusion est erronée et que le problème de la communication langagière mérite précisément une analyse d'un point de vue strictement cognitif, sans un recours aux postulats métaphysiques sur l'existence des entités indépendantes de l'esprit humain. Nous tâcherons plus tard de développer notre point

de vue sur ce sujet (CF. CHAPITRE 9). Dans le cadre que nous adoptons, la notion de référence externe est remplacée par une référence interne. Par exemple, la référence du mot pomme pourra être, dans certains contextes, une image ou un goût, plutôt qu'un élément de nomenclature botanique comme l'exigerait sans doute une position externaliste matérialiste.

La question du contenu d'un concept porte donc sur la référence du terme auquel est associé ce concept. Cette référence appartient à un autre domaine que celui des concepts, le domaine général des expériences. Or, le fait que le concept soit ainsi ancré ne nous renseigne pas sur sa propre nature. Après avoir considéré la cible de l'ancrage, nous abordons maintenant la question d'analyser la forme de ce qui est ancré.

Représentations : la forme

Les modèles représentationnalistes postulent l'existence de représentations de manière à expliquer et à prédire le comportement du système cognitif. Dans la mesure où l'on cherche à décrire ce comportement à l'aide de mécanismes généraux, il est indispensable que les représentations soient dotées d'une structure interne. C'est le moyen par lequel ces mécanismes peuvent les discerner. Lorsque des mécanismes généraux agissent sur des représentations, ils le font d'après leur forme, *id est* d'après leur structure interne, ce qui permet de les différencier des autres représentations. Dans un modèle dynamique, la structure de la représentation résume une partie de l'état du système et est nécessaire, au sein du modèle, pour prédire son trajet avenir. Dans un système computationnel, la notion de forme est encore plus importante, puisqu'elle est supposée avoir une contrepartie matérielle. Les règles du modèle, qui ne sont pas limitées à un simple moyen de description mais sont censées être constitutives du système, manipulent des représentations, elles-aussi matériellement constitutives du système, d'après leur forme. On peut, dans ce cas, considérer que la forme des représentations résume leur propriété causale au sein du système.

Il est difficile de concevoir une théorie de la communication langagière qui soit non représentationnaliste, tant on est frappé par l'apparente complexité des processus en jeu. L'introduction de représentations conceptuelles est, en premier lieu, motivée par le désir, assez naturel, de faire figurer dans le modèle une contrepartie de la notion de signification lexicale qui, nous l'avons vu, est assumée par la propriété intentionnelle des concepts. Un deuxième intérêt d'introduire des représentations conceptuelles est de doter le modèle d'un substrat sur lequel les mécanismes de combinaison sémantique peuvent opérer. À l'agencement des mots dans une expression langagière peut correspondre une combinaison de significations lexicales qui permet de doter l'expression d'un sens. Ainsi, les représentations conceptuelles acquièrent un caractère compositionnel : elles entrent dans des combinaisons pour engendrer d'autres représentations plus complexes. La manière dont les représentations conceptuelles se combinent dépend de leur forme, et la forme du résultat est construite à partir de la forme des constituants. Ainsi, les concepts, de par leur forme, possèdent une propriété causale.

Une autre raison d'introduire des représentations conceptuelles pour expliquer le traitement cognitif du langage est liée au souci de reproduire les inférences qui sont faites spontanément par les sujets. Les représentations conceptuelles semblent présenter un rôle inférentiel : la représentation associée à une expression langagière peut déclencher d'autres représentations, par exemple par un processus de déduction. Là encore, il est raisonnable de tenter d'expliquer les inférences de manière causale à partir de la forme des représentations conceptuelles qui les déclenchent.

La forme des concepts peut donc être invoquée pour expliquer l'interface avec le langage et l'interface avec le raisonnement. Il est même possible de faire reposer sur la forme des concepts leur signification elle-même : pour certains auteurs, la caractérisation de la

forme des concepts est suffisamment contrainte par les interfaces avec le langage et le raisonnement, si bien que toute idée d'ancrage devient superflue. Dans ce cas, la signification des concepts, au moins en partie, est ramenée à leur seule propriété causale. Ce courant de pensée est qualifié d'internaliste. Dans sa version radicale, l'internalisme conceptuel ignore toute idée d'interface avec l'expérience, si bien que le système conceptuel apparaît comme un jeu dépourvu d'ancrage. Il existe cependant des versions modérées de l'internalisme conceptuel dans lesquelles la forme des représentations conceptuelles contribue partiellement à l'explication de l'intentionnalité. Si l'on peut encore parler de contenu dans le cas où l'on s'intéresse aux rapports que les concepts entretiennent entre eux, il s'agit d'un contenu étroit.

The idea of the two-factor version is that there are two components to meaning, a conceptual role component that is entirely "in the head" (this is narrow meaning) and an external component that has to do with the relations between the representations in the head (with their conceptual roles) and the referents and/or truth conditions of these representations in the world. [...] The internal factor, conceptual role, is a matter of the causal role of the expression in reasoning and deliberation and, in general, in the way the expression combines and interacts with other expressions so as to mediate between sensory inputs and behavioral outputs. (BLOCK 1986 [7] p. 627)

Dans cette approche, qualifiée de sémantique du rôle conceptuel (*Conceptual Role Semantics*), le concept est déterminé par sa fonction au sein du système conceptuel, et c'est cette fonction qui lui procure, en partie, sa propriété intentionnelle. Il s'agit là d'un changement de perspective par rapport à la manière dont nous avons introduit la notion de concept. Nous avons caractérisé les concepts en tant que représentations dotées d'un contenu dans le domaine de l'expérience, avant de leur conférer une fonction au sein du système conceptuel. L'internalisme conceptuel, quant à lui, procède de manière inverse en caractérisant les concepts par le biais de la fonction qu'ils remplissent dans le comportement de l'individu qui les possède.

A possession condition for a particular concept specifies a role that individuates that concept. The possession condition will mention the role of the concept in certain transitions that the thinker is willing to make. These will be transitions that involve complete propositional thoughts involving the concept. In some cases they are inferential transitions; in others they are transitions from initial states involving perceptual experience. Normally, a possession condition has several clauses, each treating of a different kind of case. (PEACOCKE 1992 [81] p. 107)

Les concepts logiques servent de modèle à cette approche. Par exemple, le concept de conjonction peut être caractérisé par sa fonction dans la réalisation de certaines inférences par les individus qui le possèdent (PEACOCKE 1992 [81]). Cette fonction peut être représentée par l'ensemble de deux règles d'élimination ($P \wedge Q \vdash P$ et $P \wedge Q \vdash Q$) et une règle d'introduction ($P, Q \vdash P \wedge Q$). La possession du concept de conjonction est ainsi caractérisée d'une manière qui garantit la production de toutes les inférences ayant la forme de l'une des trois règles citées. Si l'on généralise cette méthode, il s'agit de déterminer, pour chaque concept, l'ensemble des clauses nécessaires qui cernent sa fonction cognitive. Cette fonction cognitive s'étend au pouvoir d'évocation du concept qui le lie aux mécanismes perceptifs ou effecteurs.

Certes, il nous paraît naturel que le postulat des concepts s'accompagne d'une description exacte de la fonction qu'ils sont censés remplir dans le système cognitif humain. Cependant, nous pensons que, dans une approche représentationnaliste, cette tâche ne peut être entamée qu'en définissant, de manière claire, d'une part le domaine où les concepts

acquièrent leur contenu, et d'autre part les mécanismes qui agissent sur les concepts d'après leur forme. Dans notre schéma de base, le système conceptuel retient trois interfaces avec la perception, le langage et le raisonnement. Notre internalisme méthodologique exige que les concepts acquièrent leur contenu dans le domaine de l'expérience. Ainsi seule l'interface avec la perception fournit aux représentations conceptuelles l'ancrage nécessaire pour qu'ils possèdent une propriété sémantique. Les deux autres interfaces, quant à elles, permettent d'établir des liens associatifs, des liens de déclenchement par appariement ou des liens de contrôle, qui lient les mécanismes à l'œuvre dans le système conceptuel aux mécanismes du langage et du raisonnement. C'est l'explication de ces mécanismes qui nécessite l'analyse de la forme des représentations conceptuelles⁵. Or, selon l'hypothèse de l'internalisme conceptuel, l'ensemble des relations que les représentations conceptuelles peuvent entretenir avec d'autres entités mentales intervient dans la détermination de leur contenu. Notamment, les relations inter-conceptuelles prennent un rôle déterminant dans la définition du contenu étroit des concepts. Nous estimons qu'il s'agit ici d'un emploi métaphorique de la notion de contenu : dans un modèle représentationnaliste du système conceptuel, les relations inter-conceptuelles ne peuvent intervenir que dans la détermination de la structure interne des représentations conceptuelles ; ce n'est que d'une manière indirecte que cette structure peut être pertinente pour l'ancrage des représentations structurées⁶. Répétons que ce que nous appelons internalisme méthodologique n'est pas lié par l'exigence posée par l'internalisme conceptuel qui veut que le contenu des concepts soit déterminé, pour l'essentiel, de manière fonctionnelle. Notre position laisse par exemple la possibilité que chaque concept soit directement ancré dans l'expérience.

La notion de forme, en conférant un caractère fonctionnel aux représentations conceptuelles, réintroduit l'idée du sens des mots. La référence d'un mot est donnée par le contenu du concept correspondant. Son sens est donné, concrètement, par la forme de ce concept. La forme du concept précise la fonction du mot dans la communication langagière, c'est-à-dire les rapports qu'il entretient par son sens avec les autres expressions langagières. Le sens du mot pomme est ainsi représenté par la forme du concept associé qui lui donne la possibilité d'intervenir dans un syntagme comme une pomme rouge, ou dans une assertion comme la pomme est un fruit sucré.

À partir du moment où les concepts possèdent une forme, il faut se poser la question du rapport entre cette forme et celle de la représentation non conceptuelle qui l'évoque et qu'il évoque. On peut imaginer qu'elles soient de même nature, qu'elles soient qualitativement différentes mais que leurs structures se correspondent, ou enfin qu'il n'y ait aucun lien entre les deux. Dans les deux premiers cas, l'ancrage devient trivial. Dans le troisième cas, le problème de l'ancrage se révélera particulièrement délicat à résoudre.

3.2. Percepts intégrés

L'une des fonctions remarquables attribuées aux concepts est celle de la catégorisation. Les concepts nous permettent d'avoir un jugement sur les données de la perception, jugement qui est présenté comme binaire dans certains modèles. Ainsi, le concept LIVRE nous permet de catégoriser les objets du monde perçu de manière à en extraire, de manière plus ou moins précise, tout ce qui mérite d'être qualifié de livre.

⁵ Nous développerons cette analyse en posant des questions sur, d'une part, le rôle inférentiel, et d'autre part, le caractère compositionnel des concepts (CF. CHAPITRES 4 & 5).

⁶ Nous aurons l'occasion de reparler en détail des motifs et des méthodes pour structurer les concepts lexicaux, et des conséquences d'une telle entreprise (CF. CHAPITRES 7).

La manière la plus simple d'expliquer la capacité de catégorisation est sans conteste de l'imaginer comme une activité de comparaison. Ainsi, est une instance du concept LIVRE tout ce qui ressemble au livre typique. La caractéristique fondamentale de cette approche est que les deux éléments de la comparaison, le percept à catégoriser et le concept qui sert d'étalon sont de même nature. Son intérêt est de résoudre de la manière la plus simple qui soit la question de l'intentionnalité : non seulement le percept typique est le meilleur candidat pour constituer le contenu du concept, mais il n'y a plus de raison de distinguer le concept de son contenu. Les concepts, selon cette approche, ne sont rien d'autre que des percepts moyens, issus d'une intégration d'expériences multiples. Il s'agit donc, à la base, d'une théorie purement empirique des concepts.

Avantages de l'approche empiriste des concepts

L'intérêt premier de l'approche empiriste est de résoudre le problème de l'ancrage. Les concepts étant des percepts moyens, *id est* des invariants perceptifs, les concepts sont *ipso facto* ancrés dans l'expérience. Cette forme d'ancrage présente l'avantage d'être économique : contrairement aux théories qui postulent l'existence d'un espace conceptuel qualitativement distinct de celui des expériences, il n'y a pas, dans l'approche empiriste, de duplication des représentations. Ainsi, nous pouvons exprimer le fait qu'un zèbre possède des rayures parce que les rayures sont présentes dans la perception du zèbre. Alors qu'un système symbolique est obligé de dupliquer cette information sous une forme explicite, l'approche empiriste se contente d'exploiter l'information qui est disponible dans les données de la perception. Du point de vue de la modélisation, il s'agit là d'un avantage considérable, car toute duplication pose le problème de la complétude et de la cohérence.

L'ancrage empiriste résout un autre problème délicat, celui de l'acquisition. L'enfant acquiert de nouveaux concepts en généralisant ses percepts. Pour réaliser cette opération inductive, plusieurs mécanismes simples ont été proposés. Le mécanisme empiriste par excellence est un mécanisme statistique d'extraction de régularités. Ainsi, ne sont retenues dans le percept moyen que les caractéristiques qui sont fréquemment présentes dans les instances à partir desquelles il est construit. C'est ainsi que fonctionnent les réseaux de neurones artificiels classiques, qu'ils soient supervisés ou non.

One of the basic principles which drives learning in neural networks is *similarity*. Similar inputs tend to yield similar outputs. Thus, if a network has learned to classify a pattern, say 11110000, in a certain way then it will tend to classify a novel pattern, *e.g.*, 11110001, in a like fashion. Neural networks are thus a kind of analogy engine. The principle of similarity is what lets networks generalize their behaviors beyond the cases they have encountered during training. (ELMAN & BATES & JOHNSON & PARISI 1996 [32] p. 59)

Cette capacité de généralisation est à la base de la formation des percepts moyens. Le système extrait les régularités des exemples qui lui sont soumis. Les aspects contingents, qui varient d'un exemple à l'autre, ne laissent pas de trace, alors que les aspects invariants sont mémorisés ensemble. Une fois que cet apprentissage a eu lieu, le système peut reconnaître un objet qu'il n'a jamais perçu, à condition que cet objet possède un nombre suffisant de caractéristiques en commun avec l'un des ensembles invariants qui ont été mémorisés. Certains problèmes cognitifs ne possèdent pas cette propriété de continuité selon laquelle des entrées proches correspondent à des sorties proches. Un contre-exemple bien connu est celui de l'induction de la grammaticalité des énoncés du langage (CHOMSKY 1975 [15]). Néanmoins, la formation des catégories semble posséder, pour l'essentiel, cette propriété de continuité que les réseaux de neurones permettent de reproduire.

Noter que l'apprentissage statistique par extraction de régularités n'est pas le seul mécanisme permettant la formation de catégories. La théorie de la Gestalt et le constructivisme peuvent être présentés comme des approches fondées sur des mécanismes d'induction de "bonnes formes" à partir d'un ensemble réduit d'exemples (DESSALLES 1998 [29]). Dans ce cas, les concepts sont ces bonnes formes acquises par l'expérience qui servent ensuite à catégoriser les situations.

Le fait de ramener l'ancrage des concepts à un mécanisme de catégorisation possède plusieurs autres mérites. Le moindre n'est pas de pouvoir gérer la gradualité des jugements de catégorisation, ce que bien des systèmes symboliques sont incapables de faire. Ainsi, bien qu'un gland ne soit pas un chêne, le gland qui germe et qui pousse constitue chaque jour une meilleure instance du concept CHÊNE, sans qu'aucune discontinuité ne soit jamais franchie. De même, les liens de ressemblance qui peuvent unir les concepts, les airs de famille, sont naturellement pris en charge par la ressemblance des représentations dans l'espace perceptif. Ainsi, nous n'avons pas besoin d'explication pour percevoir la ressemblance entre une puce électronique et une vraie puce. Un tel appariement sera beaucoup plus délicat à effectuer pour des représentations conceptuelles ne comportant que des descriptions symboliques explicites.

Inconvénients de l'approche empiriste des concepts

Si, malgré le caractère impressionnant de ces points positifs, de très nombreux auteurs ont imaginé un système conceptuel disjoint du système perceptuel, c'est que les inconvénients de l'approche empiriste leur ont semblé rédhibitoires. Une partie de ces inconvénients est liée aux deux autres interfaces, laissées de côté par l'approche purement empiriste : l'interface avec le langage et l'interface avec le raisonnement.

Le principal problème que la formation empirique des concepts ne permet pas de résoudre est celui des phénomènes systématiques. Les concepts sont impliqués dans des opérations systématiques, que ce soit dans la catégorisation, les inférences ou la composition, qui sont incompatibles avec le caractère fondamentalement graduel inhérent aux mécanismes statistiques postulés dans l'approche empiriste. Par exemple, une entité incluse dans le concept VIVANT doit être systématiquement exclue du concept MORT. Or, pour un système empiriste, toute entité perçue est plus ou moins vivante, certaines franchement, d'autres presque pas, mais le lien entre l'entité perçue et le concept VIVANT n'est jamais strictement nul, car la statistique ménage toujours une ressemblance, même très faible, entre deux représentations quelconques. La frontière entre les concepts VIVANT et MORT n'est donc pas systématique comme il se devrait. De la même manière, la systématisme de l'inférence selon laquelle certaines entités perçues peuvent mourir dès lors qu'elles sont vivantes n'est pas assurée, car l'induction statistique dont elle dépend n'offre aucune garantie. Enfin, la systématisme de composition est également absente : rien ne garantit un sens systématique à une construction comme X est malade dès lors que la variable X désigne une entité vivante.

La nature statistique de l'apprentissage empiriste pose un deuxième problème. Un système empirique est peu sensible aux exceptions et ne leur associe pas d'explication. Dans un réseau de neurones classique, les exceptions n'étant pas représentatives de la classe, sont ignorées en tant que points aberrants de l'échantillon statistique. L'apprentissage d'une exception, par exemple une entité vivante immortelle, demande donc un renforcement répété, d'autant plus important que le lien conceptuel VIVANT - MORT a été fortement imprimé, jusqu'à ce que l'exception puisse constituer une classe à elle seule. Cet apprentissage ne produit jamais un résultat strict, et surtout mémorise l'exception sans pouvoir en stocker la raison. Or, un enfant humain qui lit une fiction comprend aisément, en une seule fois, que tel personnage peut traverser les âges sans vieillir parce qu'il possède un don particulier. Si son

fonctionnement conceptuel obéissait aux seules lois de l'apprentissage statistique empiriste, l'enfant serait incapable de comprendre ce type d'histoire.

Un autre grand problème rencontré par la conception empiriste du système conceptuel est la dépendance du système par rapport à son histoire. Pour que deux individus aient des concepts comparables, il faut qu'ils aient connu des expériences similaires en moyenne, qualitativement et quantitativement. Or, cette exigence est problématique dans les domaines où les individus n'ont eu accès qu'à des expériences limitées. Par exemple, un individu ne pourra pas communiquer à propos de ce qui est nouveau pour les autres. Il devra se limiter aux expériences communes, ce qui offre une piètre modélisation de la communication humaine. Pour expliquer que les enfants parviennent à des états de connaissances similaires malgré des expériences diverses et limitées, certains auteurs ont émis l'hypothèse que ces états de connaissances constituaient de "bonnes formes", dotés par exemple de la propriété de clôture opérationnelle, *id est* de fermeture par composition (PIAGET 1945 [83]). L'intérêt des bonnes formes est que leur apprentissage ne nécessite que peu d'exemples. Par exemple, le concept de justice, développé en plusieurs étapes par les enfants, présente de nombreuses symétries. L'enfant passe d'un état égo-centré, dans lequel est juste tout ce qui lui profite, à un état de justice réciproque dans lequel il considère comme injuste vis à vis des autres ce qu'il n'aimerait pas qu'on lui fasse (PIAGET 1932 [82]). Ce concept de justice, que l'on peut noter juste(acteur , patient , acte), est invariant pour toutes les substitutions d'individus aux deux variables acteur et patient. Ceci inclut en particulier la possibilité d'adopter le point de vue de l'autre, ce dont le très jeune enfant est encore incapable. C'est parce que le concept adulte de justice présente de nombreuses symétries qu'il est atteint de manière identique par des enfants différents, soumis à des expériences différentes. Les concepts produits par généralisation statistique à partir de peu d'exemples constituent également des bonnes formes : la catégorie construite à partir de la généralisation des percepts est invariante pour toutes les transformations qui affectent les dimensions qui sont laissées de côté dans la généralisation. Ainsi, un concept empirique POMME peut ignorer la couleur précise de la pomme, sa taille, la présence de taches, *et cætera*. La catégorie créée à partir d'un percept moyen de pomme sera invariante pour toutes les transformations affectant ces aspects "anecdotiques".

Le fait que différents apprenants, exposé à des expériences limitées et différentes, convergent vers des bonnes formes n'est pas fortuit. Les mécanismes d'apprentissage de portée générale, comme les mécanismes empiristes ou constructivistes, permettant d'induire des catégories, ont la propriété d'être isotropes : ils sont insensibles à des changements de repère. Or, de tels mécanismes isotropes produisent des bonnes formes, *id est* des formes invariantes pour de nombreuses transformations (DESSALLES 1998 [29]). Dans le domaine conceptuel, il existe certainement des bonnes formes. Pour autant, on ne peut pas considérer que le système conceptuel se résume à un catalogue de bonnes formes. En particulier, les rapports qu'un concept donné entretient avec d'autres concepts ne présentent pas la belle symétrie qu'on est en droit d'attendre d'un système d'induction général. Par exemple, les relations conceptuelles évoquées au cours d'une conversation déterminent un ensemble de situations possibles de très faible symétrie (DESSALLES 1993 [28]). Si le seul mécanisme d'apprentissage disponible était un mécanisme inductif isotrope, empiriste ou constructiviste, la constitution des ensembles de connaissances que nous mobilisons dans les conversations serait inexplicable. Cette situation reproduit, au niveau conceptuel, la difficulté d'expliquer l'acquisition des connaissances syntaxiques par des moyens empiriques ou constructivistes (CHOMSKY 1975 [15])⁷.

⁷ Nous aurons l'occasion de développer cette critique plus tard (CF. ANNEXE).

Le fait de considérer qu'il n'y a pas de différence de nature entre les percepts et les concepts, et que ceux-ci sont identifiables à des percepts moyens, est une idée à la fois très séduisante et très décevante. À la fin de ce travail, nous essaierons de montrer que cette idée et ses avantages peuvent être récupérés, à condition de compléter le dispositif inductif par un mécanisme dynamique de production de symboles. Auparavant, nous allons considérer en détail l'option opposée, selon laquelle les concepts sont d'une autre nature que les percepts.

3.3. Symboles

L'idée de symbole s'oppose par plusieurs aspects à la notion de percept intégré. Un percept intégré ne peut, au mieux, que représenter une catégorie d'objets perçus, alors qu'un symbole est avant tout un élément combinable ayant, par ailleurs, des propriétés représentationnelles.

[...] Nor can categorical representations yet be interpreted as « meaning » anything. It is true that they pick out the class of objects they « name », but the names do not have all the systematic properties of symbols and symbol systems [...]. They are just an inert taxonomy. For systematicity it must be possible to combine and recombine them rulefully into propositions that can be semantically interpreted. « Horse » is so far just an arbitrary response that is reliably made in the presence of a certain category of objects. There is no justification for interpreting it holophrastically as meaning « This is a (member of the category) horse » when produced in the presence of a horse, because of the other expected systematic properties of « this » and « a » and the all-important « is » of predication are not exhibited by mere passive taxonomizing. (HARNAD 1990 [46] p. 343)

La tradition empiriste postule des représentations conceptuelles qui entretiennent des rapports de ressemblance avec les données de la perception. La topologie, voire la métrique de ressemblance, au sein de l'espace des concepts, est la même que dans l'espace des percepts. Dans beaucoup de modèles, ceci est dû au fait qu'il s'agit du même espace. Cette situation contraste avec les autres modèles, dans lesquels ce lien de ressemblance n'existe pas. Lorsque l'on met en avant le caractère symbolique des concepts, que l'on dote les concepts de propriétés formelles pour permettre leur inclusion dans des calculs dictés par les mécanismes grammaticaux ou déductifs, les représentations obtenues perdent nécessairement tout lien de ressemblance et toute relation topologique ou métrique avec leur contenu. Cette perte, qui se constate dans les différents modèles qui ont pu être proposés, est due au fait que les mécanismes généraux qui conservent les relations de distance et de voisinage possèdent la propriété d'isotropie, alors que les mécanismes compositionnels qui engendrent les représentations conceptuelles complexes en sont dépourvus (DESSALLES 1998 [29]).

Si aucune ressemblance n'existe entre les concepts et les percepts, si l'organisation des concepts ne reproduit en rien l'organisation des percepts, alors l'ancrage des concepts doit être le fruit d'un "apprentissage", soit ontogénétique, soit phylogénétique. Or l'établissement de ce lien d'ancrage pose le problème de la connexion. Par quel mécanisme les concepts, qui sont ici des entités symboliques, peuvent-ils recevoir un contenu dans l'espace des percepts ? On comprend aisément comment un mot M peut se retrouver associé à un percept P . Il suffit que M soit prononcé en présence du sujet lorsque celui-ci perçoit P . Si le mot M doit renvoyer en même temps au concept C , il faut déterminer le mécanisme par lequel le triangle $M \square P \square C$ se retrouve complété. Or C , en tant qu'entité symbolique propre au système conceptuel, est en quelque sorte hors d'atteinte.

Ce problème de la connexion concerne toutes les théories symboliques des concepts. Certes, certaines théories parviennent à dériver l'ancrage des concepts de l'ancrage de leurs composants. Ainsi, si le concept TULIPE est défini, au sein du système conceptuel, à l'aide du concept FLEUR, il hérite une partie de son ancrage de ce dernier. Cette méthode peut, en principe, résoudre le problème de l'ancrage pour la plupart des concepts. Il n'en reste pas moins que la connexion restera inexplicée pour certains concepts, dits primitifs, à partir desquels tous les autres sont construits.

Nous allons considérer successivement deux types d'explication de l'établissement de l'intentionnalité, invoquant respectivement la communication et l'information. Au départ, ces explications ont été produites dans un contexte externaliste, pour expliquer que les concepts se trouvent reliés à des entités du monde. Dans chaque cas, nous examinerons comment ces cadres théoriques peuvent être transposés dans notre propre cadre, en prenant l'explication de l'intentionnalité comme un mécanisme d'ancrage. Même dans l'hypothèse d'un monde objectif, notre expérience reste le seul lien que nous pouvons avoir avec ce monde. L'intentionnalité des concepts, comprise dans ces théories comme le lien qui unit ces concepts au monde, vient donc nécessairement de la composition de deux mécanismes, l'intégration de l'expérience perceptive et l'ancrage.

3.4. Intentionnalité et communication

Certaines approches n'abordent pas la question de l'intentionnalité comme un problème de connexion entre des représentations mentales, mais comme un problème d'association entre des mots et des catégories du monde. Pour de telles approches, qui adoptent une perspective externaliste, il est essentiel d'expliquer comment le langage parvient à décrire le monde de manière "correcte". Comment un individu donné parvient-il à connaître l'ensemble des liens mot - référence, même pour des entités du monde qu'il n'a jamais rencontrées ? La réponse suppose que le lien causal entre le mot et sa référence a généralement une origine qui dépasse l'histoire de l'individu. Si je désigne un objet précis de mon champ visuel en prononçant le mot *arbre* et que mon interlocuteur saisit mon intention, c'est parce que mon interlocuteur a appris ce mot quand il était enfant et que je l'ai moi-même appris en cours de français quand mon professeur désignait une image dans le livre de cours. Pour que je puisse communiquer le contenu du mot *arbre*, ce contenu a dû m'être communiqué. Il a été aussi communiqué à mon professeur, et ainsi de suite le long d'une chaîne d'usage jusqu'à la première fois où un individu a glosé l'objet par le mot. La communication assure le lien entre le mot et sa référence et le propage.

Le cas typique, qui sert de modèle à cette approche, est celui des noms propres. Les noms propres sont le résultat d'un acte de baptême. En donnant un nom à un individu, la communauté décide de le désigner par une étiquette qui glosera désormais l'enfant pour ceux qui le connaissent. Le nom constitue alors le moyen le plus simple pour pouvoir communiquer verbalement à propos de cet individu. Grâce à cette étiquette conventionnelle, on peut même parler de l'individu avec quelqu'un qui ne le connaît pas. La personne sait que l'étiquette ne fait que remplacer sa référence.

Someone, let's say a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard about Feynman or from whom he ever heard of Feynman. (KRIPKE 1972 [62] p. 91)

L'élargissement de la métaphore de baptême initial aux noms communs semble assez intuitif quand il s'agit des espèces naturelles ou des artefacts avec lesquelles les individus sont en interaction directe. On peut imaginer que l'on apprend des mots comme chat ou voiture quand ils sont utilisés pour désigner des objets de la vie quotidienne. Grâce à la chaîne communicationnelle, il n'est même pas obligatoire que l'individu ait eu une expérience directe des objets désignés, si bien que la communication par ostension n'est pas nécessaire pour l'apprentissage individuel.

[...] the species-name may be passed from link to link, exactly as in the case of proper names, so that many who have seen little or no gold can still use the term. Their reference is determined by a causal (historical) chain [...]. (KRIPKE 1972 [62] p. 139)

Si les individus ne peuvent pas profiter de l'ostension pour établir le lien mot - référence, il existe un autre moyen, la description, par lequel ils accèdent au moyen d'utiliser le mot dans les circonstances adéquates. Dans le cas des entités abstraites ou imaginaires désignées par des mots comme bourse ou Pégase, le baptême initial est plus difficile à concevoir. Il suffit dans ce cas d'imaginer que la description sert non seulement à contraindre le lien mot - référence pour tous les individus de la chaîne de communication, mais qu'elle sert également à effectuer le "baptême" de l'entité pour la communauté (KRIPKE 1972 [62]). Ainsi, la première description qui a été faite de Pégase en tant que cheval ailé a permis d'établir un lien mot - référence qui s'est ensuite propagée d'individu en individu, si bien que le mot Pégase, sans être équivalent à la description qui lui est associée, possède une référence pour tous les individus de la communauté.

Cette approche de la notion de contenu repose sur deux considérations de base. La première est liée au caractère indexical d'une grande partie des expressions langagières (PUTNAM 1975 [87]). Selon cette idée, la fonction principale du mot arbre est de désigner sa référence. La seconde est d'ordre social : le fait que l'on utilise le terme arbre et non un autre terme pour désigner sa référence est le résultat d'un accord dans la communauté de communication (PUTNAM 1975 [87]).

La question qui nous importe, dans ce travail, est de savoir comment une telle théorie du contenu peut s'interpréter dans une approche cognitive représentationnaliste du langage naturel. Notre but n'est pas de savoir comment les individus parviennent à employer les mots d'une manière objectivement correcte, mais plutôt de rendre compte de la capacité des individus à comprendre les énoncés langagiers de manière à effectuer certaines inférences et à produire d'autres énoncés appropriés. Dans un cadre représentationnaliste, le mot, qui n'est autre qu'une forme phonologique propre à la langue parlée dans la communauté, peut évoquer une représentation mentale chez les locuteurs de cette langue. C'est cette représentation mentale, le concept, qui déclenche les processus compositionnels et inférentiels. Pour que cela soit possible, il faut que le concept soit lié, à la suite d'un apprentissage, au mot qui l'évoque. Si l'on accepte que le concept est lié à un certain type de perception, par exemple la perception d'un lien causal entre un phénomène et ses effets, alors il suffit que le mot soit associé une fois à cette situation causale pour qu'il se retrouve associé au concept pour tous les individus de la communauté linguistique.

They could all use the term “electricity” [...] what they do have in common is this: that each of them is connected by a certain kind of causal chain to a situation in which a *description* of electricity is given, and generally a *causal* description – that is, one which singles out electricity as the physical magnitude *responsible* for certain effects in a certain way. [...] Let us call this event – my acquiring the ability to use the term “electricity” in this way – an *introducing event*. It is clear that each of my later uses will be causally connected to this introducing event, as long as those uses exemplify the ability I required in that introducing event. (PUTNAM 1975 [87] p. 199)

De cette manière, le lien mot - concept peut s'expliquer, au même titre que le lien mot - objet considéré par ces auteurs, par une chaîne historique causale qui associe, à chacun de ses maillons, une représentation phonologique et une représentation conceptuelle. Le point essentiel que l'on peut retenir de cette approche historique et conventionnelle de la signification est que le fait que les mots aient un sens suppose que les individus interagissent de manière causale avec leur environnement et interagissent de manière causale entre eux, par le biais de la communication.

Cette manière de concevoir la création de nouvelles significations et leur propagation laisse une part minimale à l'apprentissage et aux difficultés qu'il présente. Le problème est de rendre plausible le fait que deux individus finissent par accorder la même signification au même mot. Même dans le cas des noms propres, la réponse ne va pas de soi. Lors de l'événement baptismal, les personnes présentes ont l'occasion d'associer le nom non pas à l'enfant lui-même, en tant qu'entité objective, mais à une perception qu'ils en ont. Il faut qu'un mécanisme leur permette de généraliser le nom à l'individu quel que soit son âge, leur évite d'associer le nom à l'enfant seulement lorsqu'il est dans son landau, *et cætera*. La situation est encore plus problématique dans le cas des espèces naturelles. À partir du moment où l'on renonce à la simplification qui consiste à penser que les entités du monde nous apparaissent directement, détachées de toutes les autres, l'établissement des associations mots - concepts, médiatisées par les associations mots - percepts, devient moins évident. Il s'agit d'expliquer ce qui nous permet de réemployer le même mot chat dans des situations différentes dans lesquelles le contexte d'apparition de l'animal peut varier du tout au tout. Il faut postuler l'existence de capacités inductives puissantes qui permettent à deux individus, ou à un même individu dans deux situations différentes, de repérer des éléments communs qui provoquent l'utilisation appropriée d'un même mot. L'explication du lien historique causal de dénomination y perd la simplicité qui faisait son attrait. S'il existe un mécanisme psychologique complexe par lequel les individus parviennent à déterminer la référence d'un mot en contexte, le fait que ce mot soit prononcé en leur présence dans des circonstances appropriées, bien que nécessaire, ne joue qu'un rôle marginal dans l'établissement de la signification.

Il est également difficile de se contenter d'invoquer l'histoire des interactions lorsqu'il s'agit d'expliquer l'établissement du sens des mots, notamment les mots abstraits, dont la signification est obtenue par des descriptions utilisant d'autres mots. L'interprétation des descriptions suppose là encore des mécanismes riches, laissant à l'origine historique des associations mot - objet une part anecdotique. L'importance des mécanismes cognitifs est encore plus flagrante lorsque la description sert à désigner une situation dont l'individu est censé inférer le concept décrit. Ainsi, une description comme ce qui surnage lorsqu'on fait cuire du lait suppose des capacités à constituer un concept à partir d'une situation imaginée ou remémorée. Tout sujet qui apprend un mot nouveau pour lui est confronté à un problème de même difficulté que celui qu'ont connu ses ancêtres lors de la création du lien mot - objet dans la communauté.

La question de l'émergence de significations dans une communauté a été explorée expérimentalement à l'aide de simulations (KAPLAN 2001 [57]). Il apparaît que l'acte initial de dénomination d'un objet perçu dans l'environnement par un agent ne suffit pas à l'acquisition du concept par les autres agents. Il faut que ceux-ci aient une perception comparable, puis affinent leurs catégories lors des utilisations ultérieures du même mot. Les agents parviennent ainsi à former des catégories conceptuelles pertinentes en affinant des divisions selon une procédure dichotomique, par exemple sur la couleur ou la forme des objets, de manière à discriminer les objets présents. Il peut en résulter des dialogues de sourds : tel objet sera signalé par le mot *gorewa* par l'agent 1, ce qui, pour lui, signifie "rouge" ; l'agent 2, entendant *gorewa*, comprendra par exemple "triangle". La communication peut réussir si le seul objet rouge se trouve être triangulaire. Heureusement, confrontés à d'autres situations, l'un des deux agents au moins sera amené à réviser son lexique.

Ce type d'expérience montre que l'histoire de la communication entre agents, qui ressemble plus à un réseau complexe d'interactions qu'à une simple chaîne de communication, se révèle déterminante pour l'adoption d'un mot plutôt qu'un autre. L'histoire des interactions en contexte permet donc, en principe, d'expliquer l'émergence d'un ensemble donné de concepts dans une communauté. En revanche, il serait illusoire de vouloir fonder une théorie des concepts en se limitant à l'historique des interactions. Les concepts que forment les robots préexistent en puissance dans leur capacité à affiner leurs catégories et à les associer avec des mots. L'observateur pourrait être tenté de penser que la signification des mots qu'ils emploient se situe dans le monde simplifié qui s'offre à leur caméra. Ainsi, le mot *gorewa* semble désigner le triangle rouge sur le tableau. Pourtant, ces robots diffèrent suffisamment de nous pour que nous soyons obligés de reconnaître qu'il n'en est rien. La signification de leurs mots se situe dans leur mémoire, dans les structures qu'ils ont affinées pour répondre aux besoins de la communication qui joue davantage le rôle de déclencheur des mécanismes internes qu'un rôle structurant. Ces robots, pas plus que les humains, ne sont télépathes. Ils ne peuvent pas se transmettre leurs catégories conceptuelles. Ils ne peuvent qu'échanger des mots en situation et employer leurs mécanismes internes pour tenter de parvenir à des catégories compatibles avec leur perception.

Il est indéniable que l'association entre les mots et les perceptions ne doit rien, la plupart du temps, à la forme phonologique des premiers. Les associations mots - percepts ne sont donc possibles que si certains individus les créent et les propagent. Cependant, comme les situations ne se présentent jamais deux fois sous le même aspect, le lien à expliquer est celui qui s'instaure entre le mot et le concept, ce dernier permettant de discriminer les percepts. L'étude des interactions entre individus ne permet pas de faire l'économie d'un mécanisme inductif par lequel ces individus parviennent à forger des concepts compatibles avec l'usage des mots par leurs partenaires.

3.5. Intentionnalité et information

L'histoire des interactions langagières ne peut suffire à rendre compte de l'intentionnalité des concepts : elle ne peut qu'expliquer pourquoi ce sont certains percepts qui sont nommés dans la communauté plutôt que d'autres. Si l'on s'intéresse au résultat de ce processus historique pour l'individu, force est de constater que celui-ci a acquis une capacité lui permettant d'employer les mots à bon escient. Une manière minimale de décrire cette capacité est de la présenter comme une disposition comportementale.

[...] a stimulation σ belongs to the affirmative stimulus meaning of a sentence S for a given speaker if and only if there is a stimulation σ' such that if the speaker were given σ' , then were asked S, then were given σ , then were asked S again, he would dissent the first time and assent the second. [...] What now of the strong conditional, the "would" in our definition of stimulus meaning? [...] what the conditional defines is a disposition, in this case a disposition to assent to or to dissent from S when variously stimulated.

(QUINE 1960 [90] p. 32)

La présence de ces dispositions chez les individus est le résultat d'un apprentissage. Dans la doctrine behavioriste, le mécanisme d'apprentissage par lequel on acquiert de nouvelles significations ne peut être qu'un renforcement positif ou négatif.

It remains clear in any event that the child's early learning of a verbal response depends on society's reinforcement of the response in association with the stimulations that merit the response, from the society's point of view, and society's discouragement of it otherwise.

(QUINE 1960 [90] p. 82)

Pour ce genre d'approche, le fait de posséder des concepts n'est rien d'autre qu'une disposition comportementale conditionnée. On ne peut donc pas parler de contenu pour les concepts. Toute notion même de sémantique est exclue d'un tel schéma (QUINE 1960 [90]). Si l'on est représentationnaliste et que l'on fait découler les dispositions comportementales langagières de la possession de représentations conceptuelles, le seul point à retenir d'une telle théorie est le lien statistique, fruit d'un renforcement répété, qu'elle postule entre les concepts et les percepts.

La reconnaissance de ce lien statistique empêche de ne voir dans les concepts que le simple résultat d'un acte de désignation propagé à travers l'histoire des usages. Selon cette hypothèse, les sujets sont capables d'identifier l'objet qu'on leur désigne et de le reconnaître dans les situations ultérieures parce qu'ils ont forgé un concept qui leur permet de le discriminer (DRETSKE 1981 [31]). Cette fonction de discrimination justifie, d'un point de vue théorique, l'existence de concepts dotés d'un contenu. Le mot désigne un objet, mais cet objet n'existe perceptivement que parce qu'un concept le discrimine. Dans ce genre d'approche où, rappelons-le, les concepts ne sont pas de même nature que les percepts, le pouvoir de discrimination des concepts est attribué au fait que le concept co-varie avec l'objet perçu (DRETSKE 1981 [31]). Cette co-variation a la vertu d'expliquer également l'acquisition de ce pouvoir de discrimination.

In teaching someone the concept *red*, we show the pupil variously colored objects at reasonably close range and under normal illumination. That is, we exhibit the colored objects under conditions in which *information* about their color is transmitted, received and (hopefully) perceptually encoded. [...] If the subject is to acquire the concept *red*, he or she must not only be shown red things (and, presumably, non-red things), they must be allowed to receive the information that they are red (and not red). The reason for this should be clear: it is the information that the object is red that is needed to shape the internal structure that will eventually qualify as the subject's concept *red*. We need information to manufacture meaning (the concept) because information is required to crystallize a type of structure with the appropriate semantic content.

(DRETSKE 1981 [31] p. 194)

Non seulement la co-variation entre le concept et l'objet permet au sujet de former le concept puis, grâce à ce concept, de discriminer les occurrences futures de l'objet, mais elle explique également le fait que le contenu d'un concept reste stable au cours du temps dans la communauté. Ainsi, dans ce type de théories, le lien causal qui unit le concept à son contenu

est de nature statistique, attribué à une co-variation repérée par les sujets entre certaines de leurs perceptions et certains de leurs états mentaux.

Le principe de co-variation remet en cause l'idée de contenu purement conventionnel. Une signification qui serait un pur produit d'une convention sociale, propagée d'individu en individu, pourrait être totalement arbitraire. Le lien qui unit le concept à son contenu, dans la mesure où il repose sur une co-variation statistique, n'est pas arbitraire. Il prend un caractère naturel. Le concept, parce qu'il co-varie avec les objets perçus, devient porteur d'information. Pour un externaliste, les conséquences semblent impressionnantes. Les concepts, par leur intentionnalité, apportent une information sur le monde. On rejoint ainsi l'idée selon laquelle le langage nous renseigne sur la réalité qui nous entoure. Malheureusement, la connaissance que l'on peut espérer avoir du monde en étudiant les concepts est limitée en raison du manque de robustesse d'un lien intentionnel de nature statistique. Dans la mesure où l'erreur est possible, où la perception d'un chat dans l'obscurité peut conduire à tort à l'évocation du concept CHIEN, l'activation d'un concept n'offre aucune certitude sur l'état du monde. Si de telles erreurs sont fréquentes, on peut arriver à ce que le concept soit évoqué dans des conditions considérées comme impropres du point de vue externaliste : le concept CHIEN co-variera avec un ensemble hétérogène d'entités et non avec la catégorie "chien", considérée comme objective. Il semble n'y avoir aucune garantie théorique sur ce que l'évocation d'un concept peut signifier sur l'état du monde. Cette situation, qui est sans conséquences si l'on se place dans le cadre de l'internalisme méthodologique, semble inacceptable pour les tenants de l'externalisme.

La question de la validité du lien intentionnel, considéré comme informationnel, a conduit bon nombre d'auteurs à rechercher une distinction formelle entre l'évocation "normales" du concept et les conditions d'évocations considérées comme erronées. Ainsi, ce n'est pas parce que, dans des cas de perception anormale, le concept CHIEN viendra à être évoqué que ce concept recouvrera la signification "chien ou grand chat dans le noir". Même si le concept CHIEN co-varie avec des situations du type "grand chat dans le noir", cette co-variation peut être éliminée du lien intentionnel dès lors qu'on est capable de filtrer les situations anormales d'évocation. La question évidente est alors de savoir selon quel critère on peut distinguer les conditions normales des conditions anormales d'évocation. On peut considérer qu'il existe une asymétrie fondamentale qui alloue une prééminence aux conditions normales.

La nature de cette asymétrie est difficile à expliciter, en dehors du constat *a posteriori* que, comme les concepts sont porteurs d'information, il faut bien que les conditions normales d'évocation l'emportent sur les conditions anormales. On peut imaginer, par exemple, en raison de la nature statistique du lien d'évocation, que les conditions anormales sont plus rares. Cette supposition peut reposer sur une présomption d'efficacité du système conceptuel qui est capable de filtrer, dans la plupart des cas, les conditions adéquates d'évocation des concepts. Les concepts sont supposés avoir une fonction cognitive propre. On peut donc s'appuyer sur cette fonction cognitive pour affirmer que, dans des conditions normales de fonctionnement, les concepts ont un lien intentionnel non ambigu. Certains auteurs justifient cette idée de fonctionnement normal non seulement sur le plan cognitif, mais également dans une perspective phylogénétique (MILLIKAN 1984 [73]). Ils en appellent à la sélection naturelle pour fonder l'idée de fonction propre pour un concept.

Si nos concepts existent, c'est parce qu'ils sont efficaces, et ils sont efficaces parce que nous descendons d'individus qui ont été sélectionnés pour les avoir possédés, ou tout du moins pour avoir possédé les moyens de les acquérir en tant que concepts efficaces. Le contenu de tels concepts, forgés par la sélection naturelle, est celui pour lequel ils ont été sélectionnés dans la phylogenèse ou dans l'ontogenèse, non les configurations qui peuvent par moments les évoquer de manière erronée. Il semble que l'on récupère, par le biais de cette

fonction propre, une notion de contenu objectif pour les concepts. Un exemple souvent cité est celui de la perception d'une mouche par la grenouille. L'état neuronal provoqué par la vue de la mouche déclenche une réaction de prédation. Dans certaines conditions, le même comportement de prédation peut être provoqué par un leurre ou par un petit objet inerte emporté par le vent. Si l'on suppose que la réaction de prédation est adaptée pour la fonction biologique de prise d'aliment, on est en droit de dire que la mouche est la cause normale de la réaction de prédation et que le leurre en est une cause anormale. Dans ces conditions, la réaction de prédation peut être considérée comme l'indice de la présence d'une mouche, même dans le cas d'une erreur.

L'intérêt de ce type de raisonnement téléologique pour la compréhension du système conceptuel humain doit être relativisé. Notons tout d'abord que l'appel à la sélection naturelle pour justifier l'efficacité des concepts est critiquable. On peut l'accepter lorsqu'il s'agit d'une caractéristique biologique, quoique dans certaines limites (GOULD & LEWONTIN 1979 [45]). Par exemple, on peut imaginer qu'il existe un petit nombre de concepts primitifs innés qui sont de bons candidats pour être le fruit d'une évolution biologique. Or, cette éventualité est loin d'aller de soi, en particulier parce que nous parlons ici de l'innéité de concepts et non de l'innéité de sensations, comme nous aurons l'occasion de le souligner plus loin (CF. CHAPITRE 8). Lorsqu'il s'agit d'expliquer la genèse de l'ensemble des concepts, considérés comme appris au cours de l'ontogenèse, nous n'avons plus affaire à des caractéristiques biologiques à proprement parler, et l'argument de la sélection naturelle ne porte pas. Il faut donc se rabattre sur l'idée métaphorique d'une sélection ontogénétique analogue à la sélection darwinienne qui se ramène aux mécanismes d'apprentissage qui ont été proposés.

Par ailleurs, si l'appel à la sélection naturelle, dans le cas de l'exemple de la grenouille, apparaît comme justifié, dans le cas des concepts est moins évident. La fonction propre du comportement de prédation observé chez la grenouille ne fait pas mystère. Comment définir une fonction propre pour le concept CHIEN ? Si l'on dit qu'il s'agit de discriminer les chiens et que le concept CHIEN est biologiquement efficace pour cette fonction et aucune autre, le raisonnement est circulaire. On définit la fonction propre du concept CHIEN à partir de la catégorie "chiens", considérée comme une donnée objective, alors que l'objectif était, à l'inverse, d'apprendre quelque chose sur le monde à partir du concept. La seule chose qui peut être dite est donc que le concept doit avoir un contenu dans le monde, qui correspond à ce qu'il discrimine dans les conditions normales, sans qu'il soit possible d'en dire davantage puisque nous sommes dans l'impossibilité de définir ce qu'est le fonctionnement normal d'un concept donné. C'est d'ailleurs cette impossibilité fondamentale de faire une différence *a priori* entre discrimination réussie et discrimination erronée qui rend le projet externaliste si difficile à transposer à la modélisation cognitive.

Une tentative pour restaurer la cohérence de la vision externaliste est de renoncer à la fonder sur le seul critère informationnel. Si l'on reconnaît l'impossibilité, dans la plupart des cas, de caractériser de manière objective une différence entre les causes idéales d'évocation du concept et les causes erronées, il est préférable de poser que le contenu d'un concept est constitué par ce qui est commun à toutes les situations qui le déclenchent (FODOR 1990 [35]). De cette manière, le concept véhicule de l'information sur toutes ces causes. La conséquence, pour un externaliste, est que l'information portée par le concept ne suffit pas, à elle seule, pour définir son contenu. Si le concept CHIEN est déclenché dans certaines conditions par un chat, alors le chat dans ces conditions fait partie de l'ensemble des situations avec lesquelles le concept CHIEN co-varie, sans faire partie de son contenu. Que peut-on dire, alors, sur le contenu externe du concept ? Écrivons que le concept C peut être causé par une propriété P du monde dans les conditions s : $C = f(a_s(P))$. La fonction $a_s(P)$ désigne par exemple l'apparence de P dans les conditions s . Le même concept peut être causé par une propriété P' dans les

conditions $s' : C = f(a_{s'}(P'))$. Si la symétrie entre ces deux formules est parfaite, on ne peut espérer distinguer les conditions normales, représentées ici par s , des conditions anormales désignées ici par s' . Cependant, si l'on peut écrire que $a_{s'}(P) = a_{s'}(P')$ sans que $a_s(P) = a_s(P')$, alors nous tenons une dissymétrie qui permet de réhabiliter la notion de contenu P pour le concept C . Ainsi, nous pouvons affirmer $C = f(a_s(P)) = f(a_{s'}(P'))$, mais $C = f(a_s(P)) \neq f(a_{s'}(P')) : P$ et P' deviennent discernables, en théorie, par rapport à C . L'écriture $a_{s'}(P) = a_{s'}(P')$ peut signifier que P et P' ont la même apparence lorsque la perception est sujette à un certain niveau d'imprécision s' , ou qu'elles évoquent la même image si on tolère un niveau d'association s' . Ainsi, un chat peut passer pour un chien dans le noir, et le lait peut faire penser à une vache si on se laisse aller à faire des associations. L'asymétrie entre P et P' vis-à-vis de C permet de penser que l'une constitue le contenu de C et l'autre non. Si C porte de l'information sur P' , c'est seulement parce que C porte de l'information sur P .

Cows cause "cow" tokens, and (let's suppose) cats cause "cow" tokens. But "cow" means cow and not cat or cow or cat because there being cat-caused "cow" tokens depends on there being cow-caused "cow" tokens, but not the other way around. "Cow" means cow because, [...] non-cow-caused "cow" tokens are asymmetrically dependent upon cow-caused "cow" tokens. "Cow" means cow because but that "cow" tokens carry information about cows, they wouldn't carry information about anything. (FODOR 1990 [35] p. 91)

Un externaliste, selon cette théorie, peut prolonger dans le monde le caractère asymétrique des évocations pour en conclure que certaines situations, parmi celles qui évoquent le concept, en constituent le contenu, alors que d'autres, qui l'évoquent tout autant, en sont exclues. Cette asymétrie ne doit rien au critère informationnel, qu'il soit phylogénétique ou ontogénétique, puisque toutes ces situations co-varient de la même manière avec le concept.

Dans le cadre de l'internalisme méthodologique au sein duquel nous nous situons, ce débat concernant le caractère robuste du contenu conceptuel perd beaucoup de son enjeu. La question devient de savoir si les concepts, considérés ici comme des états mentaux distincts de la perception, renvoient de manière non ambiguë à certains percepts, autrement dit aux percepts qu'ils sont censés normalement discriminer. L'argument téléologique revient alors à fonder cette normalité sur le fait que chez des êtres qui sont le produit de la sélection naturelle, le fonctionnement normal est celui qui a été sélectionné. Comme dans le cas externaliste, cela ne nous renseigne pas beaucoup sur le contenu du concept, car en général nous n'avons pas de moyen indépendant d'estimer l'incidence de telle ou telle discrimination sur la survie des individus qui en sont capables. L'argument de la dépendance asymétrique, quant à lui, se limite à postuler une distinction direct/indirect dans les conditions d'évocation des concepts, sans qu'il soit possible d'en dire plus sans une information indépendante sur les catégories que le concept est censé discriminer.

Notons que pour l'internalisme méthodologique, le lien entre les percepts et les concepts est nécessairement sujet à l'erreur. On ne peut pas appliquer l'espace continu des percepts sur l'espace symbolique des concepts sans qu'il se produise des distorsions. On rejoint ainsi les notions de détection en présence de bruit développées dans la théorie mathématique de l'information. La qualité de la discrimination d'un percept par un concept peut s'estimer par la mesure d'une certaine probabilité d'erreur. On peut même estimer la qualité d'un ensemble de concepts comme sa capacité à produire des discriminations qui minimise la distorsion des distances entre percepts.

Avec cette manière de concevoir l'opération de discrimination conceptuelle, l'internalisme méthodologique permet d'abandonner sans dommage l'objectif qui consiste à vouloir établir une "bijection" entre les concepts et un ensemble prédéfini de contenus

possibles. Comme suggéré plus haut, on peut définir le contenu d'un concept comme ce qui est commun à toutes les situations qui le déclenchent. De cette façon, on conserve l'idée de lien informationnel entre le concept et son contenu, puisque la co-variation est systématique. Cette définition a également le mérite de supprimer l'essentiel de ce qui constituait des "erreurs" dans le cas bijectif. Cet avantage est dû au fait que, comme dans le cas empiriste où les concepts étaient constitués de percepts généralisés, le lien intentionnel permet de respecter la gradualité inhérente à la perception. Lorsque l'on cherche une bijection entre les concepts et les catégories du monde, il faut être capable de donner avec précision la limite de la catégorie "chiens" hors de laquelle le concept CHIEN cesse d'être activé. Une telle gageure n'est pas exigée si le concept, bien qu'étant considéré comme une entité symbolique distincte des percepts, renvoie à un contenu correspondant à un percept moyen.

Un autre avantage de l'abandon du lien bijectif au profit de concepts à contenu moyen est de permettre d'envisager plus simplement un mécanisme d'apprentissage. Dans le cas externaliste, l'origine de l'adéquation parfaite entre les entités mentales que sont les concepts et les catégories du monde reste inexplicée. Comme nous l'avons vu, l'appel à la sélection naturelle ne remplit pas le vide théorique dû à l'absence de critère indépendant pour décider ce qui, dans le monde, mérite, d'un point de vue biologique, d'être isolé comme le contenu d'un concept. La conséquence est un raisonnement circulaire dans lequel ce sont précisément les concepts que nous formons qui sont supposés favoriser la survie. Si, dans le cadre de l'internalisme méthodologique, nous considérons que le contenu des concepts se trouve dans le domaine des expériences, alors l'établissement de ce contenu au cours de l'ontogenèse semble aisé à expliquer. Il suffit d'invoquer un mécanisme statistique d'extraction de régularités. Cette facilité n'est malheureusement qu'apparente, comme nous allons le constater maintenant.

3.6. Le paradoxe de la connexion

Les différentes théories du système conceptuel humain ne sont pas égales devant le problème de la connexion du concept à son contenu. Pour les empiristes, le problème n'existe pas, car les concepts sont de la même nature que les percepts et le lien intentionnel est une identité. Les mécanismes de généralisation ou de construction de bonnes formes produisent des concepts qui, restant dans l'espace des percepts, ne cessent jamais d'être identiques à leur contenu.

Le problème de la connexion se pose de manière extrême, au point de conduire à un paradoxe, dès que les concepts sont considérés comme étant d'une autre nature que les percepts.

[...] there is a further constraint that whatever theory of concepts we settle on should satisfy: it must explain why there is so generally a content relation between the experience that eventuates in concept attainment and the concept that the experience eventuates in attaining. (FODOR 1998 [37] p. 132)

La différence de nature entre les représentations perceptuelles et les représentations conceptuelles rend le problème particulièrement délicat. Les concepts, nous l'avons dit, sont supposés être des représentations discrètes et digitales. Ils n'offrent pas la gradualité qui existe dans l'espace des percepts. Ils ne représentent pas non plus les relations de voisinage qui existent dans le domaine des percepts. Le mécanisme qui les connecte aux percepts ne saurait résulter d'un simple processus d'échantillonnage. Le lien qui semble exister entre

certaines concepts et des prototypes perceptuels, loin d'être une évidence, est précisément ce qui requiert une explication.

[...] what doorknobs have in common qua doorknobs is being the kind of thing that our kind of minds (do or would) lock to from experience with instances of the doorknob stereotype. [...] We have the kinds of minds that often acquire the concept *X* from experiences whose intentional objects are properties belonging to the *X*-stereotype. [...] *Stereotype* is a statistical notion. The only theoretically interesting connection between being a doorknob and satisfying the doorknob stereotype is that, contingently, things that do either often do both.
(FODOR 1998 [37] p. 137)

Le fait de modéliser un lien intentionnel entre deux espaces distincts est déjà problématique en soi, car l'espace conceptuel n'est pas accessible en dehors de ce lien intentionnel. Certes, dans certains modèles, bon nombre de concepts peuvent être accessibles à partir d'autres concepts. Ainsi, si l'on accepte que le concept ÉTALON peut être décrit par une expression comme CHEVAL+MÂLE+ENTIER, on peut espérer définir un mécanisme par lequel le concept ÉTALON hérite son lien intentionnel de ceux des concepts CHEVAL, MÂLE et ENTIER. Toutefois, même dans un schéma définitionnel de ce type, il existe toujours des concepts primitifs qui ne font l'objet d'aucune définition. Comment leur lien intentionnel s'établit-il ? L'expérience produit des percepts, l'exposition au langage permet d'isoler des mots, mais rien, dans l'ontogenèse, ne vient de manière évidente connecter les concepts aux percepts et aux mots qui sont censés les évoquer.

Une solution communément considérée consiste à invoquer l'apprentissage inductif. Ainsi, l'enfant aurait l'occasion de cerner, en généralisant après avoir été exposé à une ou plusieurs occurrences du concept, l'étendue des expériences que le concept est censé reconnaître. Cette solution est incorrecte. Elle présuppose ce qu'elle est censée montrer. Pour reconnaître la première occurrence d'un nouveau concept *P*, l'enfant doit déjà posséder *P* ! Au mieux, le système ne peut que combiner des concepts déjà connus pour produire une formule *G* qui sera équivalente à *P*. Mais un système capable de réaliser une telle opération possède déjà *P* en puissance.

But notice that learning that [*Px is true iff x is G*] could be learning *P* (learning what *P* means) only for an organism that already understands *G*. For, and this point is critical, *G* in [the preceding] formula is *used*, not mentioned. [...] an organism can learn *P* only if it is already able to use at least one predicate that is co-extensive with *P*, viz., *G*.
(FODOR 1975 [34] p. 80)

De fait, les systèmes d'apprentissage proposés en intelligence artificielle, dans la mesure où ils produisent des représentations conceptuelles distinctes des représentations perceptuelles, fonctionnent sur le modèle de l'appariement combinatoire (CF. ANNEXE). L'exposition à des données permet de sélectionner des combinaisons conceptuelles, pas de former des concepts *de novo*. De plus, ces systèmes s'appuient tous sur des concepts primitifs dont le lien avec les données est fourni *a priori*. Il semble donc que le lien intentionnel repose, *in fine*, sur une base innée (FODOR 1975 [34]).

Le problème de la connexion des concepts à leur contenu peut-il ainsi se résoudre par l'hypothèse selon laquelle certains concepts, dits primitifs, disposeraient du privilège d'une intentionnalité innée ? Ce n'est pas certain. Il ne faut pas confondre un concept comme ROUGE, qui est souvent cité pour faire partie du cercle fermé des concepts primitifs, avec la sensation de rouge. Affirmer que le concept ROUGE possède une intentionnalité innée signifie que l'évocation de ce concept est liée, de manière non ambiguë et inamovible, à celle de sa perception. Il devrait être possible de déceler, par des tests psychologiques, la frontière

universelle qui est censée exister entre l'extension perceptuelle d'un tel concept et les percepts qui évoquent d'autres concepts ou qui n'évoquent aucun concept. Rien de tel n'a jamais été montré, et pour aucun concept. Si les sensations connaissent une certaine universalité (CLARK 1993 [20]), il n'existe pas de constante sur les segmentations auxquelles elles donnent lieu (PALMER 1999 [80]). Par ailleurs, invoquer l'innéité des concepts revient à renvoyer la charge d'établissement du lien intentionnel de l'apprentissage à la phylogenèse. Or la sélection naturelle produit des formes localement nécessaires (MONOD 1970 [75]). Il faudrait donc être en mesure d'expliquer que les concepts innés, de par leurs propriétés dans l'espace des concepts, possèdent un lien non arbitraire avec leur contenu, ce qui est loin d'aller de soi (CF. CHAPITRE 8).

Une manière d'échapper aux difficultés liées à l'innéité des concepts primitifs consiste à accepter cette innéité en refusant celle de leur lien intentionnel. L'apprentissage d'un concept pourrait ainsi revenir à une sorte de recrutement d'un concept préexistant pour répondre à un ensemble défini d'expériences. De cette manière, l'important est l'existence d'un mécanisme perceptuel capable de grouper certains percepts, par exemple sous la forme d'un percept moyen, et d'un autre mécanisme capable de connecter ce groupe de perceptions à une représentation de type conceptuel. Ainsi, le concept ROUGE ne se trouverait lié à la perception du rouge que de manière non nécessaire, par le biais d'un mécanisme non inductif de recrutement. Notons que cette solution, si elle permet de résoudre une partie des problèmes évoqués ici, n'en résout pas d'autres, au premier rang desquels celui des liens inter-conceptuels. Nous aurons l'occasion de revenir sur cette question fondamentale de la connexion des concepts à leur contenu (CF. CHAPITRE 8). La solution que nous proposerons consistera à renoncer à l'existence des concepts en tant que représentations mentales permanentes (CF. CHAPITRE 9).

Conclusion

La question fondamentale de l'ancrage des concepts concerne la nature du lien entre les concepts et le monde perçu. Cette question concerne également les théories qui s'intéressent au lien direct que les concepts peuvent avoir avec un monde simplifié considéré comme objectif, puisque la seule connexion qu'un système cognitif peut avoir avec un tel monde s'effectue par le biais de la perception.

La question de l'ancrage se révèle soit triviale, soit extraordinairement ardue. Elle est triviale dans les théories qui ne font pas de différence entre concepts et percepts. Dans ce cas, les concepts sont ancrés de manière naturelle dans la perception. En revanche, dès que les concepts sont considérés comme des représentations distinctes de celles que fournissent nos sens, il devient très difficile d'expliquer comment les concepts se retrouvent liés aux "bons" percepts. Certains auteurs, comme nous l'avons vu, supposent que ce lien est établi par le biais de la communication au sein d'une communauté linguistique, ou du fait d'une disposition cognitive qui nous prédispose à élaborer des représentations conceptuelles portant une information sur le monde. Il n'en reste pas moins que l'amorçage de tels processus reste, dans chaque cas, mystérieux. Comme nous l'avons vu, les "solutions" conventionnelles et informationnelles sont loin de résoudre la question de l'établissement du lien intentionnel. Nous aurons l'occasion de revenir sur cette question fondamentale de l'ancrage (CF. CHAPITRE 8).

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.

Chapitre 5 :
Caractère compositionnel des
concepts

Introduction

Toute langue offre la possibilité à ses locuteurs d'exprimer une variété infinie de sens à l'aide de sons. Chaque mot d'une langue permet d'effectuer un couplage entre une forme sonore et une représentation conceptuelle. La grammaire de la langue permet de prolonger ce couplage de manière à ce qu'un sens puisse être associé à des expressions comportant plusieurs mots. Ainsi, la représentation conceptuelle associée à un énoncé doit être élaborée à partir de la combinaison des concepts associés à ses composants, ce qui confère un caractère compositionnel aux concepts. Or, si la combinaison des mots qui forment une phrase apparaît de manière tangible, il n'en est pas de même de la combinaison des concepts. Ces deux combinaisons s'effectuent-elles en parallèle ? Dans quelle mesure servent-elles de support l'une à l'autre ? Un parallélisme étroit supposerait une correspondance formelle entre la structure grammaticale d'un énoncé et sa structure conceptuelle. Si l'on renonce à une telle hypothèse, il faut imaginer deux mécanismes de combinaison indépendants, l'un grammatical, l'autre conceptuel, produisant des structures de nature différente. Cela suppose l'existence d'une interface définie pour que ces deux mécanismes puissent coopérer et se synchroniser.

Parmi les contraintes qui limitent les possibilités de combiner des mots, bon nombre peuvent être considérées comme indépendantes du système conceptuel. Ainsi, la suite de mots *la brûle écrouler maison* ne sera pas acceptée par un locuteur du français, non parce qu'il est impossible de lui trouver un sens, mais parce qu'elle ne respecte pas certaines règles, de nature syntaxique, qui limitent l'agencement des types de mots. Peut-on imaginer dresser un catalogue de règles, de type grammatical, telles que leur respect garantit l'interprétation de l'énoncé ? Si ces contraintes sont suffisamment strictes tout en restant correctes, on peut imaginer que les énoncés qui respectent ces règles auront, de manière garantie, une interprétation. Certaines approches, notamment en linguistique computationnelle, ne posent pas de différence de nature entre les contraintes de type syntaxique et les contraintes de type sémantique¹. Ainsi, une règle sémantique stipulant que l'agent d'un verbe d'attitude comme *croire* doit être de type *humain* serait de même nature qu'une règle syntaxique d'accord en nombre entre le sujet et le verbe. Une telle perspective suppose que le fonctionnement du système conceptuel est analogue à celui d'une langue. De nombreux auteurs, cependant, refusent une telle perspective. Ils admettent que les mécanismes de composition du système conceptuel et les mécanismes grammaticaux puissent être qualitativement différents. Dans ce cas, il est nécessaire de définir une interface entre la combinaison des mots et celle des concepts.

Dans ce chapitre, nous examinons successivement différentes manières de concevoir le lien entre la construction grammaticale et la composition conceptuelle. Dans une première section, nous présentons la manière dont l'interface syntaxe - sémantique est conçue dans l'approche de la grammaire générative. Cette interface est caractérisée par l'existence d'une "forme logique" qui n'est pas directement observable. Dans une deuxième section, nous présentons une acception différente de la notion de la forme logique, celle de la sémantique formelle. La forme logique y apparaît comme une traduction de la forme linguistique propre à permettre la production d'inférences valides. Dans la troisième section, l'interface syntaxe - sémantique sera décrite comme un appariement précis entre la structure syntaxique et une structure conceptuelle. La quatrième section présente une autre manière de calculer le

¹ L'habitude, dans les modèles basés sur des structures de traits, consiste à regrouper dans un même champ *syn-sem* les traits syntaxiques et sémantiques d'un syntagme, en utilisant le même mécanisme d'unification pour tester les deux types de contraintes.

sens à partir de l'ordonnement des mots, en introduisant des procédures associées aux concepts. La dernière section aborde la construction du sens des énoncés dans les théories de la linguistique cognitive à l'aide des représentations schématiques.

5.1. L'apport de la syntaxe

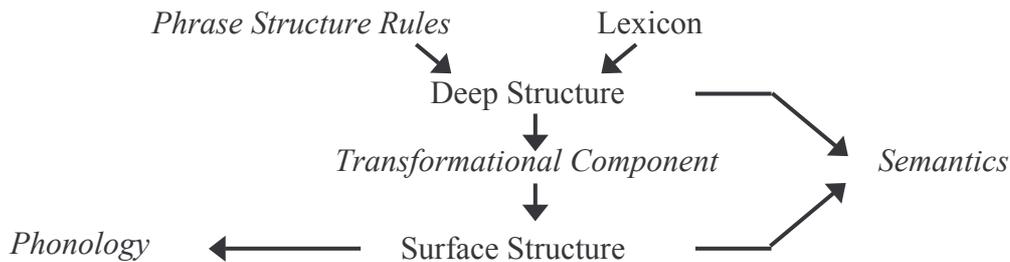
Dans la tradition de la grammaire générative, les mécanismes syntaxiques produisent une structure, appelée forme logique (LF), qui est censée être exploitable par le système conceptuel. La forme logique n'est pas directement observable et ne se confond pas avec la structure grammaticale de l'énoncé tel qu'il est prononcé. Une partie des mécanismes syntaxiques qui engendrent la forme logique est aussi qualifiée de cachée, car ces opérations n'ont pas de conséquences phonologiques (CHOMSKY 1995 [17]). La forme logique est donc l'aboutissement de l'action des mécanismes syntaxiques. Au-delà, d'autres mécanismes, de nature sémantique, sont supposés prendre le relais. Par conséquent, les propriétés prêtées à la forme logique par la théorie syntaxique ont des conséquences sur le fonctionnement du système conceptuel. Celui-ci doit posséder les propriétés suffisantes lui permettant d'intégrer la richesse structurelle de la forme logique.

Considérons par exemple la relation prédicat - argument. La thêta-théorie stipule qu'au niveau de la forme logique, un prédicat doit voir ses arguments satisfaits (CHOMSKY 1981 [16]). Elle proscrit de plus la présence d'arguments surnuméraires. Ces deux contraintes expliquent respectivement le caractère inacceptable de la phrase *lundi, il a dit* et de la phrase *lundi, il a parlé qu'il partait*. Par des règles simples de ce type, la théorie syntaxique guide l'identification des arguments des prédicats. Les marques de cas ou la position relative des syntagmes concourent au même but. Elles permettent de savoir qui est le premier argument du prédicat *frapper* dans chacune des phrases *Paulus Petrum ferit*, *Pierre frappe Paul*, et *Pierre Paul râ mizanad*. Ainsi, la forme logique doit permettre d'exprimer et de distinguer des expressions logiques comme *frapper(x, y)* et *frapper(y, x)*. C'est au niveau de cette forme logique que se révèlent les deux significations d'un syntagme comme le *dépassement de la voiture*, le concept VOITURE occupant la première ou la deuxième place du prédicat *dépasser(x, y)*.

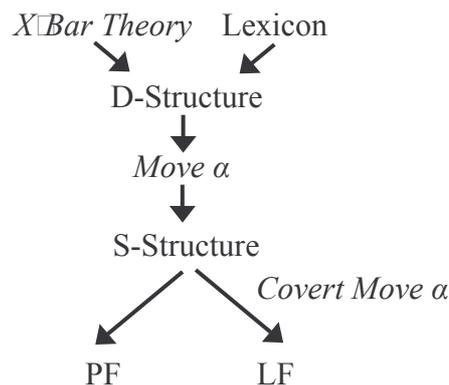
La forme logique est également supposée exprimer la quantification. Par exemple, la phrase *toutes les disquettes ont un virus* possède une structure phonologique non ambiguë qui peut conduire à deux significations. Ces deux significations sont caractérisées, au niveau de la forme logique, par l'ordre dans lequel les quantifications sont positionnées.

La forme logique, définie comme un aboutissement de l'action des mécanismes syntaxiques, nous offre un accès privilégié à la structure du système conceptuel. Il est donc essentiel de savoir en quoi elle consiste et comment elle est engendrée. La réflexion, au sein du courant de la grammaire générative, a été caractérisée par une évolution dans la manière de concevoir l'ordre et la nature des opérations qui engendrent, d'une part la forme logique, et d'autre part la forme phonologique (PF).

La théorie initiale supposait une première phase dans laquelle les mots, guidés par les règles de composition syntaxiques, prennent leur place pour former une première structure, dite profonde, puis une phase de transformations permettant de produire une deuxième structure, dite de surface, qui détermine la forme phonologique de l'énoncé construit. La forme logique, quant à elle, contient des éléments provenant des deux structures profonde et de surface (CHOMSKY 1972 [14]). Le schéma suivant résume cette conception des différents niveaux de composition.



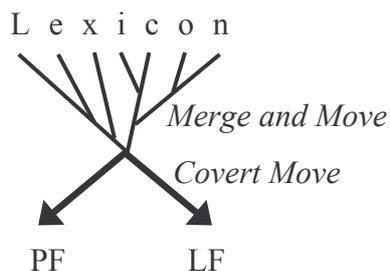
Dans cette perspective, les mécanismes de composition au sein du système conceptuel s’effectuent en deux phases distinctes, en étant soumis, dans chaque phase, aux contraintes fournies par les deux types de structure syntaxique. Cette séparation en deux phases s’est ensuite estompée. Avec l’introduction du mécanisme général du mouvement (*move α*) dans le cadre de la théorie standard de règles syntaxiques (*X \bar{B} ar*), l’idée d’une forme logique unique s’est imposée. Dans cette version de la théorie, la structure syntaxique profonde (*D \bar{S} tructure*) est modifiée par le mécanisme de mouvement pour donner lieu à la structure syntaxique de surface (*S \bar{S} tructure*), qui assure l’interface avec le système conceptuel (CHOMSKY 1981 [16]). La forme logique est le résultat de l’application, cachée au système phonologique, du même mécanisme général de mouvement à la structure syntaxique de surface. Le schéma suivant résume cette conception du processus de la production de la forme logique.



Dans les versions récentes de la théorie, une phrase est un couple (π, λ) associant une forme phonologique et une forme logique. Elle résulte d’une succession d’insertions d’items lexicaux abstraits conduisant à l’élaboration de π et de λ selon des procédures générales de la langue (CHOMSKY 1995 [17]). Conformément aux principes fondamentaux de la grammaire générative, le lien entre forme phonologique et forme logique n’est pas une correspondance, mais résulte d’un ensemble de dérivations. L’élément nouveau est l’accent mis sur le caractère optimal de ces dérivations.

It seems that a linguistic expression of [the generative procedure] cannot be defined just as a pair (π, λ) formed by a convergent derivation. Rather, its derivation must be *optimal*, satisfying certain natural economy conditions: locality of movement, no “superfluous steps” in derivations, and so on. Less economical computations are blocked even if they converge.
(CHOMSKY 1995 [17] p. 220)

Les niveaux successifs de structure syntaxique disparaissent alors au profit d’une construction directe des deux formes phonologique et logique à partir des entités lexicales. Le schéma suivant résume cette conception.



Ainsi, la forme logique est le fruit d'une construction progressive qui obéit à ses propres lois. Les opérations syntaxiques, comme le mouvement ou la fusion (*merge*) ne dépendent pas du sens, même si elles sont en partie au service du sens. L'opération de fusion confère localement au composé des propriétés calculées à partir de ses composants. Elle réalise notamment la liaison sémantique entre les syntagmes. Le principe de liaison sémantique stipule qu'un syntagme doit partager une variable avec un syntagme qui le domine (DESSALLES 2000 [30]). Ainsi, dans le syntagme le dépassement de la voiture, la fusion provoque l'unification de la variable x dans l'expression *voiture(x)* avec l'une des variables de l'expression *dépasser(y, z)*. L'existence d'un choix concernant la variable à unifier montre que les opérations de fusion, à elles seules, sous-déterminent la construction de la forme logique. Cette sous-détermination reste toutefois limitée, car la structure syntaxique impose de nombreuses contraintes structurelles à la forme logique. Prenons l'exemple de la co-référence. Dans une phrase comme *il dit que Jean est malade*, tout locuteur du français sait que le pronom *il* et le nom propre *Jean* ne peuvent pas être co-référents. Cette contrainte n'existe pas dans la phrase *le fait qu'il soit malade ennuie Jean*, où le pronom *il* et le nom propre *Jean* peuvent désigner la même personne. L'explication proposée pour ce phénomène, qui semble universel, est que la co-référence entre un pronom et une expression référentielle est bloquée lorsque le pronom est en position de c-commander² l'expression référentielle. Le phénomène est spectaculaire, car il peut agir à grande distance. Dans la phrase *il avoue que l'ami de Jean à qui Jacques a emprunté la moto avec laquelle il a renversé le livreur était absent au moment des faits*, nous sommes obligés de comprendre que l'auteur de l'aveu ne peut être ni Jean, ni Jacques, ni l'ami de Jean, ni le livreur (en revanche, l'auteur de l'accident et l'auteur de l'aveu peuvent être la même personne). Nous pouvons en conclure que la syntaxe contraint partiellement certains aspects de la forme logique, comme la co-référence, ce qui a pour effet de rendre la communication plus facile.

Il est légitime de se demander où s'arrêtent les spécifications sémantiques qui résultent automatiquement de la syntaxe. Par exemple, les différentes propositions qui ont été faites pour caractériser un système conceptuel qualitativement distinct du langage, mais capable de s'interfacer avec lui, ne peuvent éviter d'y inclure des aspects propres à la syntaxe du langage.

[...] "conceptual structure" contains not only semantic information: it also expresses thematic relations (concepts like Agent, Goal, Source, and the like) and structural relations (that is, hierarchical relations such as "argument of", "head of", "modifier of", and the like). Thematic relations represent, of course, the core materials from which syntax is made. What conceptual structure does not express [...] includes the linear ordering of words into sentences and the morphophonemic shapes that linguistic concepts must take on if they are going to be used communicatively between two individuals. (BICKERTON 1995 [6] p. 114)

² La relation de c-commande est une relation structurelle au sein de l'arbre syntaxique. Le nœud A c-commande le nœud B si le premier nœud ramifié qui domine A domine B, sans que A et B soient eux-mêmes dans une relation de dominance.

Si l'on pousse ce constat jusqu'à sa limite, il n'y a plus aucune raison de supposer l'existence de structures et de mécanismes propres au système conceptuel qui seraient différents de ceux qui nous permettent de parler.

The same biological mechanisms, developed by and for language, run both distinctively human thought and language. (BICKERTON 1995 [6] p. 105)

Si l'on accepte ce point de vue, il n'est plus question d'interface, et les calculs sémantiques doivent être conçus sur le modèle des transformations syntaxiques. Le sens serait ainsi entièrement exprimé par la forme logique. Une telle position nécessite que la forme logique soit engendrée par des transformations qui, tout en restant par nature syntaxiques, dépassent le stade de ce qui est observable. Par exemple, le mécanisme caché de montée de quantification, qui permet d'expliquer le fait que la phrase *toutes les disquettes ont un virus* puisse avoir deux sens logiquement différents, est calqué sur le mouvement des pronoms relatifs³. En extrapolant, on peut imaginer que tous les calculs conceptuels résultent des mécanismes à l'œuvre dans la syntaxe.

Contre ce point de vue, deux arguments peuvent être invoqués. L'un est que les opérations inférentielles qui portent sur les concepts (inférences déductives, généralisations inductives, détections d'incompatibilités logiques, *et cætera*) ne peuvent pas s'exprimer naturellement avec les outils offerts par le système syntaxique. Il est difficile d'envisager un modèle dans lequel les mécanismes et les structures propres à la langue offrent les moyens d'inférer la phrase *Y est mort à partir de la phrase X a tué Y*. Pour traduire les calculs de nature symbolique qui sous-tendent les processus de raisonnement, il a fallu développer des formalismes logiques, qui entretiennent peu de rapports avec les mécanismes grammaticaux des langues. Inversement, les relations syntaxiques sont, pour l'essentiel, insensibles aux dépendances structurelles entre concepts qui ont été postulées dans le précédent chapitre pour expliquer la formation des inférences. Le plus souvent, seule la catégorie syntaxique importe au niveau grammatical, et cette catégorie est en partie décorrélée des propriétés sémantiques du mot. Il est donc peu plausible de postuler d'emblée que les processus par lesquels nous combinons les concepts sont indiscernables des processus grammaticaux.

Certes, la représentation de type syntaxique s'est vue attribuer un rôle nécessaire pour rendre les représentations conceptuelles accessibles aux autres représentations mentales conscientes, notamment aux attitudes propositionnelles (CARRUTHERS 1996 [11]). En d'autres termes, nous penserions à l'aide de phrases. La représentation syntaxique serait donc le substrat sur lequel opéreraient certains processus conceptuels. L'occurrence d'une attitude propositionnelle à propos d'un contenu de pensée nécessiterait l'activation d'une représentation de type linguistique, autrement dit une phrase exprimant ce contenu de pensée. Si tel est le cas, les structures propres au langage naturel, au premier rang desquelles les structures syntaxiques, doivent jouer un rôle fondamental dans la genèse de la pensée en général, et du raisonnement en particulier, bien au-delà du simple décodage des énoncés entendus. Dans l'état actuel des réflexions sur l'organisation du système conceptuel et de ses mécanismes inférentiels, cependant, les opérations qui sont postulées restent constituées de

³ Dans le syntagme *le livre_(i) dont_(i) elle rêvait* $\tau_{(i)}$, la variable représentée par le pronom *dont* est supposée provenir de la position complément du verbe *rêver*, par un mouvement qui laisse subsister une trace $\tau_{(i)}$ co-indexée avec le pronom *dont*. En anglais, le mouvement peut laisser la préposition en place, comme dans le syntagme *the book_(i) she was dreaming of* $\tau_{(i)}$. La présence de mouvements, et des traces associées, est révélée en français par l'accord du participe passé ou par les clitiques pronominaux dans les interrogatives. De même, les deux formes logiques de la phrase *toutes les disquettes ont un virus* s'obtiennent par le mouvement caché, en tête de structure, des deux quantificateurs, donnant soit [*toutes les disquettes₁ [un virus₂ [t₁ infecte t₂]]*], soit [*un virus₂ [toutes les disquettes₁ [t₁ infecte t₂]]*] suivant l'ordre dans lequel les deux quantificateurs ont migré.

schémas de calculs logiques, de type résolution d'incompatibilité, déduction ou autre. Si les structures syntaxiques ont un rôle dans ces opérations, il ne peut être que partiel. En d'autres termes, nous sommes loin d'une réduction du conceptuel au grammatical. Même si l'on accepte l'importance du rôle de la composition syntaxique pour la genèse de la pensée consciente, la composition du sens doit conserver une partie de son autonomie.

L'autre argument contre une telle réduction est lié au fait que le système conceptuel n'est pas soumis à la même contrainte de sérialité que le canal verbal : les différents concepts d'un énoncé sont simultanément activés pour que l'énoncé soit compris. Or, bon nombre de mécanismes syntaxiques tirent leur justification de la sérialité du canal verbal. Prenons le cas du mouvement. Le mouvement du mot *dont* dans la phrase *le livre_(i) dont_(i) elle rêvait t_(i)*, laissant la trace $t_{(i)}$, se justifie par la nécessité d'amener le complément du verbe *rêver* sous la dépendance immédiate du syntagme *le livre*, afin que la liaison sémantique puisse opérer. Ce genre de mouvement ne se justifie pas dans le fonctionnement interne du système conceptuel, dès lors que des structures non linéaires comme des arbres ou des graphes peuvent être postulées. La théorie des cas offre un autre exemple de mécanismes motivés seulement au niveau syntaxique. Dans la phrase *Jean semble malade*, le syntagme *Jean* est sujet du verbe *sembler* sans en être argument au niveau sémantique. L'interprétation classique de ce phénomène est que cette phrase provient de la transformation d'une structure comme *il semble Jean malade*. Cette transformation est rendue nécessaire par le fait que l'adjectif *malade* ne peut pas donner de cas à son sujet⁴. Le nom propre *Jean*, dont le rôle sémantique est donné par l'adjectif *malade*, migre en position de sujet du verbe *sembler* afin de recevoir un cas. Nous avons donc affaire à une mécanique possédant ses lois propres, comme l'obligation pour un syntagme nominal de recevoir un cas, même dans une position dépourvue de rôle sémantique. Cette mécanique casuelle serait sans intérêt dans le fonctionnement interne du système conceptuel, puisque les arguments y sont connus. Comme le mouvement, la contrainte d'attribution de cas aux syntagmes nominaux se justifie au niveau syntaxique, non au niveau sémantique. La raison d'être du système des cas se justifie par la nécessité d'identifier, pour l'interlocuteur, les arguments d'une relation dans une suite linéaire de mots.

On pourrait passer en revue les différents mécanismes syntaxiques pour montrer que leur existence ne se justifie souvent pas sur le plan des opérations purement conceptuelles. Pour ce genre de raisons, bon nombre d'auteurs préfèrent considérer que syntaxe et sémantique correspondent à deux niveaux distincts de la compétence langagière, et que l'on se doit de définir une interface entre les deux systèmes. Les représentations formelles du sens, que nous évoquons dans la prochaine section, ont été conçues au départ comme un système totalement indépendant de l'organisation du langage naturel. Nous verrons que même dans ce cadre formel, la question de l'indépendance entre syntaxe et sémantique peut se poser.

5.2. La représentation formelle du sens

La combinaison grammaticale présente un caractère manifeste, directement observable dans l'ordonnement des mots et les transformations morphologiques. Une structure grammaticale peut être proposée à partir des unités lexicales de l'énoncé, en utilisant les règles de grammaire de la langue. La combinaison conceptuelle ne repose pas sur une base aussi tangible. Comment inférer les propriétés d'un système que l'on ne peut pas observer directement ? On peut penser que l'une des propriétés fondamentales que le système

⁴ La phrase *il croit Jean malade* n'est pas une exception à cette contrainte. Ici, le nom propre *Jean* a migré en position de complément du verbe *croire* où il peut recevoir un cas accusatif, comme le montre la paraphrase suivante : *il le croit malade*.

conceptuel se doit de posséder est la capacité d'exprimer des états de choses de manière non ambiguë, contrairement au langage qui semble bien imparfait avec ses ambiguïtés, son flou et sa sensibilité au contexte.

Une approche fréquemment adoptée consiste ainsi à prendre le problème de la composition sémantique "par le haut". Dans cette approche, un énoncé possède un sens si cet énoncé représente une situation actuelle ou possible, rencontrée ou imaginée, dans le "monde". De plus, on peut dire que le sens d'un énoncé permet, lorsqu'il est compris, d'effectuer un certain nombre d'inférences. En d'autres termes, les représentations sémantiques attribuées aux énoncés du langage sont supposées s'interpréter dans le "monde" et servir aux raisonnements qui peuvent être effectués à propos du "monde". Or, la réflexion philosophique et scientifique a produit un système de représentations dont c'est précisément la fonction : le formalisme logique.

Le formalisme logique est, en lui-même, un langage, pour lequel on peut distinguer une syntaxe et une sémantique. Dans ce langage, la valeur sémantique d'un énoncé s'obtient de manière récursive à partir de l'interprétation des expressions constituant l'énoncé. Cette évaluation s'effectue par rapport à un domaine d'interprétation, en deux temps. Dans un premier temps, les symboles non logiques reçoivent une interprétation dans le domaine. Par exemple, dans la logique du premier ordre, une constante du langage est associée à un élément précis du domaine, et un symbole de prédicat est associé à une relation définie entre certains éléments du domaine. Dans un deuxième temps, l'évaluation consiste à calculer la valeur sémantique des expressions interprétées⁵, puis à propager ces valeurs à tout l'énoncé en tenant compte des opérateurs logiques⁶. Ce type d'évaluation directe est souvent fastidieux. La raison d'être de la logique est de permettre d'éviter l'évaluation directe des énoncés, grâce à des mécanismes de dérivation qui conduisent à déduire qu'un énoncé est satisfait dans un domaine (ou peut être satisfait pour certaines interprétations de ses variables) dès que certains autres énoncés le sont. Ainsi, la règle du *modus ponens* permet de déduire la satisfaction d'énoncés partiels à partir de celle de l'énoncé englobant⁷. Certains formalismes peuvent être dotés de mécanismes de dérivation⁸ corrects et complets qui permettent de produire la liste de tous les énoncés dont la satisfaction peut être déduite de celle d'un énoncé donné.

Les qualités des formalismes logiques suggèrent une approche de la sémantique du langage naturel par laquelle on cherche à éliminer l'ambiguïté, le flou et le caractère contextuel des énoncés langagiers. Cette approche, qui est celle de la philosophie du langage, consiste à transcrire ce qui est exprimé par le langage dans un autre langage, celui de la logique. C'est au moment de cette transcription que les approximations du langage naturel sont levées. Par l'emploi d'un langage formel idéalisé, on parvient à caractériser de manière parfaite et non ambiguë les propriétés intentionnelles et inférentielles de l'énoncé de départ. En d'autres termes, un énoncé du langage naturel aurait un sens si et seulement si on peut lui associer une représentation logique et que cette représentation logique peut être interprétée dans un domaine donné. De même, une chaîne de raisonnement à partir d'un énoncé du langage naturel est fiable si et seulement si sa transcription en langage formel procède selon les étapes d'une dérivation logiquement bien formée.

⁵ Par exemple, une expression comme $p(a, b, c)$ sera évaluée comme satisfaite si le triplet d'éléments du domaine associé à la configuration de symboles (a, b, c) se retrouve dans le graphe de la relation ternaire associée au prédicat p .

⁶ Par exemple, une expression comme $p(a, b, c) \wedge q(d)$ sera évalué comme satisfaite si les deux expressions $p(a, b, c)$ et $q(d)$ sont, chacune, évaluées comme satisfaites.

⁷ Le *modus ponens* est la règle déductive qui conclut à la satisfaction d'une expression q à partir de celles des expressions p et $(p \supset q)$.

⁸ Ces mécanismes incluent les systèmes axiomatiques et les méthodes de preuve, par exemple les méthodes de preuve par réfutation.

Le fait d'adopter une représentation logique permet de franchir une étape supplémentaire, en considérant qu'un énoncé a un sens s'il peut être évalué. Comme, par hypothèse, un énoncé sensé du langage naturel peut être transcrit dans une expression logique bien formée, et que toute expression logique bien formée peut être évalué par rapport à un domaine, un énoncé sensé du langage naturel doit pouvoir recevoir une valeur de vérité⁹. Plus exactement, si l'énoncé comporte des variables, son sens peut être assimilé à ses conditions de vérité. Selon cette perspective, le sens d'une phrase devient l'ensemble des états possibles du monde dans lesquels la phrase est vraie. Ainsi, le sens de la phrase la voiture de Jean est grise est l'ensemble des états possibles du monde dans lesquels l'individu indiqué par le prénom Jean possède effectivement une voiture grise. On parvient ainsi à donner un critère d'acceptabilité sémantique permettant de distinguer les énoncés sensés du langage, en transformant la question du sens en une question épistémique.

L'idée fondamentale de la sémantique formelle consiste à remplacer la forme imparfaite de l'énoncé du langage naturel par sa transcription dans un langage formel. La forme logique ainsi obtenue a un statut radicalement différent de celle des approches générativistes du langage. Ici, la structure de la forme logique n'est pas nécessairement parallèle à la structure syntaxique. En particulier, la constituance syntaxique peut être perdue. Par exemple, si l'on traduit l'énoncé Jean casse un verre, dans un système de logique du premier ordre, par l'expression $(\exists x) \text{ verre}(x) \ \& \ \text{casser}(\text{Jean}, x)$, on constate que le prédicat *verre* n'est pas un constituant du prédicat *casser*, alors que le nom *verre* est un constituant du syntagme verbal dont la tête est le verbe *casser*. Par ailleurs, les symboles et les structures utilisés dans la forme logique ne correspondent pas nécessairement à des entités psychologiques. Ce sont, avant tout, les éléments d'un système de représentation. Dans la section précédente, la forme logique possédait un statut cognitif, dans la mesure elle était issue d'une forme syntaxique profonde par une suite supposée d'opérations mentales. Elle était censée conserver tous les aspects de la structure syntaxique utiles à l'interprétation sémantique. La forme logique considérée dans la présente section résulte d'une transcription sans prétention cognitive. L'objectif est de représenter le contenu de l'énoncé du langage naturel dans un langage artificiel où les ambiguïtés, le flou et les aspects contextuels peuvent être éliminés. De cette manière, la transcription formelle est un moyen de révéler le sens. La diversité des formalismes logiques a pour conséquence qu'un énoncé donné pourra être associé à plusieurs représentations formelles. Il serait inutile de dire que l'une de ces représentations est plus "cognitive" que les autres.

Si l'on prend à la lettre cette restriction sur le caractère non cognitif de la sémantique formelle, on parvient à un certain scepticisme. La forme logique ne serait qu'un outil commode pour exprimer le sens de manière précise en s'affranchissant des imperfections du langage naturel. Elle ne serait d'aucun secours pour modéliser les mécanismes sémantiques à l'œuvre dans le langage naturel. Ce point de vue est, d'ailleurs, celui de la tradition philosophique qui est à l'origine de la sémantique formelle. La forme logique est incontournable pour comprendre la proposition contenue dans l'énoncé langagier et l'état de fait qu'elle exprime, ce qui n'est pas le cas de la forme grammaticale. C'est donc la forme logique qui doit être retenue pour comprendre la rationalité de la pensée scientifique. Une telle attitude conduit naturellement à accepter un divorce entre la forme grammaticale et la forme logique, et à adopter une vision pessimiste concernant l'espoir de modéliser fidèlement toutes les nuances de sens exprimées par les mots et les grammaires des langues naturelles. Pour certains auteurs, cependant, le défi mérite d'être relevé, car l'adéquation de la forme logique au langage ne peut qu'améliorer son pouvoir d'exprimer la pensée humaine.

⁹ Dans cette tradition philosophique, le mot "vérité" est employé dans un sens objectif. Il n'a donc pas le statut d'attitude propositionnelle qu'on lui donne dans certaines approches cognitives du langage, ni le statut de vérification opérationnelle qu'il peut avoir en intelligence artificielle.

I reject the contention that an important theoretical difference exists between formal and natural languages. (MONTAGUE 1974 [76] p. 188)

Le projet visant à adopter une vision formelle du langage doit d'emblée aborder deux problèmes. Le premier est de concevoir une interprétation qui permette de reproduire la richesse de la sémantique du langage naturel. Le deuxième est d'organiser une correspondance systématique entre la forme grammaticale et la forme logique.

Interprétation

La première difficulté à résoudre pour rapprocher la forme logique du langage naturel réside dans le fait que l'interprétation sémantique est supposée objective, indépendante du point de vue du locuteur, ce qui la rend souvent trop rigide. Le système logique de base utilisé pour traduire les énoncés du langage naturel est la logique du premier ordre. Dans ce système, les phrases Socrate est mortel et tout homme est mortel peuvent être traduites, respectivement, par des formules comme $M(s)$ et $\forall x H(x) \supset M(x)$. Leur interprétation peut s'effectuer dans le domaine des entités "réelles". Dans ce cadre, la constante s est interprétée comme un individu particulier, connu sous le nom de Socrate. Les symboles de prédicat H et M correspondent à des ensembles d'individus, l'ensemble des humains et l'ensemble des mortels. L'évaluation du premier énoncé s'opère par un test d'appartenance, celle du deuxième par un test d'inclusion entre les deux ensembles concernés. Ces procédures sémantiques semblent opérer correctement dans un monde supposé comme objectif. Or, le langage peut exprimer des états de faits qui, bien qu'intuitivement compréhensibles, ne semblent pas entrer dans le schéma interprétatif esquissé ci-dessus. Ainsi, l'énoncé Pégase est blanc pose des problèmes, dans la mesure où le mot Pégase ne renvoie pas à une entité du monde "réel". Le domaine d'interprétation constitué par les objets "réels" ne permet pas de donner un sens à cet énoncé. L'interprétation de cet énoncé est possible à condition de pouvoir considérer un deuxième domaine d'interprétation inhabituel, incluant les entités mythologiques. De cette manière, la phrase Pégase n'existe pas sera interprétée par le fait qu'il existe un domaine, celui des entités "réelles" où aucun élément n'est associé au mot Pégase. Ainsi, en considérant plusieurs domaines d'interprétation, on parvient à donner un sens à une classe beaucoup plus large d'énoncés langagiers.

La rigidité de l'interprétation logique la rend impropre à exprimer la combinaison des significations véhiculées par le langage. Par exemple, lorsque l'adjectif ancien modifie un autre mot, le résultat ne peut pas être représenté par une conjonction logique. Le sens du syntagme ancien soldat n'est pas rendu par la formule $\text{ancien}(x) \ \& \ \text{soldat}(x)$. Un ancien soldat n'est pas un objet appartenant à l'intersection de l'ensemble des objets anciens et des soldats. Ce problème peut néanmoins recevoir une solution. Il suffit d'accepter, non seulement un changement de point de référence, comme dans l'exemple du mot Pégase, mais la possibilité de considérer tout un ensemble de domaines, chacun constituant un point de référence daté. Ainsi, on recherchera, dans l'ensemble des domaines d'interprétation du passé, ceux qui incluent un individu ayant la propriété d'être soldat, de manière à ce que le syntagme ancien soldat soit avéré dans le domaine correspondant au présent. Le fait de pouvoir adopter plusieurs points de référence augmente le pouvoir d'expression de la logique. La technique consiste à enrichir la structure d'interprétation. Au lieu de s'en tenir à un seul domaine, on considère une famille de domaines, dont chacun constitue un point de référence.

To interpret a pragmatic language L we must specify several things. In the first place, we must determine the set of all [...] complexes of relevant aspects of possible contexts of use; we may call such complexes [...] *points of reference*. [...] In the second place, we should have to specify, for each point of reference i , the set A_i of objects present or existing with respect to i . [...] In the third place, we should have to specify the meaning or *intension* of each predicate and individual constant of L . To do this for a constant c , we should have to determine, for each point of reference i , the denotation or extension of c with respect to i .

(MONTAGUE 1974 [76] p. 121)

Pour augmenter les possibilités de transcription du langage naturel en langage formel, la stratégie consiste, non à complexifier le langage formel, mais à augmenter ses possibilités d'interprétation. La conséquence est qu'un terme n'a plus d'extension en absolu. Son extension dépend du point de référence. L'analyse sémantique classique était cantonnée aux phrases ayant une structure de prédication universelle et atemporelle. Dans le cadre de la sémantique intensionnelle, l'interprétation des expressions logiques ne consiste plus à associer chaque symbole à son extension, puisque l'extension dépend du point de référence. L'interprétation passe par une fonction, appelée intension, qui doit être calculée pour déterminer l'extension. Par exemple, la prise en compte du temps dans la transcription peut être effectuée sans introduire de paramètre temporel explicite dans l'expression logique. Un énoncé comme *Jean est parti* continue à être traduit par un prédicat atemporel *partir(jean)*. Le renvoi de l'événement dans le passé est rendu au niveau de l'interprétation. Pour cela, on considère une famille de domaines indicés par le temps. La modalité temporelle de l'énoncé se traduit par le fait que le prédicat *partir(jean)* était avéré dans l'un de ces mondes passés. Toutefois, le sens de l'énoncé *Jean est parti* ne se limite pas au fait qu'il existe un monde dans lequel le départ de Jean est avéré. La phrase exprime une antériorité par rapport à l'instant d'énonciation. De telles contraintes peuvent facilement être traduites par l'emploi d'opérateurs modaux. Dans le cadre d'une logique temporelle, l'opérateur P , dans l'expression P *partir(jean)*, contraint l'évaluation à rechercher un monde passé dans lequel le prédicat s'avère.

The fourth thing we must provide is an interpretation of operators of L . To do this we associate with each operator of L a relation between points of reference and sets of points of reference.

(MONTAGUE 1974 [76] p. 122)

Dans le cas de la logique temporelle, l'opérateur P associe au monde courant l'ensemble des mondes qui lui sont antérieurs. L'emploi d'opérateurs modaux agissant sur la structure d'interprétation permet de faire figurer des contraintes sémantiques dans la syntaxe du langage logique. Cette méthode augmente considérablement la capacité du langage logique à représenter le sens des énoncés langagiers. Les opérateurs modaux associés à des mondes censés représenter les croyances de divers individus permettent de créer des logiques épistémiques capables de différencier le sens d'un énoncé en tenant compte des locuteurs. Là où la logique classique a accès à une capacité de vérification omnisciente pour évaluer les énoncés, les logiques épistémiques peuvent représenter toute la variété des croyances partielles.

De telles techniques permettent d'aborder le problème du caractère contingent de certaines connaissances nécessaires à la compréhension des énoncés langagiers. Il est habituel, en linguistique et en philosophie du langage, de distinguer les aspects qui dépendent du contexte de l'énonciation et ceux qui ne dépendent que d'une connaissance générale de la

langue et des concepts. La prise en compte du contexte est parfois appelée pragmatique¹⁰. Elle est nécessaire dès que l'on est en présence, par exemple, de termes indexicaux comme les démonstratifs.

It seemed to me desirable that pragmatics should at least initially follow the lead of semantics – or its modern version, model theory – which is primarily concerned with the notions of truth and satisfaction (in a model, or under an interpretation). Pragmatics, then, should employ similar notions, though here we should speak about truth and satisfaction with respect not only to an interpretation but also to a context of use.

(MONTAGUE 1974 [76] p. 120)

Une idée souvent admise est qu'il est possible d'opérer une distinction de principe permettant d'isoler ce qui, dans l'ensemble des connaissances utiles à l'interprétation d'un énoncé donné, dépend du contexte dans lequel l'énoncé est émis.

The interpretation of an utterance depends on the meaning of the expressions used and on various additional facts about the utterance. The truth of an utterance depends, in turn, on whether its interpretation fits the facts. [...] Using situations, one can separate cleanly the two parts of reality that intervene between meaning and truth, the facts about the utterance, and the facts about the *described situation*, i.e., those relevant to the truth and falsity of the utterance as interpreted.

(BARWISE & PERRY 1983 [4] p. 6)

L'interprétation d'un énoncé est ainsi censée se dérouler en deux temps. Dans un premier temps, le contexte de l'énonciation est élaboré en précisant la location spatio-temporelle et les entités auxquelles certaines unités lexicales comme les pronoms ou les démonstratifs font référence. Ensuite intervient la traduction des termes génériques.

Consider the sentence I AM SITTING. Its meaning is, roughly, a relation that holds between an utterance u and a situation e [...] expressed as follows:

u [[I AM SITTING]] e
iff

There is a location l and an individual a , such that
in u : at l : speaks, a ; yes
in e : at l : sits, a ; yes

[...] Now suppose we fix certain facts about the utterance. Fred is the speaker, and the location is a particular spacetime region l . Then *any* situation that has Fred sitting at l will be an interpretation of the utterance.

(BARWISE & PERRY 1983 [4] p. 19)

Dans cet exemple, le sens de l'énoncé est éclaté dans les différents éléments de la traduction. Le pronom "I" est interprété par l'unification de la variable a dans deux situations, caractérisées par les expressions "speaks" et "sits". Le fait que le temps de l'énoncé soit au présent a pour conséquence l'unification des localisations sous le même nom de variable l . Cette variable peut également contribuer à faire le lien entre cet énoncé et le discours dans lequel il se trouve.

Pour de nombreux auteurs, l'amélioration du pouvoir d'interprétation de la logique pourra permettre, à terme, de représenter le sens des énoncés langagiers de manière complète et non ambiguë, en tenant compte des conditions d'énonciation. Ce programme de recherche a produit des avancées indéniables dans l'élaboration d'un langage logique aux capacités de

¹⁰ Dans une autre tradition, issue des travaux de H. P. Grice, le terme pragmatique désigne un authentique niveau de la compétence langagière, celui des inférences et de la prise de parole pertinente.

représentation étendues. Cependant, les moyens mis en œuvre pour représenter le sens peuvent être critiqués sur le plan épistémologique et cognitif.

La première critique concerne le statut des objets utilisés dans la traduction logique. Comme dans le cas des formes profondes utilisées par les syntacticiens, certains éléments n'ont pas de contrepartie explicite dans l'énoncé. La variable *l* de l'exemple précédent n'a aucune réalité phonologique. Cependant, contrairement à l'approche évoquée dans la première section, le souci de rendre la forme logique minimale par rapport à ce qui est exprimé verbalement n'est pas central. La préoccupation est plutôt de rendre le pouvoir d'expression de la logique maximal. D'où l'introduction de variables, d'opérateurs et de structures d'interprétation complexes comprenant souvent une infinité de mondes possibles. Comme l'expression langagière sous-détermine les structures postulées par la sémantique formelle, le statut cognitif de ces structures n'est pas clair. Dans l'exemple précédent, la séparation en deux situations reliées par un partage de variable est due aux choix du formalisme. Ce qui est observable dans l'énoncé *je suis assis* n'offre pas d'indication claire selon laquelle l'auditeur d'un tel énoncé conçoit l'état décrit et l'acte d'énonciation comme deux situations traitées au même niveau. En d'autres termes, les modèles de la sémantique formelle sont conçus pour leur efficacité, non pour représenter les processus cognitifs à l'œuvre.

Une deuxième critique porte sur la possibilité, énoncée par de nombreux auteurs, d'isoler dans l'interprétation sémantique des énoncés ce qui relève du contexte immédiat de ce qui n'en dépend pas. Certaines caractéristiques liées à l'interprétation sont clairement contextuelles. C'est le cas de la localisation, de l'instant d'énonciation, de la référence des pronoms et des déictiques. Peut-on assurer que l'interprétation de certains mots, au moins, peut se faire indépendamment du contexte ? Dans l'expression *un grand vin*, l'adjectif *grand* modifie une dimension de qualité, alors que dans l'expression *un grand arbre*, la qualification opère sur la taille. L'action sémantique de l'adjectif *grand* dépend donc du contexte des autres mots de la phrase. Le mot *arbre*, prononcé par un enfant ou un botaniste expert, en Amazonie ou en Islande, recouvrera des objets différents. Il est difficile de trouver un élément lexical exempt de polysémie. Un mot comme *livre* peut désigner l'objet physique, l'œuvre intellectuelle, l'opération commerciale, l'ensemble des exemplaires, *et cætera*. La polysémie concerne également les mots grammaticaux. Dans les expressions *un livre sur une table*, *un tableau sur le mur*, *un insecte sur le plafond*, la relation spatiale désignée par le mot *sur* s'obtient par un calcul qui dépend des entités mises en relation. En résumé, il semble que le contexte sémantique puisse intervenir dans le sens de tous les éléments lexicaux. Aucun principe formel permettant de distinguer des cas d'immunité au contexte n'a été proposé, hormis ceux qui peuvent être déduits de la structure syntaxique, comme les contraintes de co-référence mentionnées dans la section précédente. Si le contexte est toujours susceptible d'intervenir, quelle sont les lois qui permettent de prédire son influence ? En l'absence d'influence du contexte, on peut espérer calculer le sens des énoncés en utilisant un dictionnaire élaboré contenant des schémas sémantiques pour chaque entrée lexicale. L'omniprésence du contexte nécessite des calculs additionnels pour lesquels la sémantique formelle n'offre pas de description générale et systématique. Il est toujours possible, au cas par cas, de postuler l'existence d'un point de référence ou d'un trait permettant de lever telle ou telle ambiguïté. Certes, la prise en compte du temps, de la localisation et du locuteur dans la structure d'interprétation contribuent significativement à préciser les situations dans lesquelles un énoncé donné est interprétable. Cependant, on ne peut exclure *a priori* aucun paramètre comme non pertinent pour l'interprétation. Ainsi, la température ambiante ou le degré d'hygrométrie peuvent modifier l'ensemble des situations auxquelles un énoncé s'applique (par exemple l'énoncé *l'eau est froide* peut désigner des situations différentes en été et en hiver). Faut-il prévoir une modalité de température, et donc multiplier l'univers des mondes

possibles, ou des situations, pour distinguer tous les cas qui ne diffèrent que par la température ? Le danger est de voir l'ensemble des dimensions concevables par l'esprit humain venir encombrer la structure d'interprétation.

Une troisième critique concerne précisément la structure d'interprétation. Le principe de la théorie des modèles consiste à faire reposer le sens sur un jugement épistémique objectif. Ainsi, le sens de l'énoncé Jean est assis est constitué par l'ensemble des situations, choisies dans les mondes accessibles de la structure d'interprétation, qui rendent l'expression *assis*(Jean) vraie. Si l'on renonce à la doctrine platonicienne intégrale, selon laquelle de telles relations seraient extérieures aux esprits tout en leur étant accessibles¹¹, le problème vient de ce que l'on renvoie toute la richesse sémantique imaginée par les sémanticiens formels dans les capacités perceptives des individus. Autrement dit, si l'on cherche à transformer les théories formelles de la sémantique en théories cognitives, les capacités perceptives humaines deviennent d'une complexité peu plausible. Pour comprendre le moindre énoncé, l'auditeur devrait effectuer des calculs sur des ensembles infinis de mondes, de manière à dresser la liste exhaustive des situations dans lesquelles l'énoncé est avéré. Il faut, de plus, que l'auditeur dispose d'une panoplie de relations non ambiguës portant sur ces mondes. Si l'on accepte une telle hypothèse, il est légitime de s'interroger sur les raisons qui font que les êtres humains utilisent un langage si ambigu et polysémique, sachant qu'ils ont à leur disposition un catalogue de relations parfaites¹².

Une théorie normative du sens, qui viserait à déterminer le sens des énoncés du langage de manière objective, indépendamment des locuteurs, pourrait en principe se satisfaire d'un système d'interprétation en forme d'oracle, quelle qu'en soit la complexité. Une telle théorie aura peu à nous apprendre sur le comportement langagier humain et sera incapable de le prédire, car l'oracle censé dispenser la vérité est inaccessible au scientifique et encore plus aux locuteurs. La situation est différente lorsque l'on utilise les formalismes logiques pour modéliser le raisonnement. Dans ce cas, on obtient de véritables mécanismes auxquels on peut assigner une contrepartie cognitive. Par exemple, il n'est pas choquant d'imaginer un mécanisme cognitif capable de transformer une incompatibilité en disjonction. En revanche, il est difficile d'imaginer un mécanisme cognitif dont la fonction serait de produire des conditions de vérité après avoir parcouru une infinité de domaines constitués d'entités et de relations qui, pour la plupart, sont extérieures au sujet et indépendantes de son fonctionnement cognitif.

Correspondance

Les critiques précédentes, qui portaient sur la nature de l'interprétation, ne sont qu'un aspect des difficultés présentées par le projet de la sémantique formelle. Une autre source de complications est liée au fait que l'on postule l'existence de deux formes concurrentes d'expression. Si une même situation, observée dans le monde, peut s'exprimer spontanément par le langage naturel et scientifiquement sous la forme d'un langage formel, il faut préciser quel lien existe entre ces deux formes d'expression. Or, dans la plupart des formalismes, ce lien n'est pas étroit. Par exemple, dans la traduction $\forall x \text{ homme}(x) \supset \text{mortel}(x)$ de la

¹¹ Une telle doctrine est incompatible avec une approche matérialiste des problèmes cognitifs. L'idée de confier à la science, par exemple aux sciences physiques, la charge de vérifier la vérité des énoncés est illusoire pour deux raisons. D'une part, la connaissance scientifique est toujours provisoire, ce qui rend la notion de vérité également provisoire. D'autre part, la plupart de ces tests de validité scientifique sont inaccessibles aux locuteurs qui, pourtant, parviennent à communiquer.

¹² L'idée qui consiste à introduire des logiques floues ou multi-valuées ne résout pas le problème. La difficulté signalée ici réside, non dans le caractère graduel et flou des catégories désignées par les mots du langage, mais dans l'ambiguïté de ces mots. Or, l'interprétation, dans ces logiques, sans être binaire, doit néanmoins être non ambiguë.

phrase tout homme est mortel, le groupe nominal tout homme se retrouve éclaté entre le quantificateur $\forall x$ et le prédicat $\text{homme}(x)$. Inversement, le prédicat $\text{mortel}(x)$ se retrouve sous la portée du quantificateur, au même niveau que le prédicat $\text{homme}(x)$, alors que l'adjectif mortel est extérieur au syntagme nominal tout homme dans la phrase. Le passage de la représentation grammaticale à la représentation formelle apparaît ainsi davantage comme une adaptation que comme une correspondance. Or, un calcul formel du sens est censé expliquer de manière exhaustive la transformation qui permet de passer de l'expression grammaticale à l'expression formelle. Si la traduction du langage sous une forme logique donne lieu à interprétation, alors l'adoption d'un système formel peut certes aider à la détermination des valeurs de vérité et à la levée des ambiguïtés, mais elle ne peut pas prétendre reproduire le calcul du sens.

Une manière de s'affranchir de cette difficulté est de postuler une correspondance stricte entre la forme grammaticale de l'énoncé et l'expression formelle. Il s'agit là d'une exigence forte, qui rend le choix du système formel particulièrement crucial. Dans ce cas, la compositionnalité devient une préoccupation centrale, puisque la structure des expressions du langage formel doivent être déduites de la syntaxe du langage naturel. Avant de profiter de la précision que permet le langage formel, il faut passer par une phase de transformation algorithmique qui produit l'expression formelle à partir de l'expression verbale. Une manière simple de formaliser cette idée est d'imaginer l'existence d'une correspondance de structures : la syntaxe présente une structure algébrique, la sémantique se caractérise, elle aussi, par une deuxième structure algébrique, et il existe un homomorphisme¹³ qui projette les éléments de la première dans la deuxième (MONTAGUE 1974 [76]).

Le mérite de cette idée est d'astreindre le système sémantique à respecter la compositionnalité grammaticale. Techniquement, on part d'un lexique organisé selon les catégories grammaticales (nom commun, verbe intransitif, adverbe, *et cætera*). On suppose donnée l'interprétation sémantique de chaque terme de ce lexique dans la structure d'interprétation considérée. Une première étape consiste à trouver des règles de compositions syntaxiques qui, à partir du lexique, engendrent de manière récursive des expressions syntaxiquement bien formées (syntagmes nominaux, phrases, syntagmes adverbiaux, *et cætera*). Une deuxième étape consiste à appliquer un ensemble d'opérations de composition sémantique qui engendrent l'interprétation d'une structure syntaxique en fonction de l'interprétation de ses composants. C'est à ce point que s'opère la mise en œuvre de l'homomorphisme. À l'application d'une règle syntaxique S_i à l'ensemble d'expressions langagières $\{\delta_j\}$ correspond, sur le plan sémantique, l'application d'une opération F_i à l'ensemble de significations $\{g_j\}$.

Corresponding to the syntactic rules S_2, \dots, S_{17} (and to given sets A and I) we introduce semantic operations $F_{2,IA}, \dots, F_{17,IA}$ with the following intuitive significance: if the expression μ is « obtained » by the n^{th} syntactic rule from the expressions $\delta_1, \dots, \delta_k [\dots]$ and if g_1, \dots, g_k are possible denotation functions of $\delta_1, \dots, \delta_k$ respectively, then $F_{n,IA}(g_1, \dots, g_k)$ is to be the corresponding denotation function of μ .

(MONTAGUE 1974 [76] p. 201)

¹³ En mathématique, on étudie les fonctions établissant des liens entre ensembles munis de structures analogues. Ces fonctions appliquent la structure d'un ensemble sur celle de l'autre. Dans le cas des structures algébriques, on peut qualifier certaines de ces fonctions d'homomorphisme. Par exemple, étant donnés deux groupes $(G, *)$ et (H, \bullet) , l'application $\varphi : G \rightarrow H$ est un homomorphisme si pour tout $g_1, g_2 \in G$ on obtient $\varphi(g_1 * g_2) = \varphi(g_1) \bullet \varphi(g_2)$. Deux ensembles structurés sont considérés comme algébriquement équivalents lorsqu'il existe un homomorphisme bijectif entre les deux. Dans ce cas, ils sont dits isomorphes.

Les lettres I et A représentent respectivement l'ensemble des mondes possibles et le domaine donnés par le modèle. La compositionnalité est obtenue par le fait que la composition sémantique qui conduit à l'interprétation de l'expression μ est entièrement calculable à partir de l'agencement de ses composants syntaxiques et de leur signification. Ainsi, la structure syntaxique de la phrase *a woman loves every man* est construite, dans la description qui suit, par l'application successive des règles S3, S10, et S9¹⁴.

[a woman loves every man]			(S9)
[a woman loves v_1]	v_1	[man]	(S10)
[v_0 loves v_1]	v_0	[woman]	(S3)
[loves]	v_0	v_1	

Les règles S3, S10 et S9 ont respectivement pour rôle la construction d'un syntagme verbal à partir d'un verbe transitif, l'introduction d'un syntagme nominal déterminé par l'article *a*, et l'introduction d'un syntagme nominal déterminé par le mot *every*. L'interprétation de cette construction à l'aide des opérations sémantiques F_3 , F_{10} et F_9 s'effectue de manière suivante.

$$F_9(F_{10}(F_3(G([\text{loves}]), G(v_0), G(v_1)), G(v_0), G([\text{woman}]), G(v_1), G([\text{man}])))) \quad (F_9)$$

$$F_{10}(F_3(G([\text{loves}]), G(v_0), G(v_1)), G(v_0), G([\text{woman}])) \quad (F_{10})$$

$$F_3(G([\text{loves}]), G(v_0), G(v_1)) \quad (F_3)$$

La fonction G donne l'interprétation d'une entité lexicale ou d'une variable. Les fonctions F_3 , F_{10} et F_9 représentent respectivement l'interprétation d'une relation binaire, d'une quantification existentielle et d'une quantification universelle. Le résultat est équivalent à la formule de la logique du premier ordre $\forall x(\text{man}(x) \supset \exists y(\text{woman}(y) \wedge \text{love}(y, x)))$. Le point important est que l'interprétation alternative de l'énoncé, représentée par la formule $\exists x(\text{woman}(x) \wedge (\forall y(\text{man}(y) \supset \text{love}(x, y)))$ provient d'une autre structure syntaxique, construite en appliquant d'abord la règle S9 et ensuite la règle S10¹⁵. Cette construction dirige en parallèle l'application de la fonction F_9 puis de la fonction F_{10} , ce qui change la portée des quantificateurs.

La volonté de mener la construction sémantique et la construction syntaxique en parallèle conduit de manière naturelle à une conception fonctionnelle des structures sémantiques. La dépendance syntaxique d'un syntagme par rapport à un autre se traduit, au niveau des significations, par le fait que l'un devient argument de l'autre. Le développement de cette idée tire profit d'un certain nombre d'outils mathématiques puissants, en particulier le calcul des lambdas.

[lambda calculus] will provide a formalism in which we can assign meanings to the basic expressions of a language and all of the larger constituents constructible from the basic expressions. In other words, it forms the basis of the compositional method of defining the meanings of expressions in terms of the meanings of their parts. This can be contrasted with first order logic, which is rather impoverished in terms of being able to assign meanings to the relevant syntactic constituents of an expression. (CARPENTER 1997 [10] p. 39)

Le choix d'un formalisme comme le lambda-calcul est motivé par le caractère artificiel de la représentation déclarative. Un syntagme comme *ballon rouge* est traduit classiquement comme une conjonction logique $\text{ballon}(x) \ \& \ \text{rouge}(x)$ dans laquelle la variable x est

¹⁴ Nous rapportons ici, en la simplifiant, l'analyse faite dans (MONTAGUE 1974 [76]) de cet exemple, qui est analogue à la phrase *toutes les disquettes ont un virus* mentionnée dans la section précédente.

¹⁵ Dans le cas du présent exemple, l'alternative syntaxique suppose un syntagme verbal ternaire où le sujet et le complément jouent des rôles symétriques. Cette possibilité est refusée par de nombreux syntacticiens.

répétée pour signaler qu'elle désigne le même objet. En revanche, une traduction fonctionnelle comme (**rouge**(**ballon**)) fait apparaître une dépendance entre les deux termes, dépendance qui est parallèle à la subordination syntaxique de l'adjectif au nom. La même idée est présente dans le formalisme de la logique combinatoire¹⁶. Ce type de langage formel fait intervenir explicitement des termes qui sont des combinateurs. La combinaison résulte alors de la juxtaposition des termes. Le combinateur se comporte comme une fonction qui se combine avec un terme voisin en le prenant comme argument. Dans ce cadre, contrairement au cas des formalismes déclaratifs, les variables n'interviennent qu'en tant que variables "muettes" pour signaler les fonctions.

Le processus de combinaison sémantique résulte ainsi de l'application récursive, dirigée par la syntaxe, de fonctions à leurs paramètres. Au lieu de constituer des arbres complexes, le décodage de la phrase procède, sous l'action des combinateurs, en instanciant progressivement les arguments des fonctions par d'autres fonctions. Pour enclencher ce processus, les catégories syntaxiques élémentaires sont décrites comme des fonctions dont le type d'argument est précisé. Par exemple, dans le cadre d'une grammaire catégorielle, un verbe peut être noté par la catégorie $(np \setminus s) / np$, ce qui signifie qu'il accepte un syntagme nominal de catégorie np à droite, le complément, puis un syntagme nominal à gauche, le sujet, pour donner une phrase de catégorie s . De la même manière, un déterminant pourra être décrit par la catégorie np / n , puisqu'il prend un nom de catégorie n à droite pour donner un syntagme nominal. De même, les structures sémantiques peuvent être construites comme des combinaisons de termes qui représentent les significations de base. L'exemple suivant montre comment les deux structures, syntaxique et sémantique, sont construites en parallèle dans un formalisme de grammaire catégorielle¹⁷. Le signe ":" sépare dans chaque cas l'expression sémantique, à gauche, de l'expression syntaxique, à droite.

<i>Someone</i>	<i>breaks</i>	<i>everything</i>	
some: $np \uparrow s$	break: $(np \setminus s) / np$	every: $np \uparrow s$	(1)
$x:$ np		$y:$ np	(2)
	break (y): $np \setminus s$		(3)
	break (y)(x): s		(4)
	every ($\lambda y.$ break (y)(x)): s		(5)
	some ($\lambda x.$ every ($\lambda y.$ break (y)(x))): s		(6)

Cet exemple permet d'illustrer à nouveau la manière dont sont gérées les quantifications. Dans cette dérivation, le verbe accepte son complément en phase (3) et son sujet en phase (4). Noter que la dérivation sémantique suit fidèlement la progression syntaxique : chaque réduction syntaxique se traduit par une nouvelle relation fonction - argument. Le traitement sémantique des quantificateurs intercale des phases supplémentaires. La phase (2) consiste à créer des variables qui jouent le rôle d'hypothèses. Sous ces hypothèses, on conclut à la dérivation de la forme (4). La réécriture du lien

¹⁶ Le formalisme de la logique combinatoire est, en tant qu'outil de génération de dérivations, équivalent au calcul des lambdas.

¹⁷ Cet exemple est adapté de (CARPENTER 1997 [10]).

hypothèse - conclusion conduit successivement aux phases (5) et (6). L'ordre respectif de ces deux réécritures explique l'ambiguïté de la double quantification¹⁸.

Cette approche de la sémantique formelle dans laquelle les processus de construction du sens sont parallèles aux processus syntaxiques présente un intérêt indéniable d'un point de vue computationnel. Sur le plan de la modélisation cognitive, elle peut cependant être critiquée à plusieurs égards.

La première critique porte sur la vraisemblance de l'homomorphisme entre syntaxe et sémantique. L'existence d'une correspondance stricte entre la syntaxe de la langue et un système formel chargé de représenter la composition sémantique semble contredire toute possibilité d'autonomie relative de la syntaxe et de ce système formel. À la limite, pourquoi parler de deux systèmes combinatoires différents ? Or, d'un point de vue cognitif, il semble bien exister deux jugements d'acceptabilité différents, l'un syntaxique, portant sur l'agencement des mots, et l'autre sémantique, portant sur l'existence d'un sens. Comment expliquer, s'il existe une correspondance stricte, que le sens d'une phrase syntaxiquement incorrecte comme *Pierre pas aimer à Jacques* puisse, dans une situation donnée, être compris, alors que dans certains contextes, la phrase syntaxiquement correcte le carrelage approuve pourra rester incomprise ?

D'un point de vue technique, la thèse de l'homomorphisme strict peut signifier qu'il n'existe, cognitivement, qu'un seul système combinatoire. Ce qui nous apparaît comme la syntaxe ne serait alors que le versant structurel visible de la combinatoire sémantique. Peut-on analyser l'ensemble des phénomènes grammaticaux comme des conséquences de contraintes liées à l'articulation du sens ? La sémantique formelle va clairement dans cette direction. En confiant au lexique la quasi-totalité des instructions de combinaison, elle ne laisse plus de place à des phénomènes de combinaison déconnectés du sens. Dans le formalisme des grammaires catégorielles, par exemple, chaque entrée du lexique comporte un élément d'algorithme qui dirige sa propre insertion dans la phrase. L'expression $(np \setminus s) / np$, qui caractérise un verbe transitif, peut être lue comme un algorithme qui attend des données de type np sur sa droite puis sur sa gauche pour produire une expression de type s . Selon le même principe, les instructions contenues dans les entrées du lexique contrôlent leur propre comportement phonologique et sémantique, si bien que le lexique est, idéalement, censé diriger l'ensemble des processus linguistiques. Une telle position extrême est cependant difficile à tenir en raison de la présence de phénomènes syntaxiques globaux sans contrepartie sémantique. Par exemple, aucune contrainte d'ordre lexical, phonologique ou sémantique n'explique que certaines langues comme le français ou l'anglais imposent systématiquement la présence d'un sujet pour les verbes conjugués, au point d'introduire des sujets explétifs comme le pronom *il* dans la phrase *il pleut*, alors que des langues comme le persan se satisfont de verbes sans sujets. De même, le mode de branchement des syntagmes diffère d'une langue à l'autre, ce qui conduit par exemple le persan à être de type $SO \square$ alors que le français est $S \square O$. Parmi les contraintes structurelles systématiques qui dépassent le lexique, on peut également citer la contrainte de liage qui agit par la relation de *c-commande*, que nous avons mentionné dans la section précédente. Ces phénomènes de structure requièrent l'adjonction de

¹⁸ Cet exemple utilise les principes du calcul de Lambek. La forme $P : np \setminus (s)$ induit une dérivation conditionnelle. Si, sous l'hypothèse $x : np$, posée lorsque P se présente, on parvient à dériver syntaxiquement $Q(x) : s$, alors on a dérivé $P(\lambda x. Q(x)) : s$. Ainsi, le quantificateur est traité comme une fonction dont l'argument est une fonction de la variable quantifiée. Noter que, contrairement à l'exemple précédent, l'ambiguïté liée à la portée respective des quantificateurs n'est pas d'origine syntaxique. La production de **every** $(\lambda y. \text{some } (\lambda x. \text{break } (y) (x)))$ n'est pas liée à l'ordre dans lequel les arguments du verbe sont intégrés, mais provient d'un ordre différent dans le "déchargement" des hypothèses.

mécanismes généraux à la grammaire, par exemple le mécanisme de montée de type¹⁹ utilisé dans les grammaires catégorielles. De telles techniques conduisent toutefois à réintroduire des opérations purement syntaxiques, au détriment du parallélisme syntaxe - sémantique qui est censé motiver l'ensemble de la démarche. Ainsi, il est difficile de concevoir techniquement un homomorphisme parfait. Or, toute différence de traitement laisse entrevoir la nécessité de deux mécanismes distincts, l'un chargé des opérations syntaxiques et l'autre des opérations sémantiques, ainsi que d'une interface entre les deux.

Une deuxième critique que l'on peut adresser à la sémantique formelle d'un point de vue cognitif est, paradoxalement peut-être, la trop grande puissance de ses formalismes. Le formalisme de la logique combinatoire ou le lambda-calcul permettent d'exprimer l'ensemble des fonctions calculables²⁰. Il faut donc imaginer que les catégories admises pour représenter la compétence langagière appartiennent à un sous-ensemble restreint de toutes les expressions formelles possibles. Or, aucun critère permettant de délimiter ces dernières ne se dégage clairement. Quel principe interdit l'existence d'une catégorie comme $(np \setminus np) / np$? Si ce principe existe, comment peut-il être justifié sur le plan de l'organisation cognitive? Il est en effet plus facile de justifier, par exemple par des explications fonctionnelles, l'existence d'une capacité que sa non existence. Si l'esprit humain possède une capacité de calcul générale capable de gérer des expressions formelles comme celles des grammaires catégorielles, pourquoi est-il restreint à ne manipuler que certaines d'entre elles? Si, en revanche, cette capacité générale de calcul n'existe pas, alors les modèles de la sémantique formelle ne peuvent être que des modèles de description, non des modèles cognitifs.

Une troisième critique peut être adressée au présupposé principal des approches formelles de la sémantique, l'idée que la combinaison conceptuelle est réalisée par une relation fonction - argument. Par exemple, la modification du nom *ballon* par l'adjectif *rouge* est rendue par le fait que le terme *ballon* devient argument du terme *rouge*. Or, le fait de traduire les constructions grammaticales par des structures fonctionnelles ne suffit pas à nous faire progresser sur le plan cognitif. La signification du syntagme *ballon rouge* n'est pas élucidée par une expression comme *rouge(ballon)*. Comme dans toutes les approches formelles, la charge de l'interprétation est repoussée dans la procédure de vérification, seule capable d'établir un lien entre les symboles *rouge* et *ballon*. La dérivation donnée plus haut pour la phrase *someone breaks everything* illustre le fait que le décodage se contente de transférer, étape par étape, la combinatoire de l'énoncé langagier dans la structure d'une expression formelle. Ce décodage ne nous renseigne pas sur l'interprétation de l'énoncé. Par exemple, il ne rend pas compte du lien systématique entre les expressions *rouge(ballon)* et *vert(ballon)*. Pour cela, il faut conférer une structure interne aux symboles utilisés dans le formalisme, de manière à ce que les fonctions *rouge* et *vert* puissent partager le fait d'être des indications de couleur. C'est ce que proposent les approches décrites dans la section suivante.

Dans la présente section, nous avons considéré l'apport des systèmes de sémantique formelle pour la compréhension de la construction du sens des énoncés. Si de tels systèmes présentent un intérêt technique parfois incontournable, ils n'en constituent pas pour autant des modèles cognitifs plausibles des processus en jeu dans la construction du sens. En particulier, ils repoussent la difficulté de l'interprétation en postulant l'existence d'une procédure générale de vérification, dont les propriétés sont cognitivement non plausibles (tests binaires dans des ensembles infinis de mondes eux-mêmes infinis). Par ailleurs, la structure

¹⁹ Le mécanisme de montée de type transforme par exemple un terme de catégorie *np* en catégorie $s / (np \setminus s)$ (*id est* une phrase attendant un groupe verbal). La catégorie *np* se voit ainsi imposer la contrainte d'apparaître en position sujet.

²⁰ Ces systèmes ont la puissance d'une machine de Turing.

sémantique utilisée dans les approches formelles, en associant un symbole sémantique différent à chaque entrée lexicale, n'explique pas les liens systématiques qui peuvent exister entre les combinaisons conceptuelles. Il est donc légitime de s'intéresser à une approche plus cognitive de la sémantique, dans laquelle le formalisme utilisé pour représenter les significations est censé modéliser les représentations telles qu'elles sont supposées être évoquées par le langage et ancrées dans la perception.

5.3. Concepts et structures

Une approche cognitive de la sémantique se doit de déterminer les représentations et les mécanismes mentaux à l'œuvre dans la procédure de construction du sens, non des opérations supposées se produire dans des espaces abstraits servant de substitut à une réalité placée à l'extérieur des locuteurs. Certaines approches cognitives de la sémantique sont ainsi guidées par l'idée que le sens d'un énoncé est une représentation mentale structurée, et que l'objectif de la recherche en sémantique est de déterminer cette structure.

[...] the act of understanding a sentence S – recovering its meaning – is to be regarded as standing S in correspondence with a mentally encoded concept C which has internal structure derivable from the syntactic structure and lexical items of S.

(JACKENDOFF 1990 [50] p. 11)

La construction du sens d'un énoncé s'explique par la composition mentale des représentations conceptuelles évoquées par les mots de l'énoncé. L'objectif est donc de décrire une interface entre deux systèmes compositionnels, le système syntaxique et le système conceptuel. L'explication de cette interface commence par le fait d'associer une structure conceptuelle à chaque entrée lexicale (LCS).

An example like [the following one] provides a good first illustration [for the relation of syntactic and conceptual argument structure].

Syntactic structure

[S [NP John] [VP ran [PP into [NP the room]]]]

Conceptual structure

[Event GO ([Thing JOHN], [Path TO ([Place IN ([Thing ROOM])])])]

[...] In order to see how [it] is put together from its parts, let us examine the lexical entries for the two words *into* and *run* [...].

into]
P	
__ NP _j	
└ [Path TO ([Place IN ([Thing]j)])]	└

run]
□	
__ <PP _j >	
└ [Event GO ([Thing]i, [Path]j)]	└

(JACKENDOFF 1990 [50] p. 45)

La LCS peut être simple, comme les concepts JOHN ou ROOM. Elle peut comporter des indications de type, comme *Thing* ou *Path*, et des fonctions comme GO ou TO qui

transforment leurs arguments en une nouvelle structure. L'énoncé *John ran into the room*, du fait de la fonction GO amenée par la LCS du verbe *run*, est interprété comme un événement. La correspondance avec la syntaxe permet de placer le concept JOHN, qui est l'interprétation du sujet grammatical, en premier argument de la fonction GO. De la même façon, la structure associée au groupe prépositionnel prend la place du deuxième argument de la fonction GO. Ces inclusions de structures conceptuelles au sein d'autres structures conceptuelles s'effectuent en respectant les types. Ainsi, le concept ROOM est du bon type *Thing* pour être argument de la fonction IN. Cette fonction transforme le concept ROOM en type *Place*, ce qui lui permet d'être l'argument de la fonction TO. L'ensemble a le bon type, *Path*, pour venir occuper le deuxième argument de la fonction GO.

La structure conceptuelle ainsi proposée est donc plus riche que celle que requiert la thêta-théorie de la grammaire générative, qui se réduit à préciser une structure argumentale, notamment pour les verbes et les prépositions. Elle est également plus complexe que la forme logique des approches formelles qui associent une fonction à chaque entrée lexicale, sans souci des liens systématiques que ces fonctions conceptuelles peuvent avoir entre elles.

Instead of a division of formal entities into such familiar logical types as constants, variables, predicates, and quantifiers, each of which has nothing in common with the others, it is argued that the essential units of conceptual structure are *conceptual constituents*, each of which belongs to one of a small set of major ontological categories (or conceptual « parts of speech ») such as Thing, Event, State, Action, Place, Path, Property, and Amount. [...] Each conceptual category has some realizations in which it is decomposed into a function argument structure; each argument is in turn a conceptual constituent of some major category. (JACKENDOFF 1990 [50] p. 22)

Cette approche de la représentation conceptuelle conserve l'aspect fonctionnel des approches formelles. Cependant, les fonctions ne sont plus l'ensemble des symboles associés aux entités lexicales de la langue, mais sont limitées à un petit nombre de primitives. Ces fonctions primitives sont supposées représenter des composantes fondamentales de notre cognition. Le fait de représenter le sens en appliquant ces fonctions aux arguments qui leur sont présentés devient cognitivement plausible, dans la mesure où grâce au typage, l'ensemble des configurations possibles reste limité. Le point fort de l'approche est ainsi que les structures conceptuelles tirent leur justification du fonctionnement cognitif, non de la structure de la langue. Le lien avec le langage prend donc la forme d'une véritable interface, dont la loi générale peut s'exprimer de la manière suivante.

[...] the fact that major syntactic phrases correspond to major conceptual constituents [...] and the basic correspondence of syntactic and conceptual structure [...] [can be formalized as the following general correspondence rules].

XP corresponds to [Entity].

$$\left[\begin{array}{c} X^0 \\ _ _ < YP < ZP >> \end{array} \right] \text{ corresponds to } \left[\begin{array}{c} \text{Entity} \\ F (< E_1 >, < E_2, < E_3 >>) \end{array} \right]$$

(JACKENDOFF 1990 [50] p. 25)

XP désigne un syntagme dont la tête est l'item lexical X^0 . Les syntagmes YP et ZP sont les compléments du syntagme global. À cette structure syntaxique correspond la structure sémantique $F(< E_1 >, < E_2, < E_3 >>)$ où F est une fonction sémantique²¹, E_2 et E_3 sont les structures conceptuelles associées respectivement aux expressions YP et ZP, E_1 est associée au spécifieur du syntagme, s'il est présent. Ainsi, la structure conceptuelle de la phrase *John*

²¹ Les crochets < > indiquent un caractère optionnel.

ran into the room conduit à l'expression fonctionnelle GO(JOHN, TO(IN(ROOM))) dont les deux arguments sont occupés par les structures conceptuelles respectives du sujet John et du complément into the room. De même, la structure conceptuelle du syntagme into the room est une expression fonctionnelle qui s'exprime comme la composée de deux fonctions primitives IN et TO et dont le seul argument est occupé par le groupe nominal the room.

L'intérêt premier de ce genre de modèle est qu'il ne fait pas de différence qualitative entre les structures lexicales et supra-lexicales. En d'autres termes, il s'agit d'un système intégralement compositionnel : les structures conceptuelles lexicales, comme les structures conceptuelles des énoncés, résultent de compositions de concepts primitifs. Grâce à ce principe, il est possible de réduire considérablement l'ensemble des éléments primitifs, ceux pour lesquels on doit spécifier directement les propriétés, notamment leur ancrage et leur rôle inférentiel. Le sens de toutes les autres structures peut, en principe, être déduit par composition. La tâche de la sémantique lexicale se ramène ainsi, pour l'essentiel, à déterminer la structure sémantique que l'on attribue aux unités du lexique. De cette manière, contrairement à l'approche formelle, il n'est pas nécessaire de postuler l'existence d'une procédure de vérification externe capable de décider de l'interprétation de toute expression. Le sens des mots, comme celui des phrases, s'obtient par une combinaison récursive de significations de base.

Le principe d'une interface systématique entre le système conceptuel et le système grammatical semble cependant mis en défaut par un certain nombre de difficultés. La structuration des entités sémantiques exige que la syntaxe fournisse des arguments convenables à ces structures, c'est-à-dire des objets sémantiques présentant le type adapté à la place à laquelle ils sont destinés. De ce point de vue, la sémantique formelle pouvait se montrer beaucoup plus permissive, en ne posant pas de contraintes dans la relation fonction - argument autre qu'une spécification de la catégorie syntaxique attendue. Or, dans l'approche de la sémantique conceptuelle, les contraintes de typage sémantique semblent souvent enfreintes. Par exemple, dans la phrase la sérénité entre dans la pièce, la définition sémantique du verbe entrer suppose que le sujet vienne occuper un argument typé comme chose, or le concept associé au mot sérénité est plutôt de type propriété. Ainsi le concept SÉRÉNITÉ ne peut pas être "substitué" à la structure du premier argument de la fonction associée au verbe entrer. Ce problème peut être résolu par l'introduction d'un mécanisme de fusion, qui remplace celui de substitution (JACKENDOFF 1990 [50]). L'idée est qu'en cas de conflit de structures, la structure conceptuelle associée au syntagme en position de sujet doit être "fusionnée" avec la structure du premier argument attendu par le syntagme verbal. Cette opération autorise un assouplissement local des contraintes d'appariement, ce qui permet en principe d'expliquer les tournures non littérales, jusqu'à certaines métaphores²².

Une autre difficulté se présente avec des phrases comme la vitre casse. La définition du verbe casser, qui permet d'interpréter une phrase comme Jean casse la vitre, peut être décrite par la structure suivante.

[Event CAUSE([Entity _],[Event GO([Entity _],[Path TO([State CASSÉ])])])]

Or, dans la phrase la vitre casse, on s'attend à ce que le sujet grammatical vienne occuper la place du premier argument de la fonction GO, contrairement à ce que prévoit l'interface standard. On peut imaginer trois manières de résoudre ces problèmes d'interface (JACKENDOFF 1990 [50]). (1) Il est possible de traiter le problème de manière purement syntaxique. Le syntagme la vitre occuperait la position sujet uniquement dans la structure de

²² Techniquement, la substitution et une partie de la fusion peuvent être assurées par le mécanisme standard d'unification. Ce même mécanisme peut également permettre de contrôler jusqu'à quel niveau les contraintes peuvent être enfreintes. Certaines métaphores, comme l'entrée de la sérénité dans une pièce, peuvent ainsi être expliquées et prédites.

surface, à la suite d'un mouvement provoqué par la vacuité de cette position dans la structure profonde. Dans ce cas, l'interface syntaxe - sémantique reste valide si elle s'effectue au niveau de la structure profonde. Cette solution est naturelle en génération, lorsque la structure profonde est élaborée avant la structure de surface. En compréhension, cependant, la reconstitution d'une structure profonde capable de recevoir un sens passerait par un échec, lorsque l'argument de la fonction GO, le thème, ne peut être apparié. Cet échec serait suivi d'un retour en arrière²³. Or il n'y a pas d'indication claire selon laquelle la compréhension de la phrase *la vitre casse* demanderait un temps de traitement supérieur à celle de la phrase *il casse la vitre*. (2) Une deuxième solution consiste à complexifier la règle de correspondance syntaxe - sémantique en introduisant une hiérarchie entre les rôles sémantiques. Par exemple, le premier argument de la fonction CAUSE, que l'on peut dénommer agent, serait prioritaire pour s'interfacer avec le sujet grammatical. Dans le cas où il est non spécifié, c'est le thème, le premier argument de la fonction GO qui pourrait se saisir du sujet. Noter qu'encore une fois, l'explication est naturelle en génération, mais l'est beaucoup moins en compréhension où elle nécessite un effet de retour en arrière. (3) Une troisième solution est de considérer des verbes comme *casser* comme polysémiques. En particulier, ils comporteraient deux définitions, l'une causale et l'autre non. Ainsi, le verbe *casser* aurait une deuxième définition décrite par la structure suivante.

[Event GO([Entity _],[Path TO([State CASSÉ])])]

La sélection de la bonne définition dépendrait ainsi de la présence d'un complément. Une solution équivalente est de marquer la fonction CAUSE comme optionnelle dans la définition initiale du verbe *casser*. Il est important de noter que la mention d'une ambiguïté dans l'entrée lexicale est une mention purement sémantique. Rien ne distingue le verbe *casser* d'un verbe comme *manger* sur le plan syntaxique, puisque les deux verbes peuvent être utilisés indifféremment avec ou sans complément. Alors que les deux premières solutions se traduisaient par une complication des mécanismes d'interface, la présente solution consiste à augmenter la complexité des représentations sémantiques de manière à prédire correctement les conditions d'emploi des mots associés.

Si l'on systématisait la logique de cette dernière approche, on en vient à classifier de manière toujours plus fine les structures lexicales. La méthodologie linguistique correspondant à ce courant part de l'observation des configurations syntaxiques où une entité lexicale apparaît, puis tente de doter cette entité lexicale des traits sémantiques minimaux qui permettent de prédire ses conditions d'emploi. La structure conceptuelle est ainsi justifiée par la seule prise en compte du caractère compositionnel des mots.

[...] knowing the meaning of a verb can be a key to knowing its behavior. Presumably, predictions about verb behavior are feasible because particular syntactic properties are associated with verbs of a certain semantic type. [...] [Our examples] suggest that general principles of grammar are at work, allowing the syntactic behavior of a verb to be predicted from its meaning. (LEIN 1993 [66] p. 5)

L'inclusion de traits appropriés dans la structure conceptuelle permet par exemple de distinguer les conditions d'emploi de verbes comme *break*, *cut* et *hit*²⁴.

²³ Il s'agit du phénomène connu sous le nom de *garden path*. Le sujet interprète une phrase comme *the horse ran past the barn fell* de manière directe, en traitant, dans un premier temps, *ran* comme un verbe conjugué. Emporté ainsi par son élan, il se heurte au verbe *fell* qui n'entre pas dans la trame syntaxique élaborée jusque-là. Dans cette situation, il effectue un retour en arrière pour réinterpréter *ran* comme le participe passé d'un verbe transitif.

²⁴ Exemples adaptés de (LEIN 1993 [66]).

Janet broke the cup / the cup broke
Janet cut the bread / * the bread cut

Carla cuts the meat / the meat cuts easily
Carla hits the door / * the door hits easily

L'idée est non seulement d'attribuer ces contrastes à des propriétés structurelles des mots, mais également de classer les mots selon leur comportement. De cette manière, les mots dérivent de classes plus ou moins générales.

By defining the functional behavior of lexical items at different levels of representation we hope to arrive at a characterization of the lexicon as an active and integral component in the composition of sentence meanings. Because of the more expressive mechanisms involved in composition, this approach will enable us to conflate different word senses into a single *meta-entry*, encoding regularities of word behavior dependent on context, and as a result, greatly reducing the size of the lexicon. (PUSTEJOŃSKY 1995 [86] p. 62)

Les différentes approches génératives du lexique voient donc dans la sémantique des mots une combinaison de composants conceptuels en nombre limité. Les versions radicales se donnent comme objectif de prédire, à l'aide de ces structures, non seulement les combinaisons grammaticales standard, mais également les possibilités d'insertion dépendantes du contexte sémantique. Dans l'exemple suivant, on constate que les combinaisons dans lesquelles le verbe *began* est admis dépendent de la nature sémantique des autres mots²⁵.

Mary began to bleed / her bleeding began
Mary began to draw / * her drawing began

Mary began the book / * the book began
Mary began the movie / the movie began

Le syntagme *begin to* + GV peut recevoir deux interprétations différentes : il peut s'agir d'un effet causal, exercé par le sujet, sur l'amorce d'un événement ; il est aussi possible que seul le début de l'événement soit pertinent, sans qu'il y ait une référence à l'origine causale de son départ²⁶. Avec le verbe *bleed*, l'interprétation non causale est préférentiellement sélectionnée, alors que l'interprétation causale est induite par le verbe *draw*. Cela se vérifie par le fait que la paraphrase *her drawing began* est peu acceptable. De même, dans les exemples de la forme grammaticale *begin to* + GN c'est la définition causale qui est préférentiellement choisie. Ceci peut être expliqué par un mécanisme de coercition. Par exemple, dans le cas du mot *book*, on peut supposer deux coercitions différentes, celle où le livre est lu et celle où le livre est écrit. Par contre aucune coercition associée à la structure conceptuelle du mot *book* ne peut donner du sens à la construction *the book began*. En revanche, parmi les coercitions acceptées par l'entrée lexicale *movie*, se trouve une interprétation qui lui permet de se trouver en position du premier argument du verbe *begin*, celle où la projection du film commence. On constate donc que le sens d'un composé s'élabore par un processus de sélection mutuelle entre les traits sémantiques des composants.

L'ambition de la sémantique lexicale est de parvenir à discriminer les combinaisons licites de mots, de la même manière que les syntacticiens tentent de discriminer les combinaisons licites des catégories syntaxiques. Pour cela, le moyen est de constituer des

²⁵ Exemples adaptés de (PUSTEJOŃSKY 1995 [86]).

²⁶ Cette alternative est analysée, dans certaines théories syntaxiques, par une distinction entre deux catégories grammaticales différentes, verbe de contrôle et verbe de montée, qui peuvent être attribuées à deux entrées lexicales différentes du verbe *begin*.

structures toujours plus riches, en ajoutant des traits restrictifs lorsque la discrimination échoue et en introduisant des disjonctions ou des traits facultatifs lorsque des configurations permises ne sont pas prédites. Cette ambition se justifie sur un plan technique. Peut-on espérer parvenir ainsi à un modèle cognitif du système conceptuel ? Il faut pour cela que la méthode lexicale parvienne à enfermer l'intégralité des conditions d'emploi des mots dans des structures fixes. Nous évoquerons plus tard le danger que l'inventaire des traits introduits ne soit pas borné, et que l'origine, ontogénétique ou phylogénétique, de ces traits soit difficile à justifier (CF. CHAPITRE 7). Pour le moment, nous pouvons nous interroger sur les limites de la méthodologie, à la lumière des exemples précédents.

La définition que nous avons donnée du verbe *casser* utilise la primitive *CASSÉ*. Si on la remplace par une mention plus générale comme *NON FONCTIONNEL*, le risque est qu'il existe des objets cassés qui restent fonctionnels (par exemple un couteau) et des objets qui cessent d'être fonctionnels sans être cassés (par exemple un ordinateur). Il n'est donc pas évident que l'inventaire des traits soit significativement plus concis que le lexique lui-même. De plus, les emplois considérés comme illicites le sont dans des situations standard ou prototypiques. Ainsi, la construction *le livre commença* est un énoncé sémantiquement acceptable dans un conte où un livre magique raconte son contenu ou s'il s'agit d'un livre électronique en défilement automatique. Un énoncé comme *il commence le tunnel*, semble exclure l'interprétation *il commence à parcourir le tunnel*, au profit de celle de *il commence à creuser le tunnel*. Cependant cette exclusion présente un caractère relatif. La phrase peut très bien figurer, avec ce sens, dans le jargon de l'équipe de surveillance du tunnel. Dans ce cas, l'exclusion est plus dans les habitudes que dans le concept lui-même, si les habitudes peuvent changer sans que le sens en soit affecté. L'exclusion sémantique n'est donc pas de même nature que l'exclusion syntaxique. Aucune modification de contexte ni aucune habitude ne peuvent rendre la phrase *les frères espèrent que Jean aime l'un l'autre* syntaxiquement acceptable²⁷. Si l'acceptabilité des combinaisons sémantiques peut toujours être remise en question par le contexte ou l'habitude, le fait d'enfermer l'acceptabilité dans des structures lexicales ne peut atteindre que les conditions d'emploi standard et prototypiques. Dans ce cas, la méthode ne peut pas conduire, seule, à un modèle de la compétence sémantique humaine.

Une manière de rendre les structures adaptables au contexte est de leur permettre de s'établir de manière dynamique, en y incluant des procédures. C'est cette solution que proposent les approches examinées dans la section suivante.

5.4. Concepts et procédures

La construction du sens peut être conçue non comme un assemblage d'unités prédéfinies, mais comme un calcul pouvant inclure divers aspects contextuels, y compris des faits contingents ou des données de la perception. L'une des idées principales de la sémantique procédurale consiste à élaborer le sens des composés de manière dynamique. Cette idée part de la métaphore des programmes informatiques, qui sont compilés puis exécutés. De la même manière, un énoncé langagier doit être analysé et compilé pour prendre la forme d'un plan d'action exécutable. Le plan peut être comparé à la forme intensionnelle dans la sémantique formelle, tandis que le résultat de l'exécution correspond à la valeur de vérité allouée à une formule dans un certain domaine (JOHNSON-LAIRD 1977 [54]). Ainsi, un

²⁷ Cet exemple, inspiré de (CHOMSKY 1975 [15]), illustre le fait que la présence du prénom Jean en position sujet bloque le lien entre les deux syntagmes *les hommes* et *l'un l'autre*, contrairement au cas de la phrase *les frères espèrent qu'ils s'aiment l'un l'autre*, qui est syntaxiquement bien formée.

énoncé possède un sens s'il peut conduire à un plan, et si, dans un contexte donné, il donne lieu à une exécution réussie.

Pour illustrer ce point de vue, considérons des énoncés simples à propos du jeu d'échecs. L'énoncé mets le fou à droite de la dame peut conduire, lorsque la dame se trouve sur la case c3, à un plan qui consiste à déplacer le fou approprié sur la case d3. Si ce plan peut être exécuté dans le contexte courant, on peut dire que l'énoncé possède un sens. Dans le cas où le fou est bloqué ou s'il n'y a plus de fou de la couleur appropriée, l'énoncé peut être considéré comme dénué de sens²⁸. Il en est de même si la case désignée n'existe pas, comme dans le syntagme à droite de h3, où la case h3 est au bord de l'échiquier. Dans ces exemples, il n'est pas besoin de considérer une procédure de vérification valable en tout temps, en tout lieu et dans tous les mondes possibles. Il suffit d'exécuter le plan suggéré par l'énoncé dans le contexte courant.

Pour que l'exécution soit possible, il faut qu'aux mots du lexique soient associées des informations nécessaires, qui peuvent prendre la forme de schémas.

We have mentioned that the compiler must have access to a lexicon, but we have yet to consider what information the lexicon should contain, or how it should be organized. [...] We took the view that a lexical concept interrelates a word, rules governing its syntactic behavior, and a schema. A schema is made up from both functional and perceptual information, and may well include information that has no direct perceptual consequence. Moreover, lexical concepts are interrelated to one another. They are organized into semantic fields that have a *conceptual core* which reflects a deeper conceptualization of the world and integrates the different concepts with the semantic field.

(JOHNSON-LAIRD 1977 [54] p. 203)

Les schémas sont des routines élémentaires qui permettent de tester la présence d'objets ou de configurations, et exécuter des actions simples. Le joueur d'échec dispose de schémas pour reconnaître un roi, pour repérer une double menace ou pour effectuer une prise en passant. Pour calculer le sens des énoncés, le sujet doit combiner certains schémas sélectionnés parmi ceux qui sont associés aux mots de l'énoncé. Lors de cette combinaison, les schémas associés aux mots se sélectionnent mutuellement de manière à s'adapter les uns aux autres, en tenant compte des rapports syntaxiques. Ainsi, pour exécuter le syntagme le fou à droite de la dame, l'auditeur commence par traiter le complément le plus interne, le groupe nominal la dame, en déclenchant un schéma par lequel il localise la dame sur l'échiquier. Le traitement du syntagme englobant, à droite de __, l'amène à enchaîner le schéma précédent avec un schéma effectuant un déplacement d'une case vers la droite. Cette case étant blanche, l'auditeur sélectionne le schéma associé au mot fou qui consiste à localiser son fou "blanc", puis, pour terminer le traitement du verbe sous-entendu, il enchaîne un schéma de déplacement de la pièce pour amener celle-ci à droite de la dame.

Les schémas ne sont pas des routines indépendantes. Par exemple, les schémas associés aux verbes mettre, déplacer, jouer, dans le contexte du jeu d'échecs, partagent un noyau de sens qui inclut l'action de déplacement d'une pièce. Il faut donc imaginer l'ensemble des schémas comme un réseau d'interrelations. Ce réseau permet de reproduire, sous une forme émergente, l'organisation en champs sémantiques qui était obtenue, dans la section précédente, par la structuration des concepts à partir de primitives. L'intérêt du réseau d'interrelations va cependant au-delà, en permettant au système de s'adapter au contexte par la sélection dynamique des schémas appropriés.

Une manière de représenter les schémas et leur sélection dynamique consiste à utiliser des trames (*frames*). Les trames sont couramment utilisées en intelligence artificielle et sont à

²⁸ Dans une perspective pragmatique, il serait plus juste de considérer l'énoncé comme non pertinent.

l'origine des langages à objets. Elles constituent un moyen puissant et souple de représenter dans un même format les données factuelles, les connaissances procédurales et les informations valables par défaut. Ainsi, une trame associée à l'individu Jean peut stocker le fait que son anniversaire est mercredi. Par héritage, la trame permet d'accéder à l'ensemble des propriétés par défaut que possède un homme adulte. Grâce à un pointeur, la trame permet de mémoriser que la fille de Jean est une jeune fille du nom de Maud dont les caractéristiques sont stockées dans une autre trame. L'utilisation de pointeurs et leur modification dynamique permet de constituer le sens d'un énoncé. Ainsi, pour interpréter un énoncé comme la fille de Jean possède une Ferrari, un tel système interprète le syntagme la fille de Jean en accédant à la trame associée à l'individu Maud, puis établit dynamiquement un pointeur entre le champ *possession*, dans cette trame, et une nouvelle trame, créée par héritage à partir de la trame générale associée au mot *voiture* et portant la mention *Ferrari* dans le champ *marque*. L'interprétation de l'énoncé passe par le succès de cette procédure. Celle-ci peut échouer, par exemple si l'on considère le syntagme *fils de Jean* sans que le pointeur nécessaire existe, où si l'on cherche à interpréter le syntagme *le fils de la voiture*, sachant que la trame associée au mot *voiture* ne peut recevoir de champ *fils*.

La notion de schéma, instanciée entre autres par les trames, présente certes une ressemblance avec les structures mentionnées dans la section précédente : il s'agit ici aussi d'expliquer le caractère compositionnel d'un concept en y enfermant des contraintes qui doivent diriger son comportement dans les compositions possibles. Cependant, les deux types de représentations se distinguent par la portée des contraintes qu'elles posent sur la composition.

In various implementations, schemata correspond to Ceccato's *constellations* (1961), Minsky's *frames* (1975), and Schank and Abelson's *scripts* (1977). [...] By enforcing selectional constraints, canonical graphs rule out anomalies like green ideas sleeping, but they allow such unlikely combinations as purple cows. [...] By incorporating more knowledge about the world, schemata favor plausible combinations and avoid less likely possibilities. In short, canonical graphs represent everything that is *conceivable*, and schemata represent everything that is *plausible*. (SOWA 1984 [99] p. 128)

Les graphes canoniques dont il est question dans cette citation déterminent, sous la forme d'un ensemble de liens inter-conceptuels, l'ensemble des combinaisons dans lesquelles un concept peut intervenir sans que le résultat soit contradictoire avec le noyau dur de nos connaissances sur le monde perçu. Les schémas, quant à eux, rendent compte de la plausibilité des combinaisons grâce à la prise en compte dynamique de l'ensemble des connaissances conceptuelles, y compris les connaissances contingentes ou celles qui relèvent de l'intégration des situations typiques. Cette alternative rejoint le débat présenté dans le chapitre précédent à propos du rôle inférentiel des concepts. Un courant important en intelligence artificielle défend la thèse selon laquelle le caractère compositionnel d'un concept ne peut pas être cerné complètement sans considérer toutes les connaissances qui le font intervenir.

La question abordée ici relève à la fois de la sémantique lexicale (c'est l'ensemble des usages du mot "livre" qu'il s'agit de définir) et de la représentation des connaissances (on veut exprimer les propriétés du concept "livre"). Contrairement aux linguistes, qui cherchent à établir une opposition entre "le linguistique" et "l'extra-linguistique", nous ne voyons pas de véritable frontière entre ces deux domaines, bien qu'il y ait des phénomènes qui se situent clairement d'un côté ou de l'autre. (KAYSER 1997 [59] p. 225)

Ainsi, pour cerner le caractère compositionnel d'un concept, *id est* les combinaisons sensées dans lesquelles le mot associé peut apparaître, il faut auparavant analyser l'ensemble des connaissances impliquant ce concept. Cette analyse produit, pour chaque concept, un ensemble de structures distinctes.

À supposer ces problèmes résolus, c'est-à-dire l'existence d'un ensemble complet de prédicats Livre, Livre¹, ..., Livreⁿ, il resterait à *déterminer* en toute occurrence le prédicat précis qui correspond à une connaissance donnée. [...] Un livre s'exprimera dans tous les cas comme un objet x vérifiant le prédicat livre(x). C'est un raisonnement qui permettra de mieux cerner, si besoin est, ce que dénote un objet vérifiant ce prédicat ; ce raisonnement pourra être mené plus ou moins loin, et il arrivera qu'il ouvre plusieurs pistes.

(KAYSER 1997 [59] p. 228)

La recherche de la structure adéquate, dans un contexte donné, peut s'effectuer grâce à une procédure qui est déclenchée systématiquement, à chaque occurrence du mot, mais dont le déroulement et le résultat peuvent être, chaque fois, différents. Dans un certain contexte, la procédure traitera le mot livre comme un prédicat concernant l'objet, dans un autre cas le même mot conduira à un prédicat relatif à l'œuvre, *et cætera*. La procédure recherche, dans l'ensemble des structures, une structure qui soit compatible avec les connaissances contextuelles déjà construites. Cette recherche de structure est une recherche à "profondeur variable", qui s'arrête à la première structure exempte de contradictions (KAYSER 1997 [59]).

Les modèles regroupés sous le nom de sémantique procédurale ont ceci en commun qu'ils contiennent, d'une part, un réseau d'interrelations de concepts d'où sont extraits les schémas et, d'autre part, un ensemble de procédures qui s'exécutent pour produire le sens. Ces modèles sont utilisés en traitement du langage naturel, mais également dans certaines autres applications où l'idée de composition conceptuelle est pertinente. Un exemple nous est fourni par le programme Copycat (HOFSTADTER 1995 [47]). Ce programme s'efforce de reproduire les analogies que des sujets humains produisent sans effort sur des suites de lettres. Par exemple, en suivant le modèle $abc \rightarrow abd$, il s'agit de produire, à partir de $ppqqrrr$, la suite $ppqqss$. La configuration d'une suite de lettres peut être décrite par un nombre limité de relations de type successeur, prédécesseur, gauche, droit, dernière milieu *et cætera*. L'architecture de ce programme, dont de nombreux aspects sont hors de propos ici, repose sur l'idée d'un partage de la connaissance entre un réseau de concepts relativement stable et une procédure chargée d'activer des schémas temporairement.

The basic image for the Slipnet [Copycat's network of platonic concepts] is that of a network of interrelated concepts, each concept being represented by a *node* [...], and each conceptual relationship by a *link* having a numerical length, representing the "conceptual distance" between the two nodes involved. (HOFSTADTER 1995 [47] p. 212)

Le système peut activer de nouveaux liens en propageant les activations dues au contexte. Par exemple, les deux concepts SUCCESSEUR et PRÉDÉCESSEUR sont liés par la relation OPPOSÉ. Ceci fait partie du noyau conceptuel de ces nœuds. D'autres couples possèdent le même lien, par exemple le couple DROITE/GAUCHE, peut-être avec une intensité moindre. L'idée est que quand la relation OPPOSÉ est activée par le contexte, toutes les paires de concepts qui sont unies par ce lien s'activent, plus ou moins fortement. Le programme parvient à réaliser une analogie lorsque l'une de ces structures, celle qui est la plus fortement activée s'il y en a plusieurs, s'applique au motif à traiter. L'activation de ces schémas est gérée à l'aide d'un grand nombre de petits programmes qui, dans Copycat, prennent le nom de "codelet".

Typical *effector codelets* do such things as: attaching a description to an object (e.g., attaching the descriptor *middle* to the *b* in *abc*); bounding two objects together (e.g., a *successor* bond between the *b* and *c* in *abc*); making a group out of two or more adjacent objects that are bonded together in a uniform manner; making a bridge that joins similar objects in distinct strings (similarity being measured by proximity of description in the Slipnet); destroying groups or bonds, and so on. (HOFSTADTER 1995 [47] p. 221)

Ce principe des codelets peut être utilisé pour rendre compte de l'interprétation des énoncés du langage. L'activation de structures conceptuelles peut être déclenchée par l'observation des structures langagières, et ceci de manière non déterministe. Dans un tel modèle, l'aspect procédural réside dans une multitude de routines dont l'action concurrente produit un résultat émergent.

L'intérêt des méthodes de la sémantique procédurale est d'offrir un modèle du système conceptuel qui autorise la "fusion conceptuelle". Dans de nombreux systèmes sémantiques, interpréter un syntagme consiste à agréger les éléments structuraux empruntés aux composants pour constituer une structure plus complexe, censée représenter le sens du syntagme. Aucune opération ne vient diminuer la complexité de la structure ainsi obtenue. Cette monotonie de la construction du sens sera critiquée plus loin (CF. CHAPITRE 8). Sans moyen de remplacer une structure complexe par une structure simple, l'interprétation d'un énoncé complexe, comme la présente phrase, diverge en construisant un échafaudage sans plausibilité théorique et sans intérêt technique. La sémantique procédurale offre un moyen d'échapper à ce problème de monotonie. En exécutant ses procédures, elle interprète un composé comme une nouvelle structure dont la complexité n'est pas corrélée à celle de l'énoncé. Dans l'exemple illustrant le formalisme des trames, le syntagme la voiture que possède la fille de Jean était interprétée par une trame aussi simple que celle associée au prénom Jean. Il existe donc, dans les systèmes procéduraux, un mécanisme de fusion permettant de fondre deux concepts en un nouveau concept, en évitant de conserver une paire conceptuelle.

L'approche de la sémantique procédurale présente ainsi deux atouts importants : (1) contrairement aux systèmes formels, elle évite de renvoyer la question du sens dans une procédure de vérification externe au système ; (2) elle offre un mécanisme de fusion conceptuelle permettant d'attribuer un sens simple aux composés langagiers.

Sur le plan de la théorie cognitive, on peut toutefois lui adresser les reproches liés à ses qualités. L'aspect dynamique et fluide des connexions entre schémas n'offre pas de critère théorique permettant de prédire quelles sont les combinaisons qui auront un sens et celles qui n'en auront pas. L'ensemble est laissé au hasard de ce qui existe à un moment donné dans la mémoire du système. Or, la puissance d'un tel système dépend de la richesse du réseau des concepts. Pour dépasser le cadre limité des applications d'intelligence artificielle, il faut imaginer un réseau de taille colossale pour que le schéma approprié puisse être, dans chaque cas, sélectionné. Certes, toutes les parties de ce réseau ne sont pas supposées actives en même temps, mais la souplesse du système exige que toutes soient accessibles. En d'autres termes, les structures rigides des sections précédentes sont remplacées par un ensemble gigantesque et redondant de concepts reliés entre eux. Lorsqu'il s'agit d'expliquer la faculté sémantique, la sémantique procédurale se contente de postuler l'existence potentielle de tous les schémas requis et des liens qui permettent d'y accéder. Le système des schémas, quelle que soit son utilité pratique, n'offre donc pas de prise à la réfutabilité théorique. Nous reviendrons sur cette critique (CF. CHAPITRE 7).

Les différents modèles évoqués depuis le début du chapitre organisent la construction du sens comme une procédure dirigée par la syntaxe. On peut imaginer, à l'inverse, que ce qui apparaît comme des relations syntaxiques formelles n'est que le reflet visible de processus

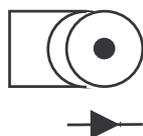
sémantiques sous-jacents. Les modèles de la linguistique cognitive vont dans ce sens, en suggérant que les relations entre les mots n'obéissent qu'à des contraintes liées au sens.

5.5. La représentation schématique du sens

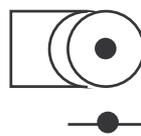
Certaines approches de la sémantique du langage, regroupées sous l'étiquette grammaires cognitives, proposent d'analyser la combinaison conceptuelle au moyen de représentations d'un type particulier, pouvant inclure des schémas graphiques. Ces schémas ont pour rôle de résumer les propriétés essentielles des représentations cognitives associées aux expressions langagières. L'idée générale est que les représentations cognitives des mots prennent leur place dans des structures schématiques associées aux constructions grammaticales pour produire la représentation cognitive correspondant à l'énoncé.

[...] we take a sentence (or a portion of discourse) to evoke in the listener a particular kind of experiential complex, here termed a **cognitive representation** or CR. The grammatical and lexical subsystems in a sentence seem generally to specify different portions of a CR. Together, the grammatical elements of a sentence determine the majority of the *structure* of the CR, while the lexical elements together contribute the majority of its *content*. The grammatical specifications in a sentence, thus, provide a conceptual framework or, imagistically, a skeletal structure of scaffolding for the conceptual material that is lexically specified. (TALMY 2000 [102] p. 21)

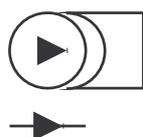
Dans cette vision de l'organisation du système conceptuel, l'élaboration des représentations cognitives allouées aux énoncés repose sur l'opposition entre structure et contenu. La structure est fournie par les éléments grammaticaux. Ces éléments incluent non seulement les entités lexicales de la classe dite fermée, comme les prépositions, les déterminants ou certains adverbess, mais également les marques morphologiques, comme l'inflexion, les cas ou les clitiques, ainsi que des entités de nature syntaxique, représentées par les catégories syntaxiques comme nom ou verbe, et les relations syntaxiques comme spécifieur ou complément. Parmi les éléments susceptibles de structurer la signification des énoncés, il faut également inclure certaines entités lexicales fondamentales qui présentent des propriétés conceptuelles particulières, comme c'est le cas du verbe *keep* dans l'exemple suivant²⁹.



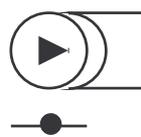
The ball kept rolling
because of the wind blowing on it.



The shed kept standing
despite the gale wind blowing against it.



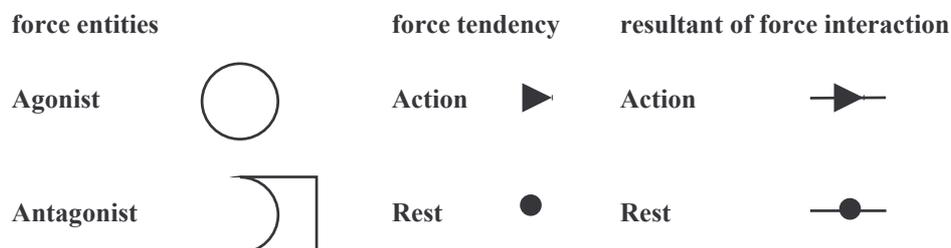
The ball kept rolling
despite the stiff grass.



The log kept lying on the incline
because of the ridge there.

²⁹ D'après (TALMY 2000 [102]).

Cet exemple montre comment le sens des énoncés peut être schématisé. L'objectif est de produire une représentation du sens de l'énoncé, en s'appuyant sur la structure introduite par des mots comme *keep*, *because* et *despite*. L'intérêt de ce type de schématisation est de se limiter à un nombre restreint d'éléments de base, de manière à présenter un caractère prédictif et productif.



La structure conceptuelle associée à la phrase *the ball kept rolling because of the wind blowing on it* prend la forme d'une action résultant de deux forces antagonistes. Dans la structure conceptuelle associée à la phrase *the log kept lying on the incline because of the ridge there*, le résultat des deux forces antagonistes est une immobilisation. Ces représentations schématiques résultent de l'interprétation d'éléments lexicaux de la classe ouverte, comme les verbes *roll* ou *lie*, et de l'interprétation d'éléments de nature grammaticale, comme la conjonction *because* ou la forme *keep + -- ing*.

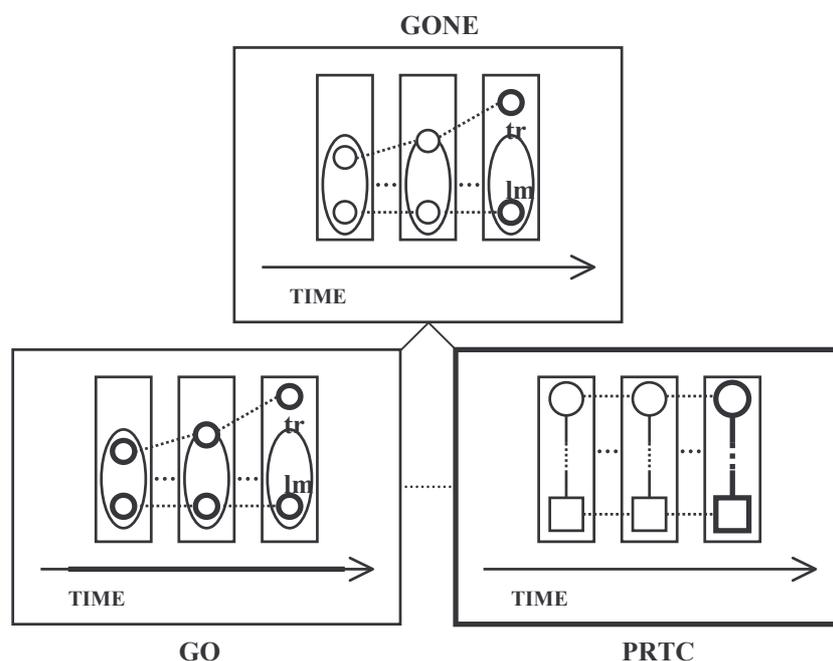
Le pari de ce type d'approche est que la structure conceptuelle contraint suffisamment l'expression langagière pour que l'une puisse être retrouvée à partir de l'autre. En particulier, les différentes langues sont supposées structurées de manière à exprimer sous forme grammaticale les composants conceptuels les plus fondamentaux. Par exemple, de nombreuses langues encodent des indications de nombre dans la morphologie. On peut penser que ces indications, qui se limitent à des distinctions singulier/duel/pluriel, jouent un rôle fondamental dans la structuration conceptuelle. Ce n'est pas le cas, en revanche, des couleurs ou des indications numériques fines comme les objets groupés "par treize" ou "par quarante-sept". Ainsi, l'étude de la grammaire des langues constitue un moyen de comprendre les composants fondamentaux du système conceptuel. Inversement, le modèle proposé pour le système conceptuel est mis à l'épreuve de sa réalisation dans l'expression de la pensée dans telle ou telle langue : la recherche de formalismes schématiques appropriés pour représenter la structure conceptuelle des énoncés a pour objectif de prédire, dans une mesure aussi large que possible, la forme langagière de ces énoncés.

Ce qui distingue les structures schématiques des structures algébriques ou logiques est qu'elles permettent d'exprimer des relations topologiques comme le voisinage ou l'inclusion, ainsi que des relations cinématiques ou dynamiques. Ce ne sont pas des représentations analogiques, puisqu'elles n'encodent ni la magnitude, ni les aspects qualitatifs comme la forme, la couleur, *et cætera*. Parmi toutes les propriétés et les relations que nous pouvons percevoir, celles qui sont retenues pour le système conceptuel sont celles qui découlent de l'observation de la grammaire.

[Cognitive grammar] claims that grammar itself serves an "imagic" function and that much of it has a figurative character. Grammar (like lexicon) embodies conventional **imagery**. By this I mean that it structures a scene in a particular way for purposes of linguistic expression, emphasizing certain facets of it at the expense of others, viewing it from a certain perspective, or construing it in terms of a certain metaphor.

(LANGACKER 1987 [64] p. 39)

La structure conceptuelle d'un énoncé apparaît comme la schématisation d'une scène. Cette schématisation constitue à la fois une simplification et un enrichissement : les propriétés métriques et qualitatives sont laissées de côté, au profit de relations explicitées comme le focus ou le rapport figure - fond, qui se traduisent directement dans la structure grammaticale. De la même manière, la structure grammaticale comporte des indications de point de vue. Grâce aux indications grammaticales liées au temps et à l'aspect, la même situation peut être considérée globalement, par son résultat ou par sa phase d'accomplissement. D'autres aspects de la richesse de la structure conceptuelle se révèlent dans le choix actif - passif, ou encore dans la forme du double accusatif : *he gave Mary a letter*, qui met en valeur le résultat de l'action alors que la forme prépositionnelle : *he gave a letter to Mary*, met le parcours accompli au premier plan. L'exemple suivant illustre le rôle de la forme du participe passé dans l'établissement d'un point de vue³⁰.



Le schéma associé au participe *GONE* diffère de celui du verbe *GO* par l'accent mis sur la partie terminale du processus. La construction grammaticale *GONE = GO + PRTC* se traduit par une combinaison au niveau des structures conceptuelles, dans laquelle la marque de participe provoque un déplacement du focus sur le résultat du processus associé à *GO*. Grâce à ce type de mécanismes de superposition, la combinaison des mots provoque la création de structures qui peuvent être davantage que la simple juxtaposition formelle de leurs composants.

Dans le type d'approche que nous considérons dans cette section, le système conceptuel est décrit à partir d'éléments qui sont supposés structurer notre fonctionnement cognitif. L'hypothèse est que cette structuration est suffisamment contraignante pour diriger l'expression langagière. Autrement dit, les combinaisons de mots reflètent des combinaisons conceptuelles, que l'on cherche à représenter par des combinaisons de schémas. Dans un tel modèle, le système conceptuel contrôle en grande partie l'expression, sans laisser de place à un niveau syntaxique défini.

³⁰ D'après (LANGACKER 1987 [64]).

Within cognitive linguistics, semantics is the primary component (which, in the form of conceptual representations, exists before, both phylogenetically and ontogenetically, syntax is fully developed). The structure of the semantic schemas puts constraints on the possible grammars that can be used to represent those schemas. (GÄRDENFORS 2000 [42] p. 165)

L'idée selon laquelle les mots et les relations entre les mots au sein d'un énoncé véhiculent des relations autres que purement algébriques nous semble fondamentale. En particulier, les mots peuvent exprimer des relations spatiales et topologiques que des formalismes purement symboliques ne peuvent pas représenter de manière naturelle. De ce point de vue, le courant de la linguistique cognitive ouvre un champ de réflexion qui nous semble incontournable. Nous devons cependant nous démarquer sur deux points.

Le premier concerne le statut de la syntaxe à laquelle ces modèles refusent toute autonomie. Les contraintes de proximité dans la liaison entre syntagmes, qui entraînent les phénomènes interprétés comme des mouvements, ou encore les contraintes positionnelles qui limitent la co-référence, semblent difficiles à expliquer par des propriétés purement sémantiques. De même, l'existence d'un jugement syntaxique, qui permet d'accepter des phrases comme bien formées ou de rejeter des phrases comme mal formées indépendamment de la signification qu'on leur accorde, serait mystérieuse si la bonne formation était subordonnée à la seule sémantique. L'ambition qui vise à éliminer du domaine de la syntaxe ce qui relève en réalité de la sémantique est louable. Nous ne souscrivons pas, cependant, à une extrapolation de ce programme qui poserait en principe l'inexistence de toute contrainte purement syntaxique. Plus généralement, l'objectif qui tend à expliquer l'ensemble des phénomènes liés à une langue par des mécanismes cognitifs de haut niveau, comme les schémas graphiques précédents, a ses limites. Il néglige le fait qu'une langue comporte de nombreux aspects non nécessaires résultant de l'habitude et de l'apprentissage par cœur. Par exemple, on dit en français *se rapprocher de la porte*, alors que dans des langues comme l'anglais ou le persan, on exprime l'équivalent de *se rapprocher à la porte*. Une interprétation cognitive consisterait à dire que le francophone mesure la distance qui reste à parcourir à l'aide d'un vecteur partant de la porte : *je suis à deux mètres de la porte*, alors que le persanophone mesurerait la même distance à l'aide d'un vecteur arrivant à la porte. Pourtant, rien n'indique, dans les paraphrases ou dans les inférences, que ces deux personnes utilisent des concepts différents. De même, lorsque le germanophone dit l'équivalent de *j'aide à mon ami* ou *je lui aide*, rien n'indique qu'il emploie un concept distinct de ce que le verbe français exprime. En revanche, si les sujets apprennent par cœur la préposition qu'il convient d'employer avec le verbe *se rapprocher* ou le cas à employer avec le verbe *aider*, l'explication cognitive peut très bien être cantonnée à une portée étymologique. Si tel est le cas, la connaissance qui permet d'employer correctement le verbe est simplement de type phonologique, un peu comme le sont les tables de multiplication. Notre souci, avec ces remarques, est donc de nous démarquer du projet qui consisterait à demander à une théorie des concepts d'expliquer la totalité des phénomènes de langue.

Le deuxième point sur lequel nous pouvons critiquer les approches courantes de la linguistique cognitive est lié à l'origine des structures mentales conférées aux énoncés. Dans le premier exemple cité, la différence des représentations schématiques associés aux phrases *the ball kept rolling* et *the log kept lying* ne doit rien à la grammaire et n'est due qu'à la différence sémantique entre les verbes *roll* et *lie*. Ainsi, comme dans les modèles de sémantique lexicale, les mots, même lorsqu'ils appartiennent à la classe ouverte, comme les verbes *roll* ou *lie*, se voient attribuer une structure conceptuelle afin que leur comportement dans les représentations schématiques puisse être distingué. Le risque est donc de devoir conférer une structure schématique précise à tous les éléments du lexique et non aux seuls éléments grammaticaux. Or une telle démarche est problématique, comme cela sera montré plus loin

(CF. CHAPITRE 7). De plus, même si l'on accepte l'idée de schématiser la totalité du lexique, le résultat n'est pas garanti. La position des forces antagonistes, dans les deux mêmes exemples, est liée à des inférences complexes sur la scène, qui veulent, dans un cas, que le vent contribue au mouvement non spontané de la balle, alors que, dans l'autre cas, le rebord s'oppose au mouvement spontané de la bûche. La méthode ne montre pas comment de telles connaissances peuvent être enfermées dans des représentations schématiques. Ce n'est pas en appliquant un schéma prédéfini qu'il aurait stocké dans l'entrée lexicale des mots *wind*, *ridge*, *roll* ou *ball* que le sujet parvient à savoir que le vent peut contribuer à un mouvement de rotation, alors qu'un rebord s'y oppose. Il semble plutôt que la direction des forces, dans ces exemples, soit le résultat d'un raisonnement, non modélisé, sur la scène, et qu'un tel raisonnement fasse appel à l'expérience de l'individu qui a vécu des situations du même type. Ainsi, l'approche par représentations schématiques ne saurait épuiser la question du sens. Elle incorpore certaines des capacités qui nous permettent de raisonner sur des scènes complexes, sans proposer de limite franche entre ce qui appartient au sens des mots et ce qui relève des capacités inférentielles générales comme le raisonnement spatial. Nous aurons l'occasion de réexaminer certaines propositions de la linguistique cognitive lorsque nous ferons nos propres suggestions sur la nature des représentations associées aux mots (CF. CHAPITRE 9).

Conclusion

Les différentes sections de ce chapitre peuvent être contrastées sur la question du lien entre la combinaison des mots et celle des concepts. Alors que la grammaire générative défend l'idée d'une syntaxe autonome, présentée comme une mécanique chargée d'apparier le sens et les séquences phonologiques, la linguistique cognitive se focalise sur la combinaison des concepts, avec idéalement un contrôle de l'ordonnancement des mots. Entre ces extrêmes, les modèles de la sémantique formelle, de la sémantique lexicale et de la sémantique procédurale proposent diverses solutions pour interfacer leur vision du système conceptuel avec le système syntaxique, allant parfois jusqu'à proposer une correspondance stricte entre les deux niveaux.

Conclusion

Cette deuxième partie de notre travail nous a permis d'étudier différentes caractérisations qui ont pu être faites concernant les concepts. Les trois chapitres de la partie correspondent aux trois interfaces que le système conceptuel entretient avec le reste de la cognition : la perception, le langage et le raisonnement. Il en est résulté des réponses contrastées, souvent incompatibles. Les propriétés qu'un concept doit posséder pour être ancré dans des relations d'évocation perceptives, pour être impliqué dans des inférences, et pour entrer dans des combinaisons dirigées par la grammaire d'une langue ne sont pas les mêmes. Il en résulte une image frustrante, selon laquelle les concepts semblent être des entités cognitives à la fois complexes et insaisissables.

La prochaine partie va se concentrer sur la caractérisation des concepts en tant que représentations internes du système conceptuel. Ce système conceptuel possède ses propres contraintes, liées notamment à l'explication des phénomènes de systématisme. Nous aurons l'occasion de constater que les solutions évoquées dans ce qui précède sont non seulement difficiles à concilier, mais présentent des défauts, parfois rédhibitoires, sur le plan de la modélisation cognitive.

