

Jean-Pierre BARTHELEMY*, *****, Michel DE GLAS**, *****,
Jean-Pierre DESCLES***, *****, et Jean PETITOT****, *****

Logique et dynamique de la cognition

Cet article trace les grandes lignes d'un programme de recherche dont l'ambition est de jeter les bases d'une intégration des paradigmes logico-symbolique et connexionniste en sciences cognitives, en approfondissant les méthodes et les modèles mathématiques qui – de la logique formelle à la géométrie différentielle – seraient susceptibles d'éliminer toute modélisation dualiste et/ou réductionniste de la cognition. Ce projet se fonde sur l'idée que les dimensions logico-symbolique et connexionniste de la cognition sont les deux versants d'un même paradigme.

Mots-clés : paradigmes symbolique et sub-symbolique, intentionnalité, catégorisation, morphodynamique et logique.

Logic and Dynamic of Cognition. This paper provides the reader with the essential features of a program of research the aim of which is to lay the foundations of the integration of symbolic and connectionist paradigms in cognitive science. It is argued that this may be done by deepening convenient mathematical methods and models – from formal logic to differential geometry – which may be used as alternatives to dualistic or reductionist models of cognition. This project is founded upon the idea that the symbolic and the connectionist dimensions of cognition are two facets of the same paradigm.

Key Words : Symbolic and Subsymbolic Paradigms, Intentionality, Categorization, Morphodynamics and Logic.

* LIASC, ENST Bretagne, BP 832, 29285 Brest Cedex.

** LAFORIA, CNRS - UPMC, 4 place Jussieu, 75252 Paris Cedex 05.

*** CAMS, CNRS - EHESS - Paris Sorbonne, 54 & 96 boulevard Raspail, 75006 Paris.

**** CREA, CNRS - École Polytechnique, 1, rue Descartes, 5005 Paris

***** GDR « Sciences Cognitives de Paris ».

INTRODUCTION

Dans leur programme de naturalisation des processus mentaux, les sciences cognitives sont confrontées à des difficultés de modélisation dues au fait qu'elles doivent articuler trois types d'objectivité : objectivité physique du monde naturel, objectivité logico-symbolique des langages, objectivité computationnelle des traitements effectifs. Les sciences cognitives doivent unifier ces trois types hétérogènes d'objectivité ; une telle unification soulève en particulier le problème de la compréhension des rapports entre *le physique et le symbolique*.

Pour mener à bien une telle entreprise, le paradigme cognitiviste a admis, au moins implicitement, des dualismes tels que le physique et le symbolique ou le neuronal et le mental et un réductionnisme du cognitif au symbolique. Le paradigme cognitiviste se fonde sur l'adhésion à la métaphore de l'ordinateur et du traitement d'informations symboliques, c'est-à-dire sur l'hypothèse que les rôles causaux sont des rôles computationnels opérant sur des symboles d'un langage interne : le cognitivisme est donc fonctionnaliste, computationnel et représentationnel. Ce *fonctionnalisme computo-représentationnel* résulte d'une tension entre deux objectifs, *a priori*, contradictoires : éviter toute forme de réductionnisme du mental au neuronal ; obéir à une version acceptable du matérialisme naturaliste en vertu duquel les concepts mentaux sont explicables en termes des concepts des sciences physiques et naturelles.

En postulant que les représentations sont indépendantes, dans leur structure formelle et leur contenu informationnel, de leur implantation dans le substrat physique, le paradigme cognitiviste souscrit à une forme de *dualisme entre le physique et le symbolique* : l'information externe, conçue de façon physicaliste, est, *a priori*, sans signification pour le système cognitif ; elle est soumise à une transduction par des modules périphériques qui la convertit en information externe computationnellement significative. Deux grands problèmes sont, dès lors, laissés sans réponse : du côté du sujet, les problèmes de l'intentionnalité et de *l'intensionnalité* sont évacués ; du côté du monde externe, le problème de la manifestation qualitative et morphologique des phénomènes et de la conscience que nous en avons est occulté. Par ailleurs, en postulant l'existence de représentations mentales neurobiologiquement implantées dans des états cérébraux et la nature symbolique de ces représentations mentales, le paradigme classique se fonde sur un *réductionnisme du cognitif au symbolique*.

Le paradigme symbolique s'est trouvé concurrencé par le paradigme connexionniste (et en particulier par l'une de ses variantes, dite sub-

symbolique) selon lequel les structures symboliques émergent des dynamiques sous-jacentes, grâce à des phénomènes collectifs, coopératifs et compétitifs, d'auto-organisation. Ainsi, l'objet d'étude devient-il une sorte de « thermodynamique neuronale du mental ». L'une des principales difficultés que l'on rencontre dans le dépassement de telles positions est d'arriver à une bonne compréhension des relations entre les bas-niveaux micro-physiques et les hauts niveaux macro-cognitifs, apparemment descriptibles symboliquement. C'est l'étude de ces relations qui constitue le défi majeur de l'ambition naturaliste.

Les problèmes posés par de telles alternatives théoriques sont, selon nous, *avant tout* mathématiques. De récentes recherches interdisciplinaires nous ont permis d'identifier des champs des mathématiques qui nous semblent particulièrement pertinents pour une meilleure modélisation des processus cognitifs. Il apparaît donc indispensable de développer les aspects techniques par une telle modélisation. Outre son intérêt intrinsèque, la mathématisation doit permettre d'établir des liens rigoureux entre les outils théoriques utilisés, de la logique formelle à la géométrie différentielle, et, ainsi, entre les approches qu'ils fondent.

Le programme de recherche qui se dessine ainsi vise, plus précisément, à approfondir les méthodes et les modèles mathématiques disponibles de façon à dépasser le dualisme entre le physique et le symbolique en proposant une intégration théorique des paradigmes symboliques et connexionnistes. Une voie de recherche consiste à introduire un niveau médiateur entre le physique et le symbolique : le niveau morphologique. Pour ce faire, il faut élaborer, d'une part, une théorie constructive des systèmes dynamiques, c'est-à-dire une morphodynamique constructive, et, d'autre part, une logique intensionnelle. Ce projet se fonde sur l'idée qu'il est possible d'appréhender les dimensions respectivement logico-symbolique et morphodynamique de la cognition comme les deux versants d'un même paradigme.

Notre projet présente trois volets :

(i) Intensionnalité : Recherche d'une « logique du sens », d'une logique constructive dont la contrepartie « géométrique » permet de pallier les limites de la topologie pour construire une « théorie des lieux » intégrable dans une théorie des représentations cognitives ; étude des liens entre logique intensionnelle (et théorie des topoi) et lambda-calcul typé ; théorie constructive des types appliquée à la description des langues et à l'analyse des processus langagiers.

(ii) Catégorisation : Dépassement du clivage extension-intension grâce aux théories mathématiques des catégories et des faisceaux, qui permettent de considérer des objets émergents comme limites de diagrammes commutatifs, de reposer les problèmes de typicalité en des termes nouveaux, de construire des structures globales à partir de structures locales.

(iii) Morphodynamique et logiques : Complémentarité entre une « logique du sens » et une « physique du sens ». Modélisation dynamique des processus de genèse et de régulation des catégorisations ; émergence et (auto)-organisation de structures ; modélisation dynamique et constructive de la constituance et de la compositionnalité syntaxique ; ancrage de la syntaxe et de la sémantique dans la perception et l'action ; modélisation du schématisme iconique des grammaires cognitives.

1. SCIENCES COGNITIVES : UN ETAT DES LIEUX

1.1. Le paradigme classique

Le paradigme classique en sciences cognitives, le *paradigme cognitiviste*, se fonde sur l'adhésion à la métaphore de l'ordinateur et du traitement d'informations symboliques, c'est-à-dire sur l'hypothèse que les rôles causaux des états mentaux sont computationnels et opèrent sur des symboles d'un système de représentation interne. Le cognitivisme repose sur trois principes que l'on peut, au prix d'une simplification extrême, résumer ainsi (Ander, 1988, 1992) :

(i) Le complexe esprit/cerveau est susceptible d'une *double description*, physique (la physique n'intervenant, en fait, que par le biais des neurosciences) et informationnelle ou fonctionnelle. Ces deux niveaux de description sont indépendants l'un de l'autre et la relation qui s'établit entre eux est à l'image de celle qui lie un ordinateur en tant que système physique à un ordinateur en tant que système de traitement de l'information.

(ii) Sur le plan informationnel, le système cognitif humain est caractérisé par des *états internes* et par les *processus* qui font passer le système d'un état à un autre. Ces états sont *représentationnels*, c'est-à-dire qu'ils sont dotés d'un contenu sémantique renvoyant à des entités externes : ils sont dits sémantiquement évaluables.

(iii) Les *représentations* internes sont des formules d'un langage interne (par exemple, le mentalais chez Fodor) proche du langage formel de la logique. Les processus sont ceux que la logique appelle

effectifs : ils sont réductibles à des opérations primitives dont l'exécution par une machine n'exige aucune interprétation. Ils s'identifient donc à l'ensemble des procédures descriptibles et réalisables en un nombre fini d'étapes élémentaires.

Le cognitivisme est donc fonctionnaliste, computationnel et représentationnel. Ce *fonctionnalisme computo-représentationnel* résulte du double souci d'éviter toute forme de réductionnisme du mental au neuronal et d'obéir à une version acceptable du matérialisme naturaliste qui exige que les concepts mentaux soient explicables en termes des concepts des sciences physiques ou des sciences naturelles.

Ecartelé entre ces deux exigences contradictoires, le cognitivisme est concurrencé par des approches rivales, qui se rangent pour l'essentiel sous la bannière du connexionnisme. Toutefois, si le cognitivisme ne fournit aucune théorie satisfaisante de l'« esprit », force est de constater que ses rivaux n'ont pas permis d'envisager ce que pourrait être une telle théorie.

Le fonctionnalisme s'efforce de concilier les avantages du behaviorisme -- à la différence duquel il admet l'existence d'états internes pouvant être causes des réponses comportementales et effets de stimuli -- et du physicalisme -- contrairement auquel il définit les états mentaux en termes relationnels -- tout en tentant d'en contourner les apories : un état mental est individualisé par les *relations* qu'il entretient avec les stimuli, les réponses comportementales et d'autres états mentaux.

Le fonctionnalisme prend sa source dans le behaviorisme. Selon celui-ci, l'organisme est une boîte noire dont le comportement, la sortie, est fonction des stimuli, des entrées, qu'il reçoit et de son état. L'objection courante adressée à cette thèse est, d'une part, qu'il est nécessaire de postuler l'existence d'états mentaux intermédiaires venant s'ajouter à ceux que l'on est censé analyser et, d'autre part et surtout, que le rôle causal de l'organisme vu comme une boîte noire ne peut être mis en évidence. En effet, selon la thèse behavioriste, il faut concevoir les termes mentaux comme désignant, non pas des objets, mais des propriétés dispositionnelles du comportement, lesquelles s'expriment par le truchement d'énoncés conditionnels contrefactuels. Il est ainsi possible de définir la signification des termes mentaux grâce à des énoncés portant sur le comportement. Une telle position est intenable dans la mesure où, pour spécifier lesdits comportements, on est contraint de recourir à d'autres états mentaux et, partant, de recourir à des définitions circulaires. Par ailleurs, les états mentaux, en tant que

propriétés dispositionnelles du comportement, ne sauraient être cause du comportement.

La thèse fonctionnaliste conserve l'idée de l'organisme vu comme une boîte noire, comme un système, mais tente de contourner les apories du behaviorisme, en introduisant dans le système des sous-systèmes ayant des comportements propres et investis de fonctions spécifiques. Le comportement de chacun de ces sous-systèmes est fonction de l'état dans lequel il se trouve et des stimuli qui lui parviennent de son environnement ou des autres sous-systèmes. A la différence du behaviorisme, le fonctionnalisme admet que les états mentaux ne peuvent pas être définis par la seule mention des stimuli extérieurs et des sorties comportementales : un état mental est individualisé par son rôle causal vis-à-vis des stimuli extérieurs, de sorties comportementales, mais également des autres états mentaux du dispositif cognitif auquel il est attribué. L'ensemble de ces états mentaux joue un rôle causal déterminé. Ils sont *fonctions* les uns des autres et ont une *fonction* globale. Ainsi un état mental occupe-t-il un rôle causal déterminé par l'ensemble de ses relations aux autres états mentaux.

Pour éviter l'écueil du physicalisme, le fonctionnalisme doit, d'un côté, proposer une méthode d'individualisation des états mentaux ne faisant pas référence aux descriptions physiques de ces états mentaux et, d'un autre côté, éviter qu'un état mental soit susceptible d'une caractérisation incompatible avec le physicalisme. Ce double objectif conduit le fonctionnaliste à souscrire à la thèse de Turing.

La thèse de Turing, qui, par définition, n'est pas susceptible d'être prouvée, est que toute procédure effective est modélisable par une machine de Turing (Turing, 1950) ; en termes plus précis, pour toute procédure, manipulation ou fonction d'entiers que l'esprit humain peut exécuter ou calculer concrètement, effectivement, il existe une machine de Turing capable d'exécuter cette procédure ou de calculer cette fonction. Le fonctionnalisme computationnel de la thèse de Turing se place ainsi au cœur du cognitivisme : il est possible, selon la thèse cognitiviste, de modéliser les traits fonctionnels d'un système par les propriétés d'une machine réelle, de telle sorte que les fonctions de transition de la machine au niveau physique coïncident avec les règles régissant le fonctionnement du système. Le fonctionnalisme computationnel «turingien» exploite l'idée que le vocabulaire utilisé pour la description abstraite d'une machine de Turing est celui qui est nécessaire à la description des états mentaux. Il promeut, en outre deux idées fondamentales : le caractère relationnel des états mentaux, la

définition abstraite d'une machine de Turing étant relationnelle ; l'inutilité, pour caractériser une machine de Turing, de faire référence aux propriétés physiques des dispositifs dans laquelle elle se réalise. Ainsi, d'un côté, toute fonction, dès lors qu'elle est exécutable par une machine de Turing, peut être effectuée par un mécanisme physique (compatibilité avec le physicalisme) ; d'un autre côté, la machine de Turing exécutant cette fonction, machine susceptible de multiples réalisations physiques, peut être caractérisée sans référence aux propriétés de ces réalisations (évitement de l'écueil physicaliste).

Ainsi le terme de « fonction » est-il pris, dans l'énoncé de la thèse fonctionnaliste, dans deux acceptions différentes. D'une part, dire que les états mentaux sont fonctions les uns des autres revient à utiliser le terme de « fonction » dans son acception mathématique classique, c'est-à-dire au sens de fonction calculable par un algorithme : les états mentaux sont ainsi assimilés aux états d'une machine de Turing (ou de tout autre dispositif computationnel), qui calcule, de façon effective, un algorithme. D'autre part, la fonction globale, dont l'ensemble des états mentaux est censé, selon la thèse fonctionnaliste, être investi doit être comprise au sens téléologique de la propriété de ce que réalise un système quand il a un but : c'est sur ce dernier point que le recours à la métaphore de l'ordinateur semble s'imposer au fonctionnaliste. En effet, bien que la fonction, au sens téléologique, d'un système présuppose une réalisation physique particulière de ce système, le fonctionnalisme n'accorde aucun rôle à cette dernière. Le fonctionnaliste est indifférent au fait qu'un programme d'ordinateur soit implanté dans une machine, équipée de composants mécaniques ou électroniques ; seul importe le logiciel qui réalise une certaine fonction. Les propriétés fonctionnelles, d'un organisme ou d'une machine, sont distinctes des propriétés physiques, mais les états fonctionnels sont réalisés dans des propriétés physiques, matérielles, de l'organisme ou de la machine.

Le fonctionnalisme est donc une forme de matérialisme se refusant au réductionnisme physicaliste : le fonctionnalisme n'entreprend pas, en effet, de réduire les propriétés mentales à leur réalisations physiques. Il admet, au contraire, le *dualisme* des propriétés fonctionnelles et des propriétés physiques : un programme, dont la fonction est d'effectuer un « calcul » peut être implémenté sur des machines ayant des propriétés physiques différentes. Le fonctionnaliste nie que les concepts neurobiologiques soient appropriés à la description des phénomènes psychologiques, que les concepts électroniques soient appropriés à la description de phénomènes informatiques, bien que les descriptions psychologiques ou

informatiques puissent être extensionnellement équivalentes, respectivement, aux descriptions neurobiologiques ou électroniques.

L'utilisation de la métaphore de l'ordinateur, à laquelle recourt le fonctionnalisme ne va pas jusqu'à munir les états mentaux d'un pouvoir représentationnel : les états mentaux, en tant qu'états fonctionnels, sont suffisants pour déterminer leur contenu, sans avoir à véhiculer une représentation spécifique exprimant ce contenu. La version *computo-représentationnelle* du fonctionnalisme, au prix d'une adhésion littérale à la métaphore de l'ordinateur, va plus loin : les rôles causaux des états mentaux sont *computationnels*, en ce sens que, comme pour l'ordinateur, ils opèrent sur des symboles d'un langage interne. L'esprit manipule des symboles d'un langage de la pensée, lequel est réalisé, implanté, dans la neurobiologie du cerveau, dans des mécanismes physiques. De plus, ce langage, proche du langage de la logique, se caractérise par une syntaxe et une sémantique sur laquelle s'articulent des *représentations*. Ainsi la thèse du fonctionnalisme computo-représentationnel stipule-t-elle un niveau de représentation intermédiaire entre le niveau fonctionnel et le niveau physique : le niveau *symbolique*.

La dimension représentationnelle du cognitivisme rend ce dernier solidaire de la psychologie ordinaire. Théorie proto-scientifique et empirique des phénomènes mentaux devant à ce titre être évaluée, c'est-à-dire défendue ou rejetée, pour les uns, schème normatif et *a priori* qu'il ne peut être question d'étudier à la manière d'une théorie scientifique en ce qu'elle ne produit pas d'explications causales ni de lois descriptives, pour les autres, la psychologie ordinaire désigne l'ensemble des principes et des concepts mentaux au moyen desquels nous nous comprenons nous-mêmes et expliquons les comportements de nos semblables. Les cognitivistes sont parmi les plus ardents défenseurs de la psychologie ordinaire. Selon Fodor, par exemple, ses trois traits fondamentaux sont (i) qu'elle attribue aux individus des attitudes propositionnelles : croyances, souhaits, désirs ... qui ont des contenus intentionnels ; (ii) qu'elle suppose que ces contenus intentionnels ont des pouvoirs causaux sur le comportement ; (iii) qu'elle forme, sur la base de ces attributions de contenus mentaux des généralisations généralement vraies pouvant prendre la forme de lois capables de prédire le comportement. Que ces généralisations ne valent que *ceteris paribus*, c'est-à-dire soient sujettes à des exceptions, n'empêche pas la psychologie ordinaire d'être, selon Fodor, un schème explicatif et prédictif du comportement remarquablement fiable. Ce schème constitue une théorie proto-scientifique du mental dans la mesure où l'on postule, à la suite de Fodor, l'existence de

représentations mentales et de relations fonctionnelles entre les états qui véhiculent ces représentations : les sciences cognitives ne feraient que rendre explicite ce qui reste implicite dans les explications de la psychologie ordinaire, en postulant l'existence, chez les individus doués d'états cognitifs, de représentations internes symboliques. La psychologie cognitive attribue un état mental (croyance, désir, ...) à un organisme sur la base d'une représentation linguistique (croire *que*, désirer *que*, ...). Les sciences cognitives explicitent cette attribution en postulant l'existence d'une représentation symbolique, d'une phrase du langage de la pensée de l'organisme. Cette « théorie représentationnelle de l'esprit » justifie *a posteriori* la psychologie ordinaire, laquelle offre, en retour, au cognitivisme une justification empirique.

La vulgate fodorienne met en lumière les points de convergence entre le cognitivisme et l'empirisme logique. L'empirisme logique et le cognitivisme partagent certaines hypothèses fondamentales : la distinction entre théorie et observation, le rôle de la syntaxe, la distinction entre syntaxe et sémantique. La remarque de Fodor selon laquelle « les sciences cognitives sont de la philosophie des sciences en petits caractères » semble suggérer que les sciences cognitives sont confrontées aux mêmes problèmes que la philosophie des sciences. Fodor soutient d'ailleurs, dans son examen des rapports entre les deux disciplines, que la psychologie cognitive actuelle légitime le bien-fondé d'une distinction, qui a joué un rôle crucial dans l'empirisme logique, entre observation et théorie ; distinction qui fonde, chez Fodor, son postulat de la modularité de l'esprit (Fodor, 1983) .

Fodor est, en fait, un épistémologue de la psychologie cognitive qui aurait chaussé les lunettes de Carnap (Carnap, 1947). Quand celui-ci met l'accent sur le pouvoir expressif du langage et sa capacité propositionnelle, celui-là, dans le même esprit, met en rapport la capacité cognitive individuelle avec la capacité linguistique et le discours produit par l'individu. En outre, tout comme Carnap, Fodor élabore une distinction entre le langage ordinaire et un langage sous-jacent : le langage de la pensée ou mentalais, qu'il présente comme inhérent à la logique-même des notions de représentation mentale et de processus cognitif. Enfin, en écho au projet logico-empirique selon lequel la construction d'un métalangage coïncide avec la construction logique du monde connu par la science, le projet fodorien postule que l'étude du langage de la pensée est l'étude du monde psychologique du système cognitif, c'est-à-dire de la réalité psychologique du système cognitif, monde défini par le langage de la pensée et sous-jacent à l'expérience phénoménale du sujet. Fodor adhère ainsi au corrolaire du

constructivisme carnapien : le métalangage postulé dessine les limites de ce qu'un individu peut connaître. L'hypothèse d'un métalangage permet à Fodor de considérer les représentations mentales comme des énoncés du langage de la pensée et ainsi de fonder sa conception syntaxique de la représentation mentale, noyau dur du cognitivisme : le comportement des symboles est réductible à leurs propriétés syntaxiques et aux règles du système.

Le logicisme et le mentalisme qui conduisent Fodor à postuler l'existence d'un langage indépendant, distinct du langage naturel mais dont celui-ci serait le résultat par calcul, rejoint, sur bien des points, la thèse selon laquelle le langage ne serait que la représentation de la pensée. Assez proche de cette thèse, Chomsky (Chomsky, 1964-1988) revendique, par ailleurs, l'existence d'une faculté rationnelle qui serait spécifique à l'homme et qui ne pourrait être qu'innée.

La conception chomskienne du langage conduit à identifier la compétence syntaxique aux automatismes et à privilégier la dimension calculatoire qui conceptualise les langues naturelles comme de simples langages formels. Une telle conception affiche de nombreuses limites intrinsèques. L'hypothèse centrale de la grammaire générative est que le comportement linguistique humain s'explique par la productivité d'un nombre fini de règles qui décrivent la compétence linguistique. C'est le fait de posséder un ensemble fini de règles (appliquées récursivement), susceptibles d'engendrer une infinité de phrases, qui expliquerait notre capacité à produire (et à comprendre) des phrases. Une telle conception réduit la créativité à la productivité. Comme le fait remarquer Auroux, « mettez un ordinateur dans une cave, programmez-le pour construire la suite infinie des nombres entiers, revenez dans mille milliards d'années (ou plus tard, si vous en avez le loisir), il n'aura toujours pas produit un nombre irrationnel » (Auroux, 1994).

Il suffit de montrer que le contenu des syntaxes naturelles n'est pas celui de la récursivité pour voir que l'interprétation de la créativité par la récursivité fait violence à des propriétés élémentaires du langage naturel. La conception chomskienne du langage appliquée à l'analyse des langues naturelles des hypothèses qui ne sont vérifiées que par des langages formels dont la générativité opère sans contrainte. La plus grande partie des concepts logiques mis en jeu dans l'analyse des rapports entre syntaxe et sémantique ne prennent sens que sous l'hypothèse d'un ensemble infini d'énoncés et des théories considérées. Or la générativité propre aux langues naturelles est terriblement contrainte : celles-ci se manifestent sous forme de séquences finies, de plus relativement courtes. Une « bonne » théorie syntaxique doit se

fixer comme objectif de traduire mathématiquement la notion de générativité contrainte et de courte portée.

L'idée fondamentale de la linguistique chomskienne est donc que les énoncés d'une langue naturelle sont dérivables par application récursive d'un nombre fini de règles et sont, par suite, les énoncés acceptables par un automate. Une deuxième conséquence d'une telle conception est l'indépendance de la signification des mots par rapport à l'activité du locuteur. Cet argument établit la liaison de la « créativité » chomskienne avec l'innéisme. Le débat Chomsky-Piaget (Piatelli, 1979) montre clairement que l'innéisme chomskien découle de l'hypothèse de l'existence d'universaux formels sur lesquels le locuteur, bien qu'il n'en ait jamais fait l'apprentissage, ne commet jamais d'erreurs et dont il n'existe aucune caractérisation dans l'univers des grammaires formelles. Le langage étant réduit aux automatismes de la compétence et aux automates qui en assurent la description, la part de la compétence linguistique, dont les automates ne peuvent rendre compte, se trouve rejeté du côté de l'inné.

En résumé, dans la perspective cognitiviste, dont Fodor et Pylyshyn (Fodor, 1975, 1983, 1990, Fodor et Pylyshyn, 1988, Pylyshyn, 1984) sont les principaux artisans, (i) les états mentaux sont des états fonctionnels de l'organisme, définis par l'intermédiaire de leur relation fonctionnelle et causale à des stimuli extérieurs et à d'autres états mentaux ; (ii) ce sont les symboles du langage interne, inscrit dans le cerveau, qui permettent de rattacher les propriétés sémantiques des représentations à des états matériels de la neurophysiologie du cerveau ; les propriétés causales des symboles sont reliées à leurs propriétés sémantiques grâce à leur structure syntaxique ; (iii) les énoncés du langage interne sont soumis, à l'aide de règles, à des transformations calculables.

1.2 Les paradigmes rivaux

Les objections dont le paradigme classique est l'objet sont nombreuses mais s'attachent toutes à montrer que le fonctionnalisme est impuissant à rendre compte de certaines propriétés des états mentaux.

En postulant l'existence de représentations mentales neurobiologiquement implantées dans des états cérébraux et la nature symbolique de ces représentations mentales, le paradigme classique se fonde sur un *réductionnisme du cognitif au symbolique*. Le projet cognitiviste d'une caractérisation des états mentaux en termes computationnels renoue avec l'idée hobbsienne selon laquelle « penser,

c'est calculer » et présuppose donc la validité de l'hypothèse de réductibilité du cognitif à un calcul symbolique, les actes de pensée étant assimilables à une manipulation de symboles selon des règles. On passera sous silence cette forme « achevée » de réductionnisme, dont les représentants, sous l'influence des théories néo-darwiniennes de l'évolution des espèces, cèdent à la tentation de construire un paradigme évolutionniste unique permettant de considérer les phénomènes biologiques, ontogénétiques et phylogénétiques comme s'il s'agissait de processus cognitifs : à force de prendre les métaphores informatiques au pied de la lettre, les organismes, après avoir été vus comme des ordinateurs, sont maintenant vus comme des systèmes experts, l'évolution comme une simple acquisition de connaissances.

Le cognitivisme s'oppose, sur ce point, à la thèse *éliminativiste* qui, elle, procède d'un *réductionnisme du cognitif au biologique*, du mental au cérébral, voire, dans sa version la plus radicale, à l'*élimination du mental* au seul profit du cérébral. L'éliminativisme postule, dans la tradition du réductionnisme classique, qu'il n'y a pas d'états intentionnels, mais seulement des états neuronaux, c'est-à-dire des configurations matérielles dans le cerveau. Les processus cognitifs, dont seules les neurosciences seraient habilitées à faire l'étude, se ramèneraient à des « cartes topographiques » et à la coordination sensori-motrice dont les espaces d'état représentent des données extérieures sans qu'intervienne une quelconque représentation symbolique.

La thèse éliminativiste, dont P.M. Churchland et P.S. Churchland (P.M. Churchland, 1979, 1984, P.S. Churchland, 1986) se sont fait, après Feyerabend et Rorty (Rorty, 1972), les ardents avocats est étayée par une critique dévastatrice de la psychologie ordinaire, visant à priver les fonctionnalistes de leur anti-réductionnisme (du cognitif au neuronal) : puisqu'il est impossible de déceler dans la cognition humaine rien qui puisse correspondre aux états mentaux postulés par la psychologie ordinaire, qu'il n'y a pas dans la cognition d'analogues des états discrets et structurés assimilables aux phrases que la psychologie ordinaire présuppose, qu'il n'y a pas d'entités telles que les croyances, ces états mentaux n'existent pas et la psychologie ordinaire est un mythe.

Le partisan de la thèse de l'identité matérialiste entre un état mental et un état cérébral s'expose à une double tentation : le réductionnisme et l'éliminativisme ; soit les états (et les processus) mentaux coïncident avec les états (et les processus) cérébraux, soit les états

mentaux n'existent pas. Le matérialisme réductionniste de la première thèse soulève de nombreuses objections : affirmer que le cerveau, ou l'une de ses aires, calcule est aussi absurde que de prétendre qu'un télescope calcule la trajectoire des rayons de lumière qui le traversent (Hart, 1986) ou que les planètes calculent leurs trajectoires, sauf à considérer que, un calcul n'étant pas nécessairement conscient, les planètes «calculent» leurs trajectoires au sens où leurs mouvements sont solutions d'équations différentielles. La seconde thèse est une forme de matérialisme éliminatif qui postule, donc, que les états mentaux n'ont pas d'existence objective ; que les concepts de la psychologie ordinaire n'ont pas à être réduits aux concepts de la neurobiologie mais à être simplement éliminés. Du point de vue de la thèse éliminativiste, prétendre que ce que, depuis Russell, on appelle attitudes propositionnelles (les pensées, les croyances, les désirs, ...) font référence à des états ou des processus mentaux inobservables exerçant une influence causale sur les conduites humaines relève d'une erreur ontologique : croire que les attitudes propositionnelles puissent faire référence à des états mentaux inobservables est, pour l'éliminativiste, une croyance ontologique fautive au même titre que la croyance en l'existence des fantômes ou du Père Noël. Au contraire, pour l'éliminativiste, une authentique science de l'esprit consisterait en la possibilité d'expliquer causalement l'ensemble de nos comportements en termes du vocabulaire de la neurobiologie, *in fine* de la physique. Pour l'éliminativisme, aucune correspondance n'est possible entre la psychologie populaire et une authentique science de la cognition. Celle-ci, que les Churchland appellent de leurs vœux, doit être fidèle à la biologie, en particulier aux caractéristiques évolutives des organismes dont la vision computationnelle à la Fodor-Pylyshyn ne peut rendre compte : la cognition ne consiste pas en des représentations internes computationnellement significatives ; la computation ne s'effectue pas sur des entités symboliques. Ainsi les modèles connexionnistes offrent-ils une solution de rechange aux modèles fournis par le paradigme cognitiviste : ils représentent la cognition en termes de réseaux parallèles et distribués sur des entités définies à un niveau sub-symbolique ; les processus computationnels ne font intervenir ni codage symbolique ni manipulation de symboles. L'information véhiculée par ces systèmes n'est pas programmée par le réseau mais «découverte» par celui-ci. Les représentations n'opèrent pas sur des symboles d'un langage de la pensée mais «émergent» du système.

Le repli éliminativiste, dans sa conception physicaliste, exprimée ici, ou dans sa conception cognitive qui postule l'existence d'états

cognitifs infra-doxastiques entre les états cérébraux et les états mentaux (Stich, 1978) est une tentation constante : les modèles de bas niveau, physiques, biologiques, semblent toujours plus « scientifiques » que les modèles de haut niveau, lesquels sont toujours guettés par l'épiphiénoménalisme.

Toutefois, la thèse éliminativiste est, sur de nombreux points, intenable. Le fait qu'on ne trouve pas dans un organisme, inscrits dans le cerveau, des croyances et des désirs n'implique pas que l'organisme n'ait ni croyances ni désirs. Rejeter le fonctionnalisme radical à la Fodor-Pyllyshyn, qui voit dans l'existence de régularités formelles entre le niveau sémantique des attitudes propositionnelles et le niveau de la causalité physique la preuve de la capacité prédictive de la psychologie ordinaire, ne saurait valablement conduire à nier l'existence des dites attitudes propositionnelles.

L'éliminativiste nous invite à choisir entre le point de vue de la psychologie et celui de la neurobiologie, tenus pour contradictoires comme l'étaient la « théorie » pré-lavoisienne du phlogistique et la théorie de l'oxygène. La neurobiologie serait à la psychologie ordinaire, ramenée au rang de théorie empirique du mental, ce que la physique est au sens commun. La physique démontrant, par ailleurs, la fausseté du sens commun, entre le concept scientifique et le concept ordinaire associés à un objet (par exemple le concept de table défini par la physique atomique et le concept ordinaire de table), il faudrait choisir le premier et renoncer au second. Tout d'abord, faire de la psychologie ordinaire une théorie empirique du mental ne résiste pas à l'examen. Elle consiste, en fait, en un ensemble de règles normatives de rationalité par l'intermédiaire desquelles nous interprétons nos comportements et ceux de nos semblables, ces comportements étant supposés obéir (Dennett, 1969, 1987, Jacob, 1992) à certaines conditions d'optimalité. Nos croyances, désirs, ..., dotés d'un contenu intentionnel, s'appliquent à des agents idéalement rationnels. Cette « posture intentionnelle » nous permet de prédire qu'un système, capable d'un comportement intelligent, réalisera tel ou tel comportement, quitte à réviser notre prédiction à la baisse. Ensuite, qu'une croyance ne soit pas justiciable d'une description physico-chimique ne permet pas de conclure, comme nous y invite l'éliminativiste, à l'inexistence pure et simple des croyances. Celles-ci donnent lieu soit à des propositions vraies, soit à des propositions fausses, soit à des propositions dénuées de sens ; la psychologie ordinaire confère donc aux croyances un rôle propositionnel, c'est-à-dire un statut d'entités abstraites. Plutôt que de les comparer aux concepts de la chimie pré-lavoisienne, il serait plus convenable de les

comparer, à la suite de Dennett, à des concepts abstraits tels que le centre de gravité : il est tout aussi absurde de se poser la question de la composition biologique d'une croyance que de se demander à quel atome identifier le centre de gravité d'une table.

En postulant qu'il est possible de découpler les problèmes de matériel de ceux de logiciel et que les représentations mentales symboliques sont indépendantes, dans leur structure formelle et leur contenu informationnel, de leur implémentation dans leur substrat physique, le paradigme classique semble souscrire, au moins implicitement, à un *dualisme* entre le physique et le symbolique. Celui-ci renvoie au dualisme entre matière et forme qui se trouve à l'origine du postulat structuraliste du primat ontologique de la forme sur la matière : le structuralisme a toujours vécu sur l'idée que les structures émanent d'une forme, laquelle s'implante dans une matière amorphe et qui lui est étrangère. Ce point de vue est, toutefois, contesté par les fonctionnalistes : selon eux, les états fonctionnels d'une machine (et d'un organisme) sont réalisés dans des propriétés physiques de la machine ; une fonction, définie à un niveau supérieur, peut être décomposée, itérativement, en sous-fonctions, sous-sous-fonctions, ..., jusqu'à atteindre le niveau physique. Par ailleurs, faire l'hypothèse que le matériel, électronique des ordinateurs ou biochimique des cerveaux, n'a aucune pertinence dans l'étude de leur capacités fonctionnelles ne peut être taxé de dualisme qu'en assimilant une distinction de niveaux d'organisation à une distinction de substances.

Toujours est-il que le fonctionnalisme n'est pas incompatible avec le dualisme : l'information externe, conçue de façon physicaliste, est *a priori* sans signification pour le système cognitif ; elle est soumise à une transduction par des modules périphériques qui la convertit en information interne computationnellement significative. Ainsi, d'une part, la transduction est cognitivement opaque et sa fonction est non symbolique ; d'autre part, la signification est le résultat des opérations effectuées par les représentations mentales symboliques, lesquelles ne sont pas déterminées par le contenu physique des états de choses externes. Deux grands problèmes demeurent sans réponse dans ce paradigme. Du côté du sujet, les problèmes de *l'intensionnalité* et de *l'intentionnalité* sont évacués. Comment des représentations mentales symboliques peuvent-elles acquérir un sens et une orientation intentionnelle vers le monde externe ? Du fait de l'hétérogénéité du physique et du symbolique, on est conduit à adopter une approche solipsiste qui substitue à l'intentionnalité conçue comme l'étude des rapports sujet-monde celle des rapports sujet-sujet. Comment, en

outre, un système cognitif peut-il agir en fonction du sens des symboles alors qu'il ne possède de relations causales qu'avec la forme syntaxique de ceux-ci ? Du côté du monde externe, le problème de la manifestation qualitative et morphologique des phénomènes et de la conscience que nous en avons est laissé sans réponse. Se limiter à poser que le monde de l'expérience est le résultat d'opérations calculatoires de l'« esprit computationnel » est insuffisant. Il convient, en outre, de comprendre la part de ces calculs et opérations qui est constitutive de la structuration qualitative du monde en choses, état de choses, événements, processus, perceptivement appréhendables et linguistiquement descriptibles, c'est-à-dire conscientes, au sens de la conscience phénoménologique.

Le paradigme classique s'oppose, sur ce point, aux conceptions *émergentielles* (Petitot, 1991, 1992, Smolensky, 1988) qui plaident en faveur de représentations mentales géométriques (et non propositionnelles) et considèrent, à l'inverse, que les structures formelles des représentations mentales émergent de processus dynamiques sous-jacents. Cette conception prône l'utilisation de résultats théoriques et expérimentaux qui démontrent l'existence de structures morphologiques et qualitatives émergeant, par un processus dynamique organisateur, des substrats physiques et l'application de ces résultats aux substrats physiques que sont les substrats neuronaux. A l'inverse du paradigme classique qui fait l'hypothèse, au moins implicite, qu'il existe des formes symboliques dans la nature, cette orientation suggère que les formes de l'esprit, du langage et du sens sont plutôt (représentables par) des formes géométriques et dynamiques dans des espaces de représentation convenables. La stratégie consiste à introduire un niveau médiateur entre le physique et le symbolique : le niveau *morphologique* et à utiliser des modèles morphodynamiques comme modèles des processus cognitifs. Cette *morphodynamique cognitive* souffre, toutefois, de certaines limites inhérentes à la démarche utilisée. D'une part, il conviendrait de montrer que ces formes géométriques peuvent être traduites symboliquement et manipulées à des niveaux cognitifs supérieurs. D'autre part, si l'on admet qu'un contenu de représentation mentale peut être modélisé par un attracteur d'un système dynamique convenable, il faut pouvoir concevoir une syntaxe d'attracteurs. C'est l'un des points principaux sur lesquels s'est cristallisé le débat Fodor-Pylyshyn / Smolensky (Fodor et Pylyshyn, 1988, Smolensky, 1988, Petitot, 1995 ; voir aussi Amit *et al* 1995).

Cette médiation entre le physique et le symbolique proposée par la morphodynamique n'est pas la seule possible. La *compilation*

généralisée (Desclés, 1990, 1996) en est une autre. Le concept de compilation se situe au confluent de deux grands courants : celui qui est né du désir de rendre les langages de programmation plus indépendants des structures des machines ; celui qui puise son origine dans la linguistique et les recherches mathématiques auxquelles celle-ci a donné lieu. La compilation est une méthode qui autorise l'informaticien à concevoir, abstraitement et indépendamment des procédés d'exécution, des langages de haut niveau et, ensuite, à écrire des interpréteurs et des compilateurs chargés de « traduire » automatiquement les programmes écrits dans des langages de haut niveau dans les codes machines qui exécutent les programmes. Cette méthodologie peut être généralisée et importée dans d'autres domaines des sciences cognitives, notamment celui du langage naturel, stratégie devant conduire à une meilleure compréhension des rapports entre le physique et le symbolique. L'idée directrice de la compilation généralisée est que, entre, d'un côté, le support neuronal où les actes et les processus mentaux sont physiquement implémentés et, d'un autre côté, les représentations symboliques de haut niveau, il existe une hiérarchie de *niveaux intermédiaires* reliés entre eux par des rapports de compilation. Pour manipuler du symbolique complexe par des composants naturels relativement élémentaires (c'est-à-dire physiques), une méthode éprouvée consiste à engendrer des représentations intermédiaires qui sont traduites les unes dans les autres par un processus de compilation. Complémentaire de la notion d'émergence des niveaux subsymboliques aux niveaux symboliques, qui se doit d'expliquer « comment le sens vient au symbole », le principe de compilation généralisée a pour ambition de ramener des représentations symboliques (manifestées, en particulier, par le langage) à des implantations supportées par des organes physiques et biologiques, lesquels sont chargés d'effectuer les opérations mises en jeu par et sur ces représentations. L'idée maîtresse de ce programme de recherche est que les différents niveaux de représentation (par exemple, dans le cadre de l'analyse du langage : linguistiques, métalinguistiques, cognitifs, neuronaux) sont reliés les uns aux autres par des processus de « compilation » ; ces niveaux ne sont pas isomorphes les uns aux autres : chaque niveau jouit d'une relative autonomie et possède une structure propre ; chaque expression d'un niveau donné est le résultat de la compilation des niveaux inférieurs et sert d'entrée au processus de compilation orienté vers le niveau supérieur ; l'ensemble des niveaux est organisé en une *architecture cognitive*. Ainsi peut-on, par exemple, concevoir un niveau cognitif où langage, perception (des mouvements spatio-temporels) et actions

motrices deviendraient compatibles (Jackendoff, 1983). Il faut, pour ce faire, concevoir un niveau de représentation où les représentations issues, respectivement, du langage, de la perception et de l'action motrice soient comparables et, de plus, reliés aux observables linguistiques, perceptifs et moteurs par des processus de compilation différents. La mise en place d'un tel niveau cognitif de compatibilité permettrait de montrer comment le langage ancre certaines catégories grammaticales (relatives, par exemple, au temps et à l'espace) sur des catégories plus phénoménologiques.

L'hypothèse des représentations intermédiaires de la compilation généralisée insiste sur le caractère médiatisé des représentations symboliques et sur une certaine architecture nécessaire à l'organisation cognitive des niveaux de représentation et des contrôles entre représentations. On postule l'existence de représentations intermédiaires entre les représentations externes et les représentations directement compatibles avec les représentations effectuées par les neurosciences. Ces représentations intermédiaires ne sont ni des données ni directement déductibles des données d'observation : elles doivent être mises en place par un processus abductif prenant pour base les données empiriques directement observables. L'hypothèse cognitive des représentations symboliques intermédiaires conduit à la stratégie suivante : (i) prendre acte des possibilités offertes par les processus de compilation ; (ii) construire, par des procédures contrôlées et répétibles, c'est-à-dire de façon abductive, des représentations symboliques intermédiaires prenant pour base des comportements humains observables, certaines de ces représentations ayant le statut de représentations mentales ; (iii) se rapprocher progressivement, par un processus complexe de compilation, de représentations compatibles avec les structures neuronales des supports biologiques observables ; (iv) tester expérimentalement les représentations intermédiaires.

1.3. La question de l'intensionnalité

Le mentalisme que l'on retrouve aux fondements du cognitivisme n'est pas simplement une thèse triviale sur l'ontologie des processus mentaux, c'est une théorie de la cognition humaine qui renoue avec les présupposés de la « logique des idées », fondement de la pensée classique en philosophie et en théorie du langage : les opérations cognitives sont celles que l'on pratique sur les idées ; le langage, privé d'autonomie, est soumis à la pensée. La pensée est unité fonctionnelle indivisible (Auroux, 1994).

L'intelligence artificielle classique offre à la perspective cognitive, en ce qu'elle présuppose l'existence de représentations mentales physiquement réalisées et de nature symbolique, un point d'ancrage : les actes cognitifs sont identifiés à des processus de traitement de l'information réductibles à des manipulations de symboles formels. L'intelligence artificielle classique peut se définir comme la résolution de problèmes par modélisation de connaissances. La résolution automatique de problèmes obéit, en effet, à l'une des deux approches suivantes : celle qui utilise un modèle du monde où le problème se pose ; celle qui utilise un modèle des connaissances disponibles sur le problème, quand le modèle du monde est insuffisant (voire inexistant) ou inexploitable sur le plan informatique en raison des temps de calcul prohibitifs qu'entraînerait l'algorithme mettant en œuvre le modèle. L'intelligence artificielle classique s'inscrit dans le deuxième terme de cette alternative (Bachimont, 1992). Toutefois, la connaissance n'est pas un phénomène observable, directement accessible, pour lequel les méthodes classiques de modélisation, celles utilisées par exemple en physique, pourraient être mises en œuvre. Ainsi le processus de modélisation des connaissances que l'intelligence artificielle se propose d'entreprendre, consiste-t-il, non seulement en la définition du modèle formel utilisé, mais également en celle de l'objet de la modélisation : la connaissance. Les connaissances que l'intelligence artificielle se propose de modéliser sont celles d'un être humain, connaissances dont le support de représentation est une langue naturelle. Celui-ci se caractérise par le fait que ses expressions sont manipulées, non seulement de façon à obéir à une syntaxe, mais également en fonction de leur sens. Par ailleurs, puisque l'ordinateur est une machine à manipuler des symboles (Newell, 1990), la modélisation des connaissances en intelligence artificielle ne peut consister qu'en leur représentation symbolique. Cette conception est rendue explicite dans l'architecture de type von Neumann : les symboles sont des pointeurs (des représentants), pointant sur des contenus (des représentés) dans une mémoire adressable. Or, les symboles manipulés par un programme informatique n'ont, pour l'ordinateur qui l'exécute, aucune signification : la représentation symbolique dans un programme se ramène à une manipulation syntaxique de signes. Le problème se pose alors de la relation qu'entretient, d'un côté, les représentations symboliques manipulées (par le langage de programmation) en fonction de leur seule forme syntaxique et, d'un autre, les connaissances exprimées dans le langage naturel, et à ce titre manipulables en fonction de leur sens, que ces représentations modélisent. La littérature offre deux réponses à ce

problème. La première (Stich, 1983) revient à postuler que les connaissances sont, elles-mêmes, des représentations syntaxiques non susceptibles d'être interprétées et, par suite, que le sens n'existe pas. La seconde, largement dominante, consiste à postuler que le sens associé à une représentation est déterminé par la forme de celle-ci, partant, que le sens est une propriété formelle de la syntaxe. Les notions de structure et de règle, formalisées et mises en œuvre, sont alors naturellement associées aux concepts logiques de symboles et de règles d'inférence et ainsi identifiés aux objets d'une syntaxe formelle. Cette correspondance est transposée aux éléments de sens dont ces symboles sont censés être porteurs, lesdits symboles étant, dans le même temps, considérés comme le support d'une intentionnalité comparable à celle qui se manifeste dans les langues naturelles. Le paradigme dont on se dote est celui du « langage de la pensée ».

A l'inverse du paradigme classique, pour lequel un système cognitif est donc un système de traitement de l'information et, à un certain niveau, l'ordinateur de type von Neumann un modèle du fonctionnement de l'esprit humain, le paradigme connexionniste propose de simuler les processus cognitifs grâce à des réseaux parallèles et distribués sur des entités définies à un niveau sub-symbolique selon des architectures et des processus sans encodage symbolique ni règles de manipulation de symboles : le modèle adéquat est celui des réseaux de neurones formels. Les modèles connexionnistes se caractérisent par des connexions entre des unités de traitement de l'information communiquant entre elles grâce à des signaux d'activation. L'information véhiculée par le système n'est pas codée par le réseau mais est découverte par lui. La représentation est une propriété, non pas des symboles d'un langage de la pensée, mais celle de collections de ces constituants simples que sont les neurones formels : lorsqu'ils sont reliés les uns aux autres, ces constituants expriment des propriétés globales qui simulent les comportements cognitifs proprement dits. La thèse principale du connexionnisme est que les représentations internes, nécessaires à l'explication causale de la production du comportement intelligent, ne coïncident pas avec les représentations (croyances, désirs, ...) que l'on utilise pour la description globale de ce comportement ; que les premières ont une granularité plus fine que les secondes. Une telle thèse opère, par rapport à la thèse cognitiviste, un renversement de perspective pour tout ce qui touche aux questions d'acquisition de connaissances, d'apprentissage, de catégorisation, ...

De plus, selon la variante subsymbolique du connexionnisme (Smolensky, 1988, Petitot, 1992), les structures symboliques émergent de dynamiques sous-jacentes à travers des phénomènes

d'auto-organisation. La thèse maîtresse de cette approche, dite *morphodynamique*, est qu'il existe un niveau morphologique capable d'opérer une médiation entre le physique et le symbolique. Elle se fonde sur l'hypothèse de l'existence d'une information morphologique, présente objectivement dans l'environnement, issue de processus dynamiques d'organisation des substrats, et qui peut être reconstituée par transduction sensorielle et servir de base aux processus cognitifs symboliques. La conséquence, du point de vue de la modélisation, est que ce sont les mêmes outils mathématiques qui permettent de modéliser les processus d'organisation des substrats physiques et les processus d'organisation des réseaux neuronaux.

Si l'on admet que l'intentionnalité est la caractéristique distinctive et irréductible des phénomènes mentaux, la crédibilité des sciences cognitives, dans la perspective cognitiviste ou dans la perspective connexionniste, dépend, pour une large part, de sa capacité à proposer une solution au problème de *l'intentionnalité*. Elles doivent montrer soit que l'intentionnalité est un faux problème soit qu'une théorie naturaliste de l'intentionnalité est possible. Dans la perspective cognitiviste, pour laquelle les actes cognitifs sont identifiés à des systèmes de traitement de l'information réductibles à des manipulations de symboles d'une syntaxe formelle, la question se pose de savoir si des systèmes de nature formelle peuvent modéliser des processus cognitifs où l'interprétation du monde joue un rôle essentiel. Le problème de l'intentionnalité, du rapport à des contenus, se pose d'abord en termes sémantiques comme le problème des rapports entre des représentations mentales de nature logico-symbolique et une objectivité physique : il est nécessaire que les symboles soient dotés d'un sens et que le système soit capable d'agir en fonction du sens. Or, comment un système peut-il agir en fonction du sens des symboles alors qu'il ne possède de relation qu'avec la forme syntaxique de ceux-ci ? Le connexionnisme (au moins dans sa variante morphodynamique) permet d'aborder la question de l'intentionnalité sous un angle différent : la relation intentionnelle, qui va du sujet au monde, est vue comme la converse de la relation de causalité, qui va du monde au sujet, laquelle s'instaure entre des structures morphologiques émergentes dans l'environnement et une information morphologique constituant leur contrepartie mentale ; le fait que la relation de causalité soit nomologiquement descriptible permet de fonder la relation converse : celle d'intentionnalité. L'intentionnalité apparaît donc originellement perceptive, secondairement comme un fait sémantique (Petitot, 1995). Cette solution au problème de l'intentionnalité, pour attrayante qu'elle soit, n'en demeure pas moins

suspendue à la validité d'une hypothèse forte, celle de l'existence d'une information morphologique dans l'environnement, et à la possibilité de donner un contenu précis à l'idée d'émergence (Pacherie, 1993).

Le problème de l'intentionnalité soulève celui de l'intensionnalité (du sens). L'hypothèse, qui fonde l'option cognitiviste, que soit le sens n'existe pas soit qu'il est une propriété formelle de la syntaxe est aporétique. Si l'on considère, en effet, qu'une représentation ne contient pas en elle-même ce qui fait d'elle une représentation, c'est-à-dire qu'elle renvoie à autre chose qu'elle représente, cette hypothèse n'est plus tenable. D'un autre côté, l'élaboration d'une « physique du sens » à laquelle le connexionnisme (en l'occurrence la morphodynamique) est susceptible de conduire, reste tributaire, comme pour le problème de l'intentionnalité, de la validité des hypothèses qui la fondent. De plus, comment les formes peuvent-elles être manipulées à des niveaux cognitifs supérieurs, apparemment descriptibles en termes logico-symboliques ?

Au-delà des problèmes posés par chacune des deux thèses (disons logico-symbolique, d'un côté, connexionniste ou morphodynamique, de l'autre), il semble que celles-ci ne soient pas diamétralement opposées et que les querelles auxquelles elles donnent lieu relèvent, pour une part, d'une double incompréhension.

Tout d'abord, les partisans des deux thèses partagent une prémisse commune : la cognition consiste en une information véhiculée par des états *réels* de l'organisme ; les processus cognitifs sont des phénomènes réels et scientifiquement appréhendables. Les sciences cognitives ont pour objectif de se constituer comme science *naturelle* de l'esprit. Elle se propose d'élaborer des modèles d'une théorie de l'esprit compatible avec le monisme matérialiste dont se réclament les sciences de la nature (avec quelque violence pour ce qui concerne le paradigme classique).

Sur le plan opératoire (Visetti, 1990), le paradigme classique présente de nombreux avantages théoriques et méthodologiques : formalismes clairs et conformes aux capacités actuelles de l'informatique ; modularité des traitements ; identification du caractère descriptif des règles et de leur rôle causal ; réflexivité ; introduction facile d'heuristiques ; donc productivité, systématisme, compositionnalité, inférence. Toutefois, les limites et les inconvénients sont tout aussi nombreux et importants : manque de souplesse et de robustesse des systèmes ; découpages conceptuels figés ; modélisation très problématique des processus d'apprentissage ; perception incompréhensible ; séquentialité des opérations et contrôle

par une unité centrale ; d'une façon générale, impossibilité de dépasser les limites du logicisme sur lequel cette approche se fonde (par exemple : réduction des structures à la relation entre un prédicat logique et ses arguments dans un langage formel). Le paradigme connexionniste offre, quant à lui, à la modélisation cognitive les avantages et les inconvénients complémentaires. Les dynamiques de processus étudiés sont massivement parallèles, l'espace d'état peut être discret ou continu. La modélisation peut mettre en correspondance les éléments de la structure considérée avec des ensembles d'unités du réseau, lesquelles sont porteuses de traits distinctifs : elles constituent les relations causales sous-jacentes aux représentations symboliques de niveau supérieur. Les avantages d'une telle approche sont la souplesse et la robustesse des systèmes ; la possibilité d'une modélisation d'une granularité plus fine ; la relative facilité d'élaborer des algorithmes d'apprentissage généraux ; une plus grande vraisemblance neurobiologique. En revanche, les approches connexionnistes rencontrent de graves difficultés, tout aussi graves que celles auxquelles se heurtent les systèmes symboliques : les phases d'apprentissage et de fonctionnement sont généralement dissociées ; le fonctionnement du réseau n'est pas descriptible en termes d'opérations de haut niveau, lesquelles n'ont plus d'équivalents au niveau causal de la dynamique ; la connaissance étant implicite dans la dynamique du réseau, sa mise en évidence est quasiment impossible.

L'une des difficultés que l'on rencontre dans le dépassement de ces positions est d'arriver à une bonne compréhension des relations entre les modèles des systèmes micro-physiques de bas-niveau et ceux des systèmes macro-cognitifs de haut-niveau. C'est l'étude de ces relations qui constitue le défi majeur de l'ambition naturaliste. Pour parvenir à se constituer en sciences naturelles de l'esprit, les sciences cognitives doivent articuler trois types d'objectivité : objectivité physique du monde, objectivité logico-symbolique des langages, objectivité computationnelle des traitements effectifs. Les sciences cognitives doivent unifier ces trois types hétérogènes d'objectivité ; une telle unification soulève en particulier le problème de la compréhension des rapports entre *le physique et le symbolique*.

L'un des préjugés des tenants de la thèse logico-symbolique est qu'il n'existe pas de physique qualitative macroscopique de la structuration et de l'auto-organisation des substrats matériels, autrement dit qu'il n'existe pas de dynamique des formes. Ce préjugé est infondé si l'on prend en compte les résultats de l'application de la théorie des systèmes dynamiques et de la théorie des singularités à l'étude des phénomènes critiques, des phénomènes collectifs et coopératifs

d'organisation, ou encore du chaos déterministe ; résultats ouvrant la voie à une « physique du sens ». D'un autre côté, les critiques de l'approche logico-symbolique, par les partisans de la thèse connexionniste, s'adressent presque exclusivement à la logique classique et font abstraction du fait que, par exemple, l'élaboration d'une « logique du sens » est dans le domaine du possible.

Le dépassement des positions qui s'affrontent, en vue d'une éventuelle intégration théorique des paradigmes symbolique et connexionniste, requiert que les concepts en jeu, concepts que des approches empiriques et expérimentales ont permis de dégager, soient clarifiés. Une meilleure modélisation des processus cognitifs passe par la mise en lumière des liens rigoureux entre les outils théoriques utilisés, de la logique formelle à la géométrie différentielle (en passant par l'informatique).

2. INTENSIONNALITE

2.1. Introduction

Le champ sémantique du *logos* se déploie selon deux axes : celui du langage, celui du raisonnement. La formalisation de la logique s'est accompagnée non seulement d'une délégation à la linguistique des problèmes ressortissant au premier axe, mais également d'une tentation constante de réduire l'étude du raisonnement à son architecture logique et de méconnaître le principe interne sans lequel le raisonnement se dilue en un assemblage de signes et une juxtaposition de formules. Ce réductionnisme, qui atteint son point culminant avec les travaux de Russell, n'est pas pour autant imputable, *stricto sensu*, à la formalisation.

La logique formelle naquit de la rupture avec la Scolastique qui, confondant grammaire et logique, ne sut éviter les ambiguïtés d'interprétation de la copule dans la syllogistique traditionnelle, laquelle admettait pour seule analyse possible le découpage syntaxique en sujet-copule-attribut. Le paradigme scolastique, se fondant sur une métaphysique qui ne reconnaît que des substances en position de sujet et des attributs en position de prédicats, réduisait toute relation au rapport copulatif. Cette réduction, logique et philosophique, de la relation au rang d'attribut prit deux formes : l'une monadique qui, à la suite de Leibniz, réduit toute proposition relationnelle à une proposition prédicative ; l'autre, moniste, qui, à l'instar de Spinoza ou de Hegel, consiste à transformer toute relation entre deux termes en un prédicat attribué à l'objet formé des deux termes de la relation. Les

réductions moniste ou monadique interdisent de prendre en compte les diverses traductions logiques de la même forme grammaticale « est ». Ce schéma, avant d'être cloué au pilori par Russell, sous le nom de « dogme des relations internes », fut battu en brèche par Frege. Celui-ci, le premier, nota la distinction qu'il convient d'opérer entre propositions singulières du type « Socrate est mortel » et propositions générales telles « Tous les hommes sont mortels ». Le premier énoncé exprime une relation de *subsumption* d'un objet sous un concept : l'objet « Socrate » tombe sous le concept « mortel ». Le second énoncé exprime une relation de *subordination* entre deux concepts, tout objet qui tombe sous le premier concept tombant sous le second ; cette relation valant pour un individu quelconque, on quantifie universellement, d'où la conditionnelle universelle : quel que soit l'objet x , si x est un homme alors x est mortel.

Cette distinction logique reçut une contrepartie en termes de classes de la part de Peano : la première phrase exprime l'appartenance d'un individu à la classe des mortels, la seconde l'inclusion de la classe des hommes dans celle des mortels. Ainsi, qu'il s'agisse du formalisme frégeén ou de celui de Peano, l'ambiguïté de la copule est-elle levée par le clivage logique entre les deux relations : subsumption-subordination c'est-à-dire appartenance-inclusion. Les fonctions d'appartenance et d'inclusion n'épuisent pas l'ensemble des fonctions de la copule. Il revient encore à Frege d'avoir montré qu'une phrase comme « Vénus est l'étoile du matin » a la même forme logique que l'énoncé « Vénus est Venus », l'expression « l'étoile du matin » fonctionnant comme un autre nom de Vénus, c'est-à-dire qu'elle exprime une relation d'*identité*. Enfin, s'inscrivant en faux par rapport à la tradition aristotélicienne, Frege fit remarquer qu'il convient de réserver la fonction d'assertion existentielle aux seules propositions existentiellement quantifiées et de dénier aux conditionnelles universelles, du fait de leur caractère hypothétique, tout caractère existentiel : l'énoncé « tous les hommes sont mortels » ne suppose pas qu'il existe des hommes. Cette affirmation est le seul moyen d'autoriser un calcul quantificationnel rigoureux.

Les travaux de Frege montrent donc clairement que la même copule « est » peut valoir, sur le plan logique, comme relation d'appartenance, d'inclusion, d'identité ou comme assertion d'existence. Ces travaux de Frege et Peano sur le calcul propositionnel et le calcul des prédicats, qui ouvrent la période devant aboutir aux œuvres maîtresses des « Principia Mathematica » de Whitehead et Russell et du « Tractatus Logico-philosophicus » de Wittgenstein, tracent l'épure de ce qu'il est désormais convenu d'appeler la logique classique. Celle-ci affiche, en

tant qu'outil d'analyse et de représentation du langage et de la pensée, des limites maintes fois soulignées par les philosophes et les linguistes, limites qui nourrissent le scepticisme de nombreux chercheurs quant à la pertinence de la formalisation de leur discipline et les conduisent à opposer modélisation formelle et modélisation sémantique.

Les limites et les *a priori* logico-philosophiques sont clairement mis en évidence avec l'analyse de la copule « est » dans les langues. En effet, les valeurs sémantiques du relateur *est* et de ses homologues à travers les langues montrent que les langues indo-européennes cumulent sous un même marqueur « est » de nombreuses valeurs qui sont exprimées dans les autres langues par différents marqueurs. De plus, les principales valeurs qui ont été dégagées par les logiciens (Peano, Frege) ne sont pas celles qui sont mises en évidence par l'observation des langues. Les distinctions fondamentales ne sont pas « identité », « appartenance » et « inclusion ». Les grandes distinctions que laisse apparaître la diversité des langues (Lyons, Benveniste, Culioli-Desclés, Desclés 1987) sont d'un côté, « identification » (et non identité) et, d'un autre côté, « localisation ». L'identification d'une entité suppose que l'on trouve un identificateur mieux déterminé, par exemple du point de vue de sa référence, ou mieux connu. Par exemple, dans l'énoncé *Molière est l'auteur du Misanthrope*, il n'y a pas identité entre « Molière » et l'« auteur du Misanthrope » mais identification de celui que dénote *Molière* à celui qui est dénoté par *l'auteur du Misanthrope* : on suppose que l'on dispose de certaines informations à propos de l'« auteur du Misanthrope » dont on ne dispose pas à propos de « Molière ». L'identification donne une connaissance supplémentaire car il n'y a pas symétrie entre l'entité identifiée et l'entité qui sert d'identificateur. Cependant, cette opération d'identification engendre une relation symétrisable qui est réflexive et symétrique mais non transitive. L'identification n'est donc pas une simple équivalence, elle est plus faible. L'identité est une équivalence puisqu'elle doit respecter les deux lois de Leibniz (deux entités identiques doivent vérifier les mêmes propriétés ; deux propriétés identiques doivent s'appliquer exactement aux mêmes objets) et le principe d'extensionnalité (lorsque deux propriétés s'appliquent aux mêmes objets, alors elles sont identiques). Il y a donc une claire différence entre les informations contenues dans l'énonciation de « $a = a$ » (identité) et l'énonciation « $a = b$ » (identification de a à b). Or, les langues expriment beaucoup plus l'identification que l'identité. Un énoncé comme *Molière est Molière* ne doit pas être interprété comme une identité (aucune information ne

serait apportée par cette énonciation) puisqu'il s'agit d'une prédication où le prédicat *est Molière* s'applique au terme nominal *Molière*.

Lorsqu'une entité n'est pas identifiable à un identificateur, en raison des différences qui bloquent toute identification, l'entité peut cependant être repérée par rapport à une autre. En particulier, une entité peut être localisée par rapport à un lieu : l'entité peut être à l'intérieur du lieu ou à sa frontière ou dans sa fermeture ou encore à l'extérieur. Les langues naturelles expriment par des marqueurs spécifiques les différents types de localisation, souvent en composant un relateur de localisation avec une spécification topologique (intérieurité, exteriorité, frontière, ...) ou cinétique (mouvement dans un lieu ou absence de mouvement, franchissement ou non d'une frontière, ...). Les relations d'attribution d'une propriété à une entité, les attributions d'un objet à une classe (appartenance) ou les inclusions entre classes sont très souvent exprimées dans les langues par des relateurs apparentés directement à la localisation. Ces relateurs sont souvent associés à d'autres valeurs apparentées comme « possession », relations d'« ingrédience » et « partie-tout ». Dans certaines langues, par exemple le français ou l'anglais, le relateur « être » (« be ») entretient des relations avec d'autres relateurs comme « avoir » (« have ») : « y a x » peut être considéré comme le converse de « x est à y », avec cependant des conditions précises sur les déterminations référentielles sur x et sur y. Par exemple, nous avons la relation entre les énoncés : *Jean a un livre* et *un livre est à Jean*. L'ensemble des valeurs de la copule « est » se structure dans une architectonie de valeurs sémantiques qui spécifient un schème général de « repérage » : « x est repéré par rapport à y », qui apparaît comme un invariant traversant la diversité des langues (Desclés, 1987).

On peut ainsi isoler trois grands domaines du repérage : (i) les *identifications* plus ou moins fortes ; (ii) les *différenciations*, dont les localisations, les attributions, les possessions, les relations « partie-tout » sont des spécifications ; (iii) la *rupture* (non réflexive et symétrique) qui indique qu'une entité ne peut être ni identifiée ni différenciée d'une certaine entité posée comme un repère. Les différents relateurs de repérage sont constitutifs de certains processus de catégorisation puisque construire une catégorie revient à la fois à identifier par rapport à un identificateur, à différencier entre eux les éléments de la catégorie et à placer à l'extérieur de la catégorie toutes les entités qui n'ont rien de commun avec les autres éléments de la catégorie, ces entités étant en rupture avec ces éléments. L'étude des représentations cognitives et des processus de catégorisation ne peut pas se contenter des seules relations de subsomption entre objet et

concept ou de subordination entre concepts, représentées en intelligence artificielle par le seul relateur « *is-a* ». En effet, d'une part, comme nous venons de le voir, d'autres relateurs structurent les réseaux sémantiques et, d'autre part, les concepts d'un réseau ne sont pas homogènes : entités individuelles, entités massives, entités collectives, lieux, (voir section 3)

D'autres limitations, pour ce qui concerne la question du sens, peuvent être illustrées par quelques exemples très simples. Soit un prédicat linguistique (par exemple « est un homme ») qui exprime un concept. Le prédicat s'applique à des objets pour former des propositions dont la valeur de vérité est, pour certains objets, « vrai » et, pour les autres, « faux ». La classe des objets auxquels le prédicat s'applique pour donner la valeur « vrai » s'appelle l'extension du concept : l'extension de « homme » est constituée de tous les objets qui appartiennent à la classe des hommes. La notion d'extension renvoie à celle, frégréenne, de *Bedeutung* (référence, dénotation) ; la *Bedeutung* d'une proposition est donc, *in fine*, sa valeur de vérité. L'opérativité de la logique classique se fonde entièrement sur le principe d'extensionnalité c'est-à-dire sur la possibilité de substituer les uns aux autres les termes co-extensionnels (c'est-à-dire ayant la même extension, la même *Bedeutung*), les propositions ayant la même valeur de vérité. D'où l'importance accordée à la référence et la tentation d'y réduire la notion de sens ou d'exclure celle-ci de la logique. Cette exclusion du sens au seul profit de la référence n'est pas sans poser de sérieux problèmes dès lors que l'on assigne à la logique une tâche autre que la seule formalisation de la connaissance mathématique. Ainsi, les deux propositions « x est un homme » et « x est un bipède rationnel » sont-elles extensionnellement équivalentes ; toutefois, si la substitution de l'une à l'autre vaut *salva veritate*, elle ne vaut pas *salva significatione*. L'exigence d'extensionnalité de la logique classique exclut donc toute approche formelle de la question du sens, c'est-à-dire du *Sinn* frégréen ou, à la suite de Carnap (Carnap, 1947) et Church, de l'*intension* des concepts : le concept « homme » admet pour intension les caractères essentiels définissant un homme : animal, bipède, rationnel, ... (et pour extension la classe des individus possédant ces caractères). Le langage formel de la logique classique n'existe que pour fournir un matériau adéquat à la manipulation de signes, aveugles à la signification. De plus, la logique classique tolère, outre le principe de substitution des identiques, « $(p(a) \text{ et } a = b) \text{ } \emptyset \text{ } p(b)$ », la règle de généralisation existentielle « $p(a) \text{ } \emptyset \text{ } \exists x p(x)$ ». L'extensionnalité exclut donc de toute approche formelle les discours qui présentent cette propriété que l'application de ces règles altèrent la valeur de vérité

d'une formule composée. Or, les verbes d'attitude propositionnelle (savoir, croire, souhaiter que, ...) permettent de construire des contextes où la valeur de vérité des formules composées n'est pas préservée. Ainsi, puisque la formalisation ne peut s'exercer que sur des discours de type extensionnels et que les concepts intentionnels s'expriment sous forme de verbes engendrant des contextes non extensionnels, la logique classique n'autorise aucune approche formelle de l'intentionnalité, *a fortiori* de l'intensionnalité. Les limites du principe d'extensionnalité et, partant, de la logique classique justifient l'élaboration d'une logique de l'intensionnalité.

2.2. L'intuitionnisme

La sémantique des mondes possibles, développée par Kripke (Kripke, 1965, 1980) et par Hintikka (Hintikka, 1967, 1969, 1991 (traduction française Lavand)), prend sa source dans les travaux de Carnap. Celui-ci définit, à partir d'un langage prédicatif du premier ordre, la *description d'état*, analogue, par son mode de construction et le double usage auquel il se prête, à l'*état de choses possible* de Wittgenstein. Elle rend compte, comme celui-ci, dans le langage objet, d'une situation, relativement à laquelle une formule se trouve soit validée soit falsifiée. Grâce à cette notion fondamentale, Carnap est en mesure de développer une sémantique des notions comme la nécessité et la possibilité, prises dans des formules du type «il est possible que ... », «il est nécessaire que ... ». Pour ce faire, il faut spécifier des critères de substitution en contexte intensionnel : il suffit que deux expressions soient équivalentes dans toutes les descriptions d'état pour que la substitution de l'une à l'autre dans une formule modale n'altère pas la valeur de vérité de ladite formule. De tels critères constituent un analogue formel aux critères extensionnels. La sémantique des mondes possibles existe sous deux versions qui, bien que présentant des points communs, n'en diffèrent pas moins par leur esprit. Déterminer la valeur de vérité d'une formule dans une situation, c'est déterminer sa valeur de vérité dans un modèle. La sémantique des mondes possibles naît de deux mutations dans la théorie des modèles : on établit entre eux une relation d'ordre, on se donne la possibilité de les différencier en assignant à chacun une « ontologie locale ». Une telle mutation permet à Kripke et Hintikka de moduler le concept de nécessité en fonction des propriétés d'une relation R , dite d'*alternativité*. Nécessité et possibilité sont définies en fonction de règles du type : (i) *Nec* p est vrai dans un monde W ssi, dans tout monde W' tel que WRW' , p est vrai ; (ii) *Pos* p est vrai dans un monde W ssi il existe un monde W' tel que WRW' et p est vrai. Ainsi la relation R dissocie-t-elle les deux

dimensions de chaque couple de modalités : leur homologie formelle avec le couple dual de quantificateurs, la spécificité de leur signification conceptuelle. En particulier, si R est une équivalence, on obtient des sémantiques correspondant aux axiomatiques modales de Lewis. Il suffit d'ailleurs de modifier les propriétés de R pour généraliser l'appareil sémantique à des modalités comme les modalités déontiques ou épistémiques. La différence entre les deux versions de la sémantique des mondes possibles développées par Kripke et Hintikka concerne l'interprétation des opérateurs modaux dont la sémantique des mondes possibles permet de tester les formules. Kripke n'utilise la sémantique des mondes possibles que pour l'étude des modalités aléthiques (nécessité, possibilité) où l'on décèle difficilement l'intentionnalité plus immédiatement reconnaissable dans les concepts de connaissance, croyance, ... Hintikka tente de tirer de la sémantique des mondes possibles une théorie de l'intentionnalité. De plus, selon Hintikka, les expressions rapportant des phénomènes intentionnels présentent en général la propriété d'intensionnalité : l'intensionnalité serait le symptôme logique de l'intentionnalité. Kripke fait une lecture métaphysique des modalités aléthiques ; Hintikka entreprend une lecture transcendantale des modalités. La notion fondamentale de la sémantique hintikkienne est celle d'*ensemble modèle*. Celui-ci joue le même rôle que la description d'état dans la sémantique carnapienne. Hintikka a prouvé que la description d'état peut être réduite à l'ensemble modèle au titre de support d'une évaluation sémantique. Ce résultat permet de diminuer le nombre de conditions sémantiques définissant la satisfiabilité d'une formule. On détermine celle-ci par un test d'inclusion : une formule satisfiable est une formule que l'on peut inclure dans un ensemble modèle comportant toutes les sous-formules de ladite formule. L'établissement de la preuve par la recherche avortée du contre-exemple, méthode sous-jacente à la démarche hintikkienne, rend cette dernière extrêmement puissante. Par ailleurs, puisque la sémantique des ensembles modèles ne contient que des règles d'inclusion, l'ensemble des objets constituant l'univers du discours n'y est pas spécifié. Or, l'ensemble modèle étant défini à partir d'un sous-ensemble de l'ensemble des règles déterminant la description d'état, les règles sur la quantification ne permet pas de s'assurer de la portée référentielle du discours. La parade d'Hintikka, portant sur les quantifications existentielle et universelle, fait de sa sémantique une logique libre de présupposés existentiels. La portée référentielle du discours est assurée par l'appareil de quantification. Pour traiter les modalités, Hintikka élargit la sémantique des ensembles modèles en sémantique des systèmes modèles, famille d'ensembles modèles munie

d'une relation d'alternativité. Les propriétés de cette dernière sont spécifiées par des règles d'inclusion conditionnelles : une formule est satisfiable ssi elle peut être incluse dans un ensemble d'un système modèle, valide ssi sa négation ne peut être incluse dans aucun ensemble d'aucun système modèle. Dire qu'un monde W est alterne au monde W' c'est dire que W est concevable pour le porteur d'attitudes relativement à ce que sont ses connaissances, ses croyances, ... dans W' . L'interprétation de la relation R ouvre sur la distinction entre deux types de pertinence, la distinction entre les deux sens de savoir et de croire. L'approche hintikkienne, si elle permet de produire de manière élégante certaines analyses et de réfuter les critiques acerbes de Quine pour qui l'extensionnalité est non seulement une exigence technique mais aussi un principe ontologique, n'offre que peu d'arguments à opposer à Searle dont les critiques portent essentiellement sur le lien supposé entre intensionnalité et intentionnalité. Selon Searle (Searle, 1983), l'affirmation que Jean croit que le Roi Arthur a tué Lancelot est un rapport intensionnel, mais la croyance de Jean, comme telle, n'est pas intensionnelle. Elle est purement extensionnelle : elle est vraie ssi

- (a) il existe un x unique tel que $x = \text{Roi Arthur}$,
- (b) il existe un y unique tel que $y = \text{Lancelot}$,
- (c) « x a tué y » est vraie.

La croyance de Jean est extensionnelle même si l'affirmation que l'on porte sur cette croyance est intensionnelle. L'intentionnalité est la propriété immanente à l'esprit de pouvoir représenter d'autres choses ; l'intensionnalité désigne l'impossibilité pour certaines assertions de satisfaire les principes logiques de l'extensionnalité. Identifier l'intentionnalité à l'intensionnalité procède d'une confusion entre les caractéristiques des rapports et celles des choses rapportées.

La vision catégorique de la sémantique de Kripke (Goldblatt, 1983, Fourman, 1977) permet une lecture non modale de la sémantique des mondes possibles. Soit I l'ensemble des états de la connaissance, variables dans le temps, d'un agent donné (ou, de façon équivalente, l'ensemble des états de la connaissance, à un instant donné, d'un ensemble d'agents). On peut associer à tout prédicat p , en tout état i de la connaissance, la classe p_i des objets x pour lesquels $p(x)$ est vraie. Sous l'hypothèse que la vérité persiste dans le temps, le fait d'associer à tout état i la collection des objets pour lesquels une proposition est vraie revient à définir un foncteur de I dans la catégorie des ensembles. La catégorie dont les objets sont ces foncteurs est un *topos*, objets qui

peuvent être vus comme des « ensembles intensionnels », comme l'intension du concept dont p est l'extension.

La sémantique de Kripke (Kripke, 1965), ainsi reformulée en termes catégoriques, rend complète la logique intuitionniste de Heyting (Heyting, 1966), laquelle est obtenue, à partir de l'axiomatique de la logique classique, par abandon de l'axiome du tiers exclu.

La réfutation du tiers exclu, pour spectaculaire qu'elle soit, ne se justifie pas moins par une série de raisons précises (Largeault, 1992). Mais la place qu'occupe cette réfutation dans les écrits des intuitionnistes tient en partie à des raisons historiques : ce fut l'arme utilisée par les intuitionnistes contre le formalisme hilbertien. Le programme formaliste de Hilbert consistait

(i) à transformer les théories mathématiques en un jeu de formules dont la production est assurée par une axiomatique et des règles mécaniques de dérivation ;

(ii) à démontrer, par des moyens combinatoires, l'impossibilité de dériver ainsi une contradiction.

Selon cette méthode axiomatique, les seuls objets mathématiques sont les signes et les relations entre ces signes, dérivés grâce à ce mécanisme de production. Les axiomes définissent une existence *potentielle*, la preuve de non-contradiction ayant pour objet d'assurer le passage du potentiel à l'actuel. Ainsi, dans une théorie formalisée, le non-contradictoire se substitue-t-il formellement au vrai. Le premier théorème d'incomplétude de Gödel de 1931, qui établit que « vrai dans un modèle de l'arithmétique élémentaire » et « prouvable » ne coïncident pas, mit fin à une telle conception des mathématiques : en vertu de ce théorème, la notion de non-contradiction (c'est-à-dire formellement de consistance ou cohérence) est incomplète. En d'autres termes, aucun moyen n'existe, si ce n'est celui de se laisser guider par l'intuition de ce qui est vrai, pour le choix d'une axiomatique et du mécanisme d'inférence d'un système formel. L'intuition mathématique ne se laisse pas réduire en termes algorithmiques.

Le programme de Hilbert, s'il avait été couronné de succès, aurait permis de montrer que la logique classique, et toutes les mathématiques, obéissent à une forme, implicite, de constructivisme, puisque la démonstration du non-contradictoire n'utilise que des moyens constructifs, que les preuves de non-contradiction n'utilisent que des moyens finitistes. Il s'agissait, dans l'esprit de Hilbert de vider de sa substance le point de vue constructiviste de Brouwer, selon lequel les mathématiques doivent se limiter à des procédés constructifs. Le

second théorème de Gödel de 1931, qui établit l'impossibilité de démontrer la non-contradiction d'une théorie grâce aux seuls moyens que la théorie fournit, porta un coup fatal au programme hilbertien.

Il faut noter, toutefois, que le programme de la «*reverse mathematics*», lancé en particulier par Takeuti et Feferman, qui consiste à résoudre le problème suivant : «*Etant donné un théorème T des mathématiques ordinaires, quels axiomes d'existence ensembliste sont-ils nécessaires pour prouver T ?*», va dans le sens d'une relativisation de l'échec du programme de Hilbert, lequel devient ainsi un programme de réduction des parties significatives de l'arithmétique du second ordre à des systèmes significativement plus faibles.

Le rejet par Brouwer du tiers exclu trouve sa source dans le caractère en devenir des objets mathématiques intuitionnistes : une assertion quelconque sur des objets en devenir ne saurait être, d'une manière définitive, vraie ou fausse. Plus précisément, si l'on donne à « *p est vraie*» le sens constructif, c'est-à-dire si « *p vraie*» signifie qu'on dispose d'une preuve de p , «*non- p* » signifie « *p est contradictoire*» (ou impossible, ou réductible à l'absurde). Ainsi le non-contradictoire intuitionniste se substitue-t-il au vrai classique et le contradictoire intuitionniste au faux classique. Une telle conception puise sa légitimité dans le fait, comme l'a montré Gödel, que la double négation (c'est-à-dire la non-contradiction) de toute loi de la logique propositionnelle classique est valide en logique intuitionniste. En d'autres termes, le contradictoire intuitionniste du vrai, c'est-à-dire du prouvé vrai est le non-vrai, au sens du non prouvé vrai, non pas au sens de prouvé non vrai. Suivant une telle interprétation, « *p ou non- p* » signifie « *p est prouvée vraie ou p est prouvée contradictoire*», c'est-à-dire que l'on dispose d'une preuve que p est vraie ou d'une preuve que p est contradictoire. Accepter que tel n'est pas le cas de toute proposition p vaut rejet du tiers-exclu.

Il est clair qu'un tel rejet ne relève pas d'un désir d'échapper à l'alternative vrai-faux, mais du refus, d'une part, d'admettre la *symétrie* du vrai et du faux, et, d'autre part, d'adhérer à un *principe d'omniscience*. En effet, la conception classique de la vérité et de la fausseté fait fi du processus permettant d'établir la vérité ou la fausseté. En séparant, ainsi, l'univers logique en deux moitiés symétriques, elle occulte le caractère *local* des assertions et des négations. Cette symétrisation induite est la cause principale des problèmes rencontrés par la logique classique : récursivité du vrai et non-récursivité du faux, non effectivité de la sémantique et, par suite, divorce entre la syntaxe et la sémantique, théorème de non-catégoricité

de Löwenheim-Skolem, ... Par ailleurs, le tiers exclu équivaut à un principe d'omniscience. Son rejet vaut refus du postulat de la résolubilité de tous les problèmes : le tiers-exclu est une catégorie de l'être (« être ou ne pas être » est une tautologie) non de la connaissance.

On voit donc que la sémantique de Kripke exploite, à partir de la logique de Heyting, l'idée d'états de la connaissance, ou d'états mentaux, variables dans le *temps*. A chaque instant, l'état de la connaissance, ou l'état mental qui lui est associé, consiste en un ensemble de vérités établies, c'est-à-dire prouvées. Ces connaissances persistent dans le temps et ne peuvent que s'enrichir d'autres connaissances. L'introduction d'une dimension temporelle dans la sémantique des mondes possibles renvoie à l'hypothèse brouwerienne d'objets mathématiques en devenir.

2.3 L'intuitionnisme revisité

Toutefois, la sémantique de Kripke soulève de nombreuses questions. Est-elle, en particulier, vraiment intensionnelle ? La notion de vérité locale, sous-jacente à la logique des mondes possibles n'introduit-elle pas une extensionnalité locale ? Le fait que le topos permettant de fonder la sémantique des mondes possibles est équivalent (dans un sens mathématique précis) au topos des faisceaux sur un espace convenablement choisi semble plaider en ce sens : c'est la notion intuitive de vérité locale qui a présidé à la naissance du concept de faisceau. Par ailleurs, la sémantique de Kripke n'opère-t-elle pas une inflexion par rapport à la philosophie constructiviste ayant présidé à l'élaboration de la logique intuitionniste de Heyting (*a fortiori* par rapport au programme de Brouwer) ? Le déclasserement de la prouvabilité de la disjonction de deux propositions (laquelle requiert un mécanisme de bifurcation entre les preuves de chacune des deux propositions constitutives) en l'affirmation que ces deux propositions sont tautologiques semble le montrer.

Ces questions suscitent, en amont, d'autres questions, plus fondamentales, sur les liens réels entre théorie des topoi et théorie des ensembles. Le point de vue de base de la théorie des catégories consiste à concentrer l'identité ensembliste d'une application dans une flèche. La théorie des topoi opère une mise à plat des propriétés qu'une catégorie doit posséder pour se comporter « comme » la catégorie des ensembles. Elle permet ainsi de développer une théorie des modèles pour la logique du premier ordre parallèle à la théorie classique des modèles, généralisant cette dernière en ceci que ses modèles sont des

objets dans un topos plutôt que des ensembles. Elle fournit également une élégante réponse au problème des objets dont l'existence est incertaine, par introduction d'un prédicat existentiel dont l'interprétation sémantique est la « mesure » du degré auquel un objet existe. Plus précisément, on définit deux relations d'égalité. La première associe à tout couple d'objets une quantité, élément d'une algèbre de Heyting complète, interprétée comme le degré d'incertitude sur l'égalité entre ces deux objets ; si les deux objets coïncident, la mesure est celle du degré d'existence de l'objet. La seconde ne permet pas de distinguer les objets au regard de leur existence : deux objets sont équivalents si aucun d'eux n'existe ou s'ils existent et sont égaux. Cette approche généralise, par ailleurs, les travaux sur la méthode du forcing, utilisée par Cohen pour montrer l'indépendance de l'hypothèse du continu : les travaux de Cohen, une fois reformulés en termes de modèles booléens, au sens de Scott-Solovay, de la théorie des ensembles, conduisent à la notion de topos. Toutefois la théorie des topoi présuppose largement la théorie des ensembles, puisque la collection des flèches d'un objet vers un autre est un ensemble (on parle dans ce cas de « petite » catégorie). Rappelons que la notion de topos a émergé des travaux des géomètres, lesquels n'entendaient pas renoncer totalement à la puissance de la théorie des ensembles. Ainsi n'est-il pas certain que le projet global de remplacement de la doctrine ensembliste par une autre, devant conduire à une logique de l'appartenance, de l'infini, du continu, différente de celle de la théorie des ensembles puisse être réalisé au sein de la théorie des topoi et que celle-ci puisse être vue comme l'aboutissement de l'idéal constructiviste de Brouwer et Heyting.

Toujours est-il que la nécessité de construire une vraie logique intensionnelle, une vraie logique du « sens », prenant en compte, sur son versant géométrique, les représentations logiques des lieux, autorisant, via un assouplissement des axiomes de la topologie, la construction d'une véritable théorie des lieux, intégrable dans une théorie des représentations cognitives, est patente. Il est clair que le problème du continu est l'un des principaux points d'ancrage de cette construction : le continu substantif, identifié comme une certaine sorte de multiplicité dans le discours mathématique classique ; le continu au sens adjectival permettant de distinguer entre les processus continus et discontinus dans la langue naturelle, entre les applications continues et discontinues dans le cadre mathématique¹. Le premier est susceptible

¹ Cf. Salanskis [1991], p. 144.

d'une approche en termes de cardinalité, le second est justiciable d'une approche de type topologique. Mais ces deux dimensions de signification sont étroitement liées, le problème du continu étant celui du substrat de la droite réelle, dont toutes les propriétés, les topologiques et celles de cardinalité, interviennent conjointement.

La *locologie* (De Glas, 1992, 1993) offre des éléments de réponse à ces questions. Une locologie sur un ensemble X est une relation sur X dont les propriétés sont la traduction formelle des notions intuitives de « ressemblance » ou d'« indiscernabilité » sur X . En donnant un contenu formel aux notions intuitives de « ressemblance » et de « granularité », la donnée d'une locologie permet de définir les opérateurs de *cœur* et d'*ombre* (d'un ensemble) qui, tout en présentant certaines similarités avec ceux d'intérieur et d'adhérence en topologie, ne diffèrent pas moins de ceux-ci, en ce qu'ils ne sont pas idempotents. Ils peuvent, de plus, être vus comme la contrepartie « ensembliste » des connecteurs modaux de nécessité et de possibilité. Tous les concepts fondamentaux de la topologie peuvent être redéfinis dans un cadre locologique, concepts dont certaines des propriétés sont remarquables : continuité d'une fonction présentant des « sauts » négligeables eu égard à un certain critère, préservation de la continuité d'une fonction après discrétisation (sans hypothèse particulière sur la locologie dont l'espace est muni), connexité de certains ensembles dénombrables, convergence de certaines suites oscillantes, ... Ce versant « géométrique » de la locologie, qui obéit à une vision non atomiste du continu, se double d'un versant logique. La locologie obéit aux axiomes d'une logique sub-intuitionniste (abandon de l'axiome $p \emptyset (q \emptyset p)$). Sa sémantique donne une traduction formelle de « je sais, dans un état donné de la connaissance, que p (est vraie) ». Se trouve ainsi formalisée, dans un cadre non modal, la notion de « point de vue » (ou, pour reprendre la terminologie de Kripke, de « monde possible »). La sémantique locologique est une généralisation formelle de la sémantique de Kripke (connaissance locale vs vérité locale) : à la différence de cette dernière, elle n'est pas, même localement, compositionnelle. La jonction entre ces deux versants est opérée par l'analyse des fondements catégoriques de la locologie.

L'approche locologique, qui autorise un « discret-dense », ou, si l'on préfère, un « continu-discret », pour peu que l'on regarde \mathbf{Z} « de loin », conduit à une lecture nouvelle du continu. La construction utilisée procède d'une approche post-cantorienne, commandée par des exigences constructives qui visent à construire le continu à l'aide de stratégies discrètes et à la remise en cause de la représentation ensembliste du continu à la Cantor-Dedekind. L'approche locologique

présente certaines similarités avec d'autres approches qui, depuis les travaux fondateurs de Brouwer et Heyting, tentent d'échapper à la juridiction ensembliste : l'analyse non standard (Robinson, 1974), l'*alternative set theory* (Vopenka), la méréologie (Lesniewski), la théorie des ensembles empiriques (Benabou). On peut voir, dans toutes ces approches, une revanche posthume du constructivisme brouwérien sur le formalisme hilbertien, à tout le moins la possibilité offerte de reprendre sur de nouvelles bases le programme formaliste de Hilbert en contournant les limites imposées par les théorèmes d'incomplétude de Gödel. A la doctrine formaliste, qui, dans sa volonté d'autofondation, a conduit à dénier aux mathématiques tout contenu sémantique objectif, et qui, de ce fait (?), a dominé les mathématiques de ce siècle, vient se substituer une méthode constructive pour qui les mathématiques sont conçues comme une science d'objets. Il convient, toutefois, de ne pas en tirer la conclusion que la transcendance objective des contenus sémantiques mathématiques peut se réduire à celle de structures finitairement accessibles.

Inconcevable, en tant que modalité mathématique, aux yeux de la tradition aristotélicienne, les concepts d'(ensemble) infini, au sens de l'infini actuel, et, corrélativement, de continu, furent portés, grâce aux travaux de Cantor et Dedekind, puis ceux de Zermelo et Fraenkel, sur les fonts baptismaux. Les mathématiques du vingtième siècle sont, pour l'essentiel, les héritières de cette révolution cantorienne qui fait de l'infini actuel un objet définissable, un objet d'investigation. Ce qui, au premier chef, caractérise cet infini actuel, c'est son irréductibilité au fini et, partant, son caractère « transcendant ». Cette rupture confine toutefois au paradoxe : dans l'ordre de l'infini, il y a toujours une possibilité de dépassement ; un ensemble peut être équipotent à l'une de ses parties propres. La théorie des cardinaux, qui consacre la rupture, est intrinsèquement incomplète. La preuve, établie par Cohen, de l'indécidabilité de l'hypothèse du continu, outre qu'elle fournit un exemple d'application du théorème de Gödel, montre l'impossibilité d'une hiérarchie complète des cardinaux transfinis et, via la méthode du *forcing*, l'irréductibilité de la vérité ensembliste à la validité constructive d'une théorie. En termes plus techniques, les axiomes ZFC (c'est-à-dire l'axiomatique de Zermelo-Fraenkel plus l'axiome du choix) ne permettent pas de dominer axiomatiquement le continu, parce qu'ils ne permettent pas de caractériser univoquement (de façon « absolue ») les cardinaux et en particulier la fonction $2^{\aleph_a} = \aleph_{f(a)}$. Dans un univers ZFC satisfaisant l'axiome de constructibilité, les ensembles pathologiques impliqués par l'axiome du choix et qui devraient être inconstructibles et indéfinissables sont forcés à être constructibles et

définissables : c'est le cas, par exemple, d'un bon ordre sur \mathbf{R} ou d'un sous-ensemble de \mathbf{R} non mesurable au sens de Lebesgue. Plus précisément, dans la hiérarchie de Luzin des sous-ensembles projectifs de \mathbf{R} (c'est-à-dire les sous-ensembles définissables à partir des intervalles ouverts par itération des opérations de réunion, d'intersection, de complémentation et de projection), l'axiome de constructibilité force certains projectifs à être très irréguliers. De plus, dans ZFC, la régularité des projectifs n'est démontrable que pour les premiers niveaux de la hiérarchie projective ; aux niveaux supérieurs, on rencontre des indécidables métamathématiques. Guidé par la nécessité d'obtenir une théorie plausible du continu, c'est-à-dire de rendre la transcendance du continu (transcendance qui renvoie le continu au-delà du constructible) intramathématique, Gödel a conçu l'idée, grâce à une inversion de l'exigence de constructibilité, d'enrichir l'univers axiomatique en introduisant des axiomes d'existence de grands cardinaux : cardinaux inaccessibles (Tarski), cardinaux mesurables (Ulam, Solovay, Martin), axiomes garantissant la régularité des projectifs. Pour accéder à des propriétés de détermination de ces projectifs, il faut, de plus, introduire des axiomes d'existence de cardinaux encore plus grands.

Une autre dimension du continu est celle qui renvoie à la structure topologico-algébrique de la droite réelle : \mathbf{R} est un corps ordonné, connexe, complet dans lequel \mathbf{Q} est dense. La densité de \mathbf{Q} rend compte formellement de l'idée que, aucun intervalle ouvert de \mathbf{Q} n'étant vide, le continu n'admet pas de «trous ». La connexité et la complétion de \mathbf{R} renforcent cette idée : d'une part, \mathbf{R} ne peut s'écrire comme réunion de deux sous-ensembles ouverts disjoints (connexité) ; d'autre part, dans \mathbf{R} , toute suite de Cauchy converge (complétion). Que toutes ces propriétés soient, par construction, déniées à \mathbf{Z} illumine le propos de Cohen évoquant «l'incroyable richesse du continu ». Cette incroyable richesse (topologique) se trouve, toutefois, privée de sa splendeur, l'indécidabilité de l'hypothèse du continu l'interdisant d'être transposée dans la hiérarchie des cardinaux.

La théorie des grands cardinaux règle, certes, formellement le problème de l'insertion de la transcendance de l'infini au sein des mathématiques, mais la conséquence fondamentale d'une telle option est la quasi impossibilité, déjà inscrite dans la théorie naïve des ensembles à la Cantor, d'une appréhension intuitive du transfini : il est permis, sur le plan épistémologique, de questionner le bien fondé du platonisme de Gödel. Sans perdre de vue que la signification profonde des résultats d'incomplétude et d'indécidabilité est qu'il existe un écart irréductible entre le concept computationnel de prouvabilité et celui de

vérité mathématique objective et que, par suite, la logique, quand elle opère comme un organon, se délite en une logique dialectique (au sens péjoratif du terme) qui nous interdit de comprendre l'accès que nous avons aux objectivités mathématiques, il n'en reste pas moins que le platonisme gödelien qui a inspiré la théorie des grands cardinaux, se fonde sur le postulat que les objets mathématiques ne sont pas des constructions mentales. Cette conception, transcendantale (Petitot, 1992), de l'intuition ouvre sur la question de l'intuition pure.

La locologie amène, également à repenser le concept de localité. A la différence de la localité topologique, qui est une localité indéterminée pouvant être n'importe quoi entre le vide et l'espace sous-jacent tout entier², la locologie renoue avec l'idée, leibnizienne, de localité infinitésimale. Elle ouvre, de plus, dans sa dimension logique, vers une vérité locale exprimable dans une logique géométrique/catégorique. Evidemment, ce local, autonomisé par rapport au continu et même au topologique, ne permet pas d'embrasser toutes les mathématiques (la théorie des systèmes dynamiques, dans ses développements récents, voit dans la localité une composante décisive sans pour autant remettre en cause le continu réel et les outils de la topologie), mais conduit à la question de la faisabilité d'une géométrie différentielle « discrète ».

3. CATEGORISATION

3.1. Introduction

Les sciences cognitives, en particulier dans leur version computationnelle, ont renouvelé considérablement la problématique dans laquelle la question de l'organisation des connaissances et celle, corrélative, des relations entre langage et pensée étaient abordées tant en psychologie cognitive qu'en linguistique, renouvellement ayant conduit à un renforcement de la conception rationaliste et logiciste de l'analyse de ces questions. Toutefois, les difficultés de la validation, en psychologie cognitive, des modèles de mémoire sémantique et celles, en intelligence artificielle, liées à la gestion des grandes bases de connaissances conduisirent à la mise en échec des formalismes classiques hérités de la logique du premier ordre et des organisations hiérarchiques (tels que les réseaux de Collins et Quillian, 1969). Selon le formalisme classique, l'appartenance catégorielle est définie par des conditions nécessaires et suffisantes portant sur les propriétés des

² Cf. Salanskis [1991], pp. 175-179.

catégories. Le schéma utilisé met donc en dualité l'intension (les propriétés) et l'extension (les exemplaires) des catégories. En conséquence les divers exemplaires d'une catégorie sémantique doivent être considérés comme «équivalents» (voir Miller, 1969, Miller et Johnson-Laird, 1976, Keil, 1979, et ses «concepts ontologiques», Genter, 1981, et ses «composantes de verbes»). Les années 70 et 80 virent l'éclosion de formalismes (ou pseudo-formalismes) visant à l'élaboration de modèles «opératoires» (voir, par exemple, Dubois 1993).

Les recherches de Rosch (Rosch, 1975, Rosch et Lloyd, 1978, Mervis et Rosch, 1981) s'inscrivent, à la fois, dans le courant des modèles de traitement de l'information fondés sur une conception classique des concepts et dans celui des recherches anthropologiques. La perspective roschienne obéit au double souci de munir les sciences cognitives d'une dimension écologique, naturaliste, et de s'opposer au relativisme culturel, *a contrario*, de mettre au jour les principes «universels» de la cognition indépendants des systèmes des langues. Cette double influence se traduit, chez Rosch, par la prise en compte, comme modèle de l'organisation des connaissances en mémoire, du modèle taxonomique de classification systématique du monde naturel.

Les recherches de Rosch et de ses collaborateurs sont à l'origine d'un regain d'intérêt pour l'étude de la catégorisation comme activité cognitive fondamentale. L'idée fondatrice de Rosch, assez simple *a priori*, est qu'il nous est impossible d'appréhender la réalité, dans sa diversité quasiment infinie, dans son individualité intrinsèque, et que, pour la penser, il nous faut réduire sa multiplicité et sa complexité grâce à une organisation catégorielle. Le problème qui se pose alors est de modéliser la façon dont nous opérons un découpage de la réalité. Dans une telle perspective, catégoriser revient à organiser un domaine en y introduisant des continuités et des discontinuités et à y faire valoir des principes d'agrégation des objets du domaine en catégories, la catégorisation étant le reflet de la capacité d'abstraction de l'esprit humain. Ces principes de regroupement d'objets en catégories sont, dans l'esprit de Rosch, des principes psychologiques. Pour y accéder, la psychologie étudie les jugements prédictifs de subsomption d'un objet sous une classe d'appartenance. Le cadre d'analyse de ce processus de catégorisation est celui d'une théorie représentationnaliste de l'esprit : la catégorisation procède d'une comparaison de deux représentations mentales, celle de l'objet et celle de la catégorie, ou celle du token et celle du type (Jackendoff, 1983). Dans cette problématique, les catégories sont des classes d'objets. Mais, une classe d'objets ne constitue pas nécessairement une catégorie. C'est sur ce point que

l'analyse roschienne de la catégorisation semble rompre avec une vision logique de la catégorisation remontant à la tradition aristotélicienne. Pour qu'une classe d'objets accède au statut de catégorie, il faut que ses membres appartiennent à un même ensemble sur les bases d'un critère de ressemblance, qu'ils manifestent des traits fondamentaux partagés. La catégorisation est un jugement de similarité, jugement formé par comparaison de traits. Le rejet de la théorie classique se manifeste, également, dans la théorie du prototype : certains objets sont de meilleurs représentants de la catégorie que d'autres ; l'agrégation d'un objet à une catégorie ne se décide pas en examinant s'il présente chacune des propriétés nécessaires et suffisantes qui définissent la catégorie mais en comparant l'objet au(x) prototype(s) de ladite catégorie. En outre, plus le jugement de catégorisation est rapide, plus l'objet est typique de la catégorie. Un dernier point qui sépare la catégorisation roschienne de la théorie classique est que les catégories sont prises dans une taxonomie : les catégories les plus spécifiques sont incluses dans les catégories plus englobantes, lesquelles sont plus abstraites et correspondent à une détermination moindre. De plus, de ce point de vue, l'idée de Rosch est qu'un certain niveau de catégorisation, le niveau de base, est spontanément choisi dans le processus de catégorisation, celui à partir duquel on peut déduire le plus grand nombre de propriétés susceptibles de le déterminer. L'analyse roschienne s'insère donc dans une étude psychologique des concepts : la catégorisation est une théorie de la formation et de l'organisation des concepts ; un concept est la représentation mentale, mise en jeu dans la catégorisation, des propriétés de la classe d'objets qui constitue son extension.

Plus précisément, l'objectif premier de Rosch est d'identifier les principes psychologiques qui constituent les structures mentales des catégories d'objets naturels ou manufacturés. La première hypothèse qui fonde les recherches de Rosch est que « la catégorisation humaine ne doit pas être considérée comme le produit arbitraire d'un accident historique, mais comme le résultat de principes psychologiques de catégorisation » (Rosch, 1975). Ces principes sont au nombre de deux. Selon le premier, celui d'*économie cognitive*, le rôle du système de catégorisation est de fournir le maximum d'information pour le moindre effort cognitif. Catégoriser un objet revient à le considérer, non seulement comme équivalent aux autres objets de la même catégorie, mais également comme différent de ceux qui n'appartiennent pas à cette catégorie. Le second principe, celui dit de *réalité*, revient à postuler que le monde perçu n'est pas un ensemble non structuré d'attributs de co-occurrences équiprobables ; que, au contraire, il existe

dans le monde des discontinuités et des corrélations ou co-occurrences de propriétés. Le monde des objets constitue, ainsi, un échantillon de l'ensemble des possibles, les patterns de co-occurrence d'attributs n'étant pas équiprobables ou réguliers mais possédant une structure *corrélationnelle*. La perception humaine est à même d'identifier ces discontinuités et ces corrélations de formes.

Ces deux principes ont des conséquences importantes sur la structure catégorielle. Les systèmes des catégories se déploient sur deux dimensions : la *dimension verticale* qui consiste en le niveau d'inclusion des catégories (chien, mammifère, animal, être vivant) ; la *dimension horizontale* qui concerne la segmentation des catégories à un même niveau d'inclusion. Pour ce qui concerne l'organisation verticale, les deux principes qui fondent l'approche roschienne ont pour conséquence que tous les niveaux possibles de catégorisation ne sont pas également pertinents et que, au contraire, il existe un niveau de catégorisation privilégié, le niveau de base, qui se définit comme le niveau d'abstraction pour lequel les catégories possèdent le plus d'attributs en commun. En amont, se situent les niveaux subordonnés et, en aval, les niveaux superordonnés. Pour ce qui est de l'organisation horizontale, les exemplaires d'une catégorie tendent à être situés sur un gradient de typicalité, certains exemplaires, les prototypes, étant plus représentatifs de la catégorie que les autres.

La théorie roschienne, brossée ici à grands traits, à l'intersection de la psychologie du langage et de la mémoire sémantique, est venue donner un contenu empirique à des idées qui s'étaient développées dans d'autres champs (par exemple les « photographies mentales mélangées » de Galton, 1879, ou encore la notion d'« air de famille » des membres d'une même catégorie, développée par Wittgenstein, 1953). Elle s'inscrit aussi comme héritière de travaux de psychologues sur la similitude (Attneave, 1950), l'apprentissage (Attneave, 1957) et la « genèse des idées abstraites » (Poster et Keele, 1968, 1970).

La théorie de Rosch a été l'objet de nombreuses critiques (voir, par exemple, Kleiber, 1990). Au delà de ces critiques, somme toute mineures, il convient de pratiquer un examen approfondi de ces divers concepts : les modèles proposés, dans la mesure où ils font l'impasse sur les possibilités de formalisation, semblent, en fait, prisonniers d'une formalisation naïve et par trop simpliste pour rendre compte de concepts aussi complexes que ceux de typicalité, atypicalité, prototypicalité, « dualité » intension-extension, ... C'est ainsi que l'on peut construire des modèles de la typicalité qui restent dualistes : les prototypes y sont définis à partir d'une moyenne de traits (propriétés)

pondérés (ou bien à partir de traits représentables). Rosch elle-même a d'ailleurs proposé une définition de la typicalité à partir de «traits empiétants». Ces approches, que l'on peut qualifier de statistiques, conduisent à des paradoxes tels que l'«homme moyen» de Quetelet (1985) : dans une population définie par plusieurs caractéristiques mesurables, on considère un «être moyen» dont les caractéristiques sont les moyennes de celles de la population. Un tel être (qui correspondrait à un prototype défini comme une moyenne calculée sur des traits numériques) n'existe que si la population décrite est une partie connexe de l'espace vectoriel engendré par les mesures de traits ! Une autre approche, développée par J.-D. Kant (1996), consiste à considérer des disjonctions de conjonctions «typiques» de traits (ce qui évite l'impasse des conditions nécessaires et suffisantes) et à rendre cette approche opératoire à l'aide d'une implémentation connexionniste.

Ces études sur la catégorisation mettent en lumière nombre de problèmes : (i) la genèse des catégories et des concepts ; (ii) l'émergence et l'(auto-) organisation ; (iii) les relations subtiles entre intension et extension ; (iv) les notions d'intériorité et d'extériorité et la capacité d'un système symbolique à engendrer la «logique» qui marque son fonctionnement ; (v) la discrimination des objets qui fondent une catégorie selon un niveau de représentativité (on parle, dans ce cas, de gradient de typicalité). Une approche naïvement ensembliste et taxinomique laisse ces problèmes non résolus et en crée d'autres qui viennent les bruyers artificiellement : (i) le recouvrement partiel des classes c'est-à-dire la possibilité pour un objet d'être instance de plusieurs classes ; (ii) l'amalgame entre degré de représentativité et sa (mauvaise) traduction en termes de degrés d'appartenance ; (iii) le problème de la différence de statut théorique entre lien instanciel et lien taxinomique ; (iv) le problème des exceptions qui pose celui de la différence entre «attribut» et «propriété» ; (v) la prise en compte de l'incertitude dans le processus de classification et, donc, l'introduction de degrés d'incertitude, qu'il convient de distinguer soigneusement des degrés de typicalité ; (vi) le fait qu'il n'y a pas classification en soi mais que la classification résulte d'un processus dépendant du contexte, du point de vue de celui qui procède à la classification et des objectifs de la classification ; (vii) l'apparente contradiction entre le désir de procéder à une taxinomie et le fait que celle-ci se fonde sur la donnée d'une ressemblance.

Les formalismes mathématiques utilisés pour rendre compte de structures catégorielles s'apparentent plus à des schémas de

représentation de données qu'à des modèles de catégorisation. Dans le cadre de l'approche classique, les arbres hiérarchiques des taxonomistes ont fourni des schémas de représentation de catégories (Miller, 1969, Friendly, 1979), les correspondances et treillis de Galois introduits par Riguet en 1948 et popularisés par Wille, à partir de 1982, sous les appellations respectives de « contextes » et « treillis de concepts », permettent de représenter, dans un cadre plus général que les hiérarchies, des classes définies par des conditions nécessaires et suffisantes. Un contexte est un triplet (X, X', C) où X et X' sont des ensembles finis respectivement d'objets et de propriétés et C une correspondance (ou relation) observée entre ces ensembles (xCx' signifie que l'objet x possède la propriété x'). Ainsi que l'a noté Duquenne (1986), cette notion prend bien en compte deux dimensions descriptives : extensionnelle (les objets) et intensionnelle (les propriétés). Cependant la correspondance C induit des opérations de fermeture sur les ensembles de parties de X et de X' et ce sont ces dernières qui établissent la dualité : les treillis des intensions et des extensions sont anti-isomorphes. La construction du treillis de Galois d'une correspondance (ce treillis est constitué des rectangles maximaux du tableau 0/1 associé à un contexte, ordonnés par inclusion de ligne, respectivement de colonnes) devient la clé d'une approche formelle pour une dualité objets/propriétés.

Ces schémas sont cependant incapables de figurer des classes avec leur gradient de représentativité et bon nombre d'auteurs préfèrent utiliser des représentations sous forme de cartes obtenues par des techniques de « *multidimensional scaling* » (Carroll et Arabie, 1980). Plutôt que de mobiliser des traits caractéristiques des données, les représentations issues du « *multidimensional scaling* » favorisent une interprétation en termes de dimensions (les catégories deviennent des nuages de points dans un même espace vectoriel subjectif dont la structure, en général euclidienne, est inspirée de celle des espaces perceptifs et muni d'une base dont les axes sont interprétés comme les variables pertinentes permettant de paramétrer les données). C'est ainsi qu'un débat « dimensions contre traits caractéristiques » a alimenté (et alimente toujours) la communauté psychométrique (Henley, 1969, Rips *et al.*, 1973, Shoben, 1976, Shepard *et al.*, 1972, Shepard, 1974, 1982, Pruzansky *et al.*, 1982). Les « *feature models* » ont été développés à partir du modèle du contraste de Tversky (1991) qui plonge ses racines dans des travaux psychométriques (Restle, 1959) et trouve actuellement d'intéressants échos en intelligence artificielle (Vakarelov, 1991). Le modèle du contraste rend compte, entre autres, des effets de typicalité sur la similitude perçue entre différents stimuli.

Il cherche en particulier à éclaircir et à formaliser l'idée de gradient de représentativité. Les représentations arborées offrent des versions simplifiées du modèle du contraste et une description des catégories permettant la comparaison aussi bien intra-catégorielle (via la distance arborée qui rend ces éléments plus ou moins représentatifs) qu'inter-catégorielle (la structure arborescente). Pour plus de détails et des exploitations de cette approche, voir Abdi (1985), Abdi *et al* (1984, 1989), Barthélemy (1991), Barthélemy et Guénoche (1991), Ténin (1996).

3.2. Catégories et catégories

Hérités des taxonomies de la biologie systématique, les schémas généralement utilisés pour l'organisation catégorielle ne laissent apparaître entre les objets aucune relation autre que l'appartenance à une même catégorie (c'est la fameuse relation « *is-a* » des représentations sémantiques en intelligence artificielle). Or, il semble que ce soit des relations diversifiées entre les objets (identification, appartenance, inclusion, ingrédience, partie-tout, attribut-valeurs, ...) qui contribuent à l'émergence des catégories, des concepts et des systèmes symboliques.

Comment, dans un tel cadre, concilier le caractère purement local de l'organisation interne d'une catégorie, fondé sur la typicalité et la ressemblance, et la rigidité globale d'une organisation intercatégorielle, sachant que le tout est piloté par la notion d'air de famille, dont on ne sait pas comment elle peut s'instancier, de manière cohérente, dans chacun de ces niveaux ? Comment concilier le caractère générique des prototypes et leur caractère agrégatif, tout en maintenant leur représentativité ? La géométrie algébrique nous enseigne que généricité et représentativité ne vont pas nécessairement de pair : les points génériques d'une variété algébrique sont, en quelque sorte, situés à l'« extérieur » de la théorie algébrique où l'on exprime cette variété. Enfin, comment gérer le rapport intension-extension au sein d'une approche en rupture avec une analyse radicalement intensionnelle des catégories ? Ces questions sont la trace d'un malaise mathématique : l'approche de Rosch (ou celles qui préconisent par rapport à celle-ci quelques inflexions) résisterait-elle à une analyse formelle de ses « axiomes » ?

Il convient, selon nous, d'envisager une démarche visant à articuler des *structures localement définies* avec un certain *constructivisme*. Trois approches sont envisageables.

Ehresmann et Vanbremeersch (1987) proposent une modélisation des systèmes vivants -- les systèmes hiérarchiques évolutifs -- basée sur la théorie mathématique des catégories. Ils utilisent de manière non triviale et non métaphorique deux notions centrales de la théorie : les limites inductives et les foncteurs. En simplifiant à l'extrême, considérons des objets (neurones, organismes, agents sociaux, ...), liés par des relations. Pour un système d'objets A_i liés, la limite inductive est un objet de la catégorie qui « recolle » les objets A_i en tenant compte de leurs liens : « elle sera caractérisée (à un isomorphisme près) par la condition : pour tout autre objet C de la catégorie, les liens de la limite vers C correspondent biunivoquement aux familles de liens individuels des objets A_i vers C "compatibles" avec les liens (...) entre les A_i ». La limite inductive joue donc le rôle de représentant du système : « Une construction de ce genre semble intervenir dans la catégorie des neurones d'un animal (...) : certains circuits neuronaux souvent excités recrutent un neurone particulier qui les représente (...). Par exemple un objet familier sera représenté par un neurone pilote (...), dont l'excitation suscite la reconnaissance de l'objet. » Un cas particulier de limite inductive est la somme des objets, le système étant alors considéré sans les liens internes. La relation qui existe entre la limite inductive (recollant le système avec ses liens) et la somme est

une mesure de l'émergence qui caractérise le *tout* comme irréductible à ses *parties*.

Un système ne possède pas nécessairement une limite inductive. On peut alors adjoindre à la catégorie de départ, si elle vérifie quelques bonnes propriétés, des limites par des constructions classiques comme la *complétion libre*. Cette opération acceptable dans le domaine du « vivant », où il peut y avoir *création* d'objets nouveaux, doit être questionnée dans le cas des représentations symboliques. On construit très aisément un modèle de système hiérarchique où les objets du niveau k sont les limites des systèmes d'objets de niveau $k-1$. Ce modèle généralise, en quelque sorte, les arbres hiérarchiques des taxonomies en considérant les classes, non pas comme des réunions d'objets ou de classes d'un niveau inférieur, mais comme induites par des relations entre les objets.

« Dans des systèmes autonomes tels que les systèmes biologiques ou sociologiques, l'évolution est sous la dépendance d'une hiérarchie d'organes de régulation internes qui analysent les contraintes internes et externes, et cherchent à modifier en conséquence l'état du système et son rapport à l'environnement (...). Dans un système évolutif, un tel organe de régulation sera modélisé par un sous-système évolutif, appelé système régulateur, dont les composants seront appelés agents. » (Ehresmann et Vanbremeersch, 1991). Cet organe a « un triple rôle : observationnel, il recueille des informations sur l'état interne du système et sur les communications en provenance de l'extérieur ; organe décisionnel, il participe au choix d'une stratégie, compte tenu de la situation et des actions possibles ; organe de commande enfin, il coopère à la mise en œuvre de cette stratégie par l'intermédiaire des effecteurs qu'il contrôle. » (Ehresmann & Vanbremeersch, 1991). Le système régulateur construit une représentation de l'état interne du système qui sera appelée *le paysage du système*. Une trop grande variation du paysage sous l'effet de contraintes internes ou externes entraîne une transformation du système par l'action des agents ayant préalablement élaboré une stratégie. Techniquement, le système régulateur et le paysage sont des catégories : le système régulateur est une catégorie en tant que système évolutif et le paysage est la catégorie dont les objets sont *les perspectives*.

La piste « limites et représentativité » inspirée des travaux d'Ehresmann et Vanbremeersch a été suivie par Barthélemy et Boldini (1994) qui, dans le cadre simplificateur des structures ordinales, ont systématiquement étudié toutes les modalités d'association

diagramme/limites dans la perspective d'une construction des foncteurs entre des catégories d'objets et des catégories de traits.

En second lieu, l'approche piagétienne mérite d'être examinée. Analysée par Piaget dans le cadre de l'épistémologie génétique, la théorie des catégories se révèle être un outil prometteur pour la formalisation de mécanismes fondamentaux de la psychogenèse. Elle est, sans doute, le formalisme le mieux à même d'exprimer rigoureusement les thèses piagésiennes. On peut envisager un traitement systématique, dans le cadre de la théorie mathématique des catégories, des thèses de Piaget et de ses collaborateurs sur la formation des catégories au sens psychologique, telles qu'elles sont exposées dans l'ouvrage *Morphismes et Catégories* (Piaget et al 1990), et approfondies par Boldini (1995).

Conformément au paradigme de l'épistémologie génétique, l'étude des «deux fonctions principales de la raison en ses créations : comparer et transformer» se fait à la fois au sein des mécanismes psychologiques les plus élémentaires et au sein des théories scientifiques les plus abstraites. Dans ce cas précis, c'est la théorie mathématique des catégories qui est objet de l'investigation épistémologique et outil pour l'étude psychogénétique. Cette fascinante circularité caractéristique de l'approche génétique a peut être trouvé là son cadre d'expression le plus naturel. Cependant les mathématiques de la psychogenèse restent à faire ; la mise en œuvre, par Piaget, des notions primitives de la théorie mathématique des catégories est suffisamment fructueuse pour nous laisser espérer qu'une mathématisation plus complète est possible. C'est Henriques qui nous fournit les clés pour un traitement formel des mécanismes de construction mis à jour par Piaget. Comparer et transformer sont les deux fonctions principales de la raison, mais celles-ci ne relèvent pas de la même approche mathématique. Comparaisons et transformations sont des opérations de nature différente sur le plan psychologique, mais elles tendent à se confondre dans les théories scientifiques les plus abstraites comme les théories algébriques où les morphismes sont à la fois instruments de transfert et instruments de comparaison de structures. Les phases principales des processus cognitifs mis en œuvre et les concepts opératoires qui en rendent raison sont les suivants.

Les objets sont tout d'abord assimilés aux schèmes d'action du sujet, les correspondances qui se tissent à ce niveau sont qualifiées *d'intramorphiques*, les objets n'ont pas d'identité individuelle et se correspondent par le seul fait d'être assimilés au même schème. Ensuite, les actions du sujet entraînent des coordinations entre

schèmes, les correspondances qui se tissent à ce niveau sont qualifiées *d'intermorphiques*. Enfin le sujet se met à opérer sur les morphismes en généralisant les opérations qu'il effectue sur les objets, les compositions de morphismes qui en résultent permettent la construction de la catégorie en tant que structure fermée pour la composition et dans laquelle la réversibilité opératoire est possible. Ces nouvelles correspondances qui se déduisent des opérations du sujet sont qualifiées de *transmorphiques*. Si chaque niveau de correspondance est le préalable génétique du niveau supérieur, la présentation grossière qui précède ne doit pas laisser penser que ces constructions se hiérarchisent strictement ; en fait il s'agit de trois processus relativement autonomes en interaction permanente et en intégration croissante.

Ces trois étapes se retrouvent dans tout processus cognitif de construction catégorielle, qu'il s'agisse de la psychogenèse ou de thématizations scientifiques et formelles. Nous pouvons préciser et motiver les caractérisations de la section précédente en mettant en évidence les invariants qui sont dégagés par l'activité cognitive. Ce sont ces constructions d'invariants, très bien décrites par Henriques, qui vont nous permettre de formaliser l'ensemble du processus. Henriques distingue deux types d'invariants : *les invariants de remplacement* et *les invariants de transformation*. Les premiers apparaissent lorsque les objets sont assimilés à un même schème, les seconds apparaissent lorsque le sujet structure ses transformations et opère sur les morphismes. On voit donc que les invariants de remplacement et de transformation apparaissent respectivement aux niveaux intramorphiques et transmorphiques.

L'ensemble des objets et des morphismes relatifs à un schème donné forme alors une *précatégorie*. Ce qui caractérise une précatégorie c'est son ouverture : ouverture à l'adjonction de nouveaux objets, de nouveaux morphismes, et à la composition des morphismes, cette dernière opération étant encore très limitée. Le sujet s'engage ensuite dans des activités de coordination puis de transformation des schèmes opératoires. Dans le premier cas, il y a élaboration des morphismes du niveau intermorphique, dans le second, il y a élaboration des morphismes qui sont les traces des transformations. Ce dernier point, tout à fait fondamental, est à l'origine des limitations du formalisme catégorique ; nous sommes en présence de deux ordres distincts relativement autonomes : celui des comparaisons et celui des transformations. On distinguera deux types de morphismes à ce niveau : *les morphismes cotransformationnels* qui suivent les transformations, et *les morphismes protransformationnels* qui peuvent servir de projet

à des transformations. Les invariants de transformation apparaissent lorsque les transformations mises en œuvre par le sujet forment une structure, c'est-à-dire lorsqu'il y a composition et transformation inverse, ceci permettant la construction des transformations identiques qui permettent de dégager les invariants de transformation.

Comme les systèmes de transformations sont liés à l'acte et au temps, il ne peut y avoir de saisie totale du système que par la « *composition intemporelle des transformations virtuelles* ». Ceci n'est rendu possible que par l'interaction entre morphismes et transformations ; le système des morphismes étant appréhendable dans sa globalité, c'est par les opérations sur les morphismes, en tant que marques des transformations passées, que se construiront les transformations identiques marquées à leur tour par des endomorphismes (morphismes de l'objet sur lui-même). Pour plus d'informations on pourra consulter Boldini (1994, 1995).

Enfin, la piste relative à la théorie des topoï a été également poursuivie par Boldini (1995). Le postulat fondamental de cette approche est que les objets sur lesquels portent la connaissance et le discours sont cognitivement structurés. Cette structuration est locale et différente de la classe au sens ensembliste du terme. A ces objets sont associés des traits, des descripteurs, définis localement, c'est-à-dire *localement pertinents*. Dans ces conditions, la structure mathématique qui s'impose est celle de *faisceau*. Ceci oblige à considérer l'espace primitif des objets comme un espace topologique, et les traits comme des fonctions booléennes définies localement. Un faisceau représente donc une structuration de la connaissance relative aux objets donnés. La théorie mathématique des catégories, et particulièrement la *théorie des topoï*, font une place essentielle aux catégories de faisceaux et permettent d'obtenir des résultats remarquables. En particulier, à la catégorie des faisceaux sur un espace topologique, on peut associer de manière intrinsèque un langage logique qui la décrit : *son langage interne*. Le fait que cette logique soit *nécessairement* intuitionniste souligne le caractère fondamentalement constructif de cette approche. On est donc amené à reprendre l'ensemble de cette démarche, pour éventuellement l'appliquer à notre cas particulier : la catégorie des faisceaux de fonctions booléennes localement définies. Bien que cette catégorie ne forme pas un topos, tous les objets qui caractérisent les topoï parmi les catégories s'interprètent aisément dans cette dernière, et le fait qu'elle soit, dans tous les cas, une sous-catégorie d'un topos lui permet d'hériter de la logique intuitionniste de celui-ci. A ce stade d'élaboration rien n'est dit sur la nature de la topologie adoptée. Notre conviction est que les notions d'objets et de propriétés émergent

conjointement d'un réseau de relations qui forme la structure topologique de l'espace de référence.

3.3. Contradictions

Une facette essentielle de la catégorisation est celle de la *contradiction* telle qu'elle apparaît, en particulier, dans l'étude de la *typicalité* (et de l'*atypicalité*).

La contradiction, selon Aristote dans *Catégories* et le livre *D* de *Métaphysique*, c'est-à-dire l'« opposition comme affirmation et négation », s'applique à des *énoncés*, et de deux énoncés l'un est vrai, l'autre faux, indépendamment de la nature des *objets* auxquels ces énoncés s'appliquent. Aristote, en effet, fait remarquer que si d'une autruche il est vrai d'affirmer oiseau, il est aussi vrai d'affirmer non-oiseau pourvu que l'autruche soit animal et non pas non-animal, de telle sorte que dire d'un objet qu'il est oiseau et non-oiseau ne modifie en rien la conclusion d'un syllogisme quelconque. Cela n'entraîne nullement qu'on puisse admettre un énoncé et sa contradiction. Ce principe de non contradiction se traduit formellement, sur le plan syntaxique, par l'involutivité de la négation ($p \not\equiv \sim\sim p$ et $\sim\sim p \not\equiv p$) et, sur le plan sémantique, par la bivalence et la symétrie entre le « vrai » et le « faux », entre présence et absence. Les divers affaiblissements de la négation classique ne remettent pas fondamentalement en cause l'édifice du principe de non contradiction aristotélicien : les négations affaiblies n'acquièrent une valeur opératoire qu'à la faveur d'une spécification, d'une détermination, des *objets* sur lesquels portent les *énoncés* qui, eux, obéissent au principe de non contradiction.

La négation intuitionniste, non seulement ne laisse aucune place à la contradiction, puisque la vérité intuitionniste est plus forte que la vérité classique, mais de plus ne se démarque de la négation classique que si elle est appliquée à des objets du discours dont le domaine est spécifié. En effet, en logique intuitionniste du premier ordre, le tiers exclu garde sa validité pour peu que les énoncés s'appliquent à des objets dans un domaine fini ; ce n'est donc que grâce à une spécification du domaine des objets que la distinction entre négation intuitionniste et négation classique apparaît.

La négation paraconsistante (da Costa, 1974) opère, il est vrai, une rupture apparemment radicale, puisque la « logique » paraconsistante entraîne l'invalidité de $\sim(p \rightarrow \sim p)$. Il s'agit, dans l'esprit de da Costa, de donner corps à la « thèse de Hegel » selon laquelle une proposition et sa négation peuvent coexister. Toutefois, pour ne pas trivialisier le système ainsi conçu (un système dans lequel toute formule bien

formée serait un théorème, en application de l'*ex falso sequitur quodlibet*), da Costa est contraint d'adjoindre à la négation paraconsistante une seconde négation qui, elle, opère classiquement, la distinction entre les deux négations se faisant, une fois de plus, à la faveur d'une spécification des objets auxquels s'appliquent les énoncés.

Ainsi, à la question de savoir si la contradiction peut s'appliquer aux faits du monde naturel et si, du point de vue épistémologique, elle peut être introduite dans les contenus mêmes de pensée, il a toujours été répondu, à l'exception des travaux d'Héraclite et de Hegel, par la négative.

L'œuvre d'Héraclite occupe une place singulière dans la philosophie présocratique (Dumont, 1994). Mais, ce qui rend la pensée héraclitéenne actuelle, c'est sa théorie du combat, thème essentiel de sa vision du monde. Pour Héraclite, la nature aime les contraires et sait en opérer la synthèse pour parvenir à l'harmonie. Mais cette harmonie se conserve au prix d'une tension entre les opposés : « Peut-être la nature aime-t-elle les contraires et sait-elle en dégager l'harmonie, alors qu'elle ne s'intéresse pas aux semblables, c'est ainsi qu'elle unit le mâle à la femelle et ne rapproche pas les êtres de même sexe [...]. Il apparaît aussi que l'art, en imitant la nature, fait de même ». A cette citation, extraite de l'ouvrage *Du Monde*, il faut adjoindre celles que l'on retrouve dans les fragments 10 et 8 et qui conduiront, à la suite de Parménide et de Socrate, à taxer l'Ephésien d'hermétisme : « Les unions sont des touts et des non-touts, rapprochement et différence, accord et désaccord » (fragment 10) ; ainsi, « les contraires s'accordent et la belle harmonie naît de ce qui diffère » (fragment 8). « La route qui monte et celle qui descend sont une et identique » (fragment 60). Ainsi, l'identité et la différence ne peuvent-elles se comprendre que dans un rapport « dialectique ». Le trait saillant de la philosophie héraclitéenne du logos est, semble-t-il, moins le fait, comme l'a retenu la tradition, qu'elle constitue une philosophie du tragique et du déchirement que le fait qu'elle milite en faveur d'une philosophie de l'être qui s'accommode de la co-présence des contraires.

Une autre dimension de l'héraclitéisme à laquelle celui-ci est souvent réduit, c'est la philosophie du devenir. Hegel verra en Héraclite le précurseur de l'idée qu'il n'y a d'être que du devenir. Il convient, toutefois, lorsque l'on invoque le devenir chez Héraclite, de ne pas oublier sa vision du logos. La vision héraclitéenne du monde s'appuie sur la transformation incessante des choses qui passent et des êtres qui meurent, le devenir étant une perpétuelle métamorphose des substances. Mais nous trouvons dans le devenir le champ où s'affrontent

les contraires : « les choses froides se réchauffent, le chaud se refroidit, l'humide se dessèche et l'aride s'humidifie » (fragment 126). Il y a donc, dans le devenir, une co-présence de l'affirmation et de la négation : « Nous descendons et ne descendons pas les mêmes fleuves, nous sommes et nous ne sommes pas » (fragment 49).

C'est à cette philosophie héraclitéenne du logos que va s'attaquer Parménide pour qui seul ce qui est vrai peut être véritablement exprimé ; toute proposition fautive contient un nom faux, c'est-à-dire un nom sans détenteur : l'être est et le non-être n'est pas. L'être étant ce qui est ne peut pas être nié, même partiellement, si bien que tout recours au devenir (de l'être) se trouve éliminé. L'opposition entre l'éléatisme et l'héraclitisme se fonde donc sur deux philosophies antagonistes de l'être et du devenir. Pour Parménide, l'être est, il est inengendré et impérissable, sans fin. Il est indivisible, immuablement fixé au même endroit. Il ignore donc la dispersion et le rassemblement. Il est éternel et immobile et ignore le temps et l'espace.

Nulle trace de la vision héraclitienne du logos n'est décelable dans les travaux consacrés à la logique depuis les socratiques. La vision parméniennienne de l'être et du non-être irrigue, depuis Aristote, la totalité des œuvres de logique. Elle va inspirer toute la théorie aristotélicienne du vrai et du faux, de la négation et de la non contradiction : « il est faux de dire de ce qui est qu'il n'est pas ou de ce qui n'est pas qu'il est, et il est vrai de dire de ce qui est qu'il est et de ce qui n'est pas qu'il n'est pas ». La théorie aristotélicienne des catégories lèguera à la logique un cadre où la référence à la prédication de la forme « x est / n'est pas y » est quasi exclusive et qui conduit à traiter toutes les propositions selon deux oppositions binaires : affirmation - négation et quantificateur universel - particulier. La théorie aristotélicienne du vrai et du faux, en s'appliquant aux « noèmes » et non aux faits linguistiques va laisser sans réponse la question de savoir si « être homme » et « être dit homme » sont des énoncés synonymes et privilégier une sémantique de l'« être » au détriment d'une sémantique du « dire ».

C'est également dans le cadre de la philosophie parméniennienne du logos que les logiciens médiévaux vont, sur des bases aristotéliciennes, concevoir le principe de l'*ex falso sequitur quodlibet*, principe qui, dès lors, se placera au cœur de la logique et qui conduira au concept de consistance.

La logique de Port-Royal, dans laquelle il faut voir la matrice de la mathématisation de la logique, subira la même influence. La *Logique* d'Arnaud et Nicole et la *Grammaire* d'Arnaud et Lancelot fourmillent, malgré leur nouveauté épistémologique, de points de vue aristotéliciens (et augustiniens).

Aucune des approches, modernes, visant à affaiblir ou à généraliser la logique classique, ne remettra vraiment en cause le principe de non contradiction hérité de la philosophie parméniennienne et aristotélicienne de l'être et du non-être. La seule tentative d'échapper à une telle vision et de renouer avec l'héraclitisme est à mettre au crédit de Hegel. C'est ce que nous enseigne la Doctrine de l'Essence de *La Science de la Logique. L'identité*, selon Hegel (Harris, 1987), est auto-relation, non pas en tant qu'immédiate mais en tant que réfléchie. En tant qu'immédiate, elle ne serait que l'être ; en tant qu'étant réfléchi, elle est la structure totale de l'être avec tous les moments saisis comme l'identité dont ils ne sont que les moments : dans l'être, pris comme identité, ces divers moments sont absorbés et nous n'avons accès qu'à une identité pure et simple. La véritable identité est celle qui inclut la différence, alors que l'identité abstraite est exclusive de toute

différence, soit qu'elle laisse hors d'elle-même « les traits multiples du concret », soit qu'elle confond toute diversité en une seule qualité. La véritable identité est l'identité du tout, ou le concept, qui est une identité des différences. Elle est médiatisée en tant qu'essence et elle est l'identité aboutissant à la diversité. Le fondement de l'identité et de la diversité est la totalité du concept, le système se déployant comme une échelle dialectique. Telle que l'envisage ce que Hegel appelle l'entendement, l'identité est donc vide de toute diversité. Elle est donc une relation à un seul terme, ce qui la met en contradiction avec soi-même : une relation doit avoir deux termes. Dans «*A est A*», le premier *A* doit être distingué du second, bien que la proposition les identifie. S'ils étaient identiques, on ne saurait formuler la proposition. Ainsi l'acception classique de l'identité est-elle contradictoire et entre-t-elle en conflit avec le principe de non contradiction. Identifier, c'est donc aussi distinguer. Pour s'identifier soi-même, soit on opère une distinction à l'intérieur de soi entre ce que l'on reconnaît comme propre et ce que l'on rejette, soit on marque une distinction entre soi et les autres.

Hegel souligne, par ailleurs, que des choses différentes ne peuvent être appréhendées comme telles que si on les compare et les relie. La différence est, elle-même, une relation dont les deux termes ne sauraient être différents au point qu'ils n'entretiennent aucune relation. De plus, s'ils sont différents, c'est sous un certain point de vue qui est commun aux deux. Ils relèvent d'un tout, unique, saisi dans son ensemble, à l'intérieur duquel ces distinctions sont faites.

On ne peut faire des comparaisons qu'en termes de détermination des choses comparées. Lorsque deux choses ont les mêmes caractéristiques, nous disons qu'elles sont semblables ; lorsque leurs caractéristiques diffèrent qu'elles sont dissemblables. Similitude et dissimilitude sont donc des sous-catégories de l'identité et de la différence. Mais ces deux termes doivent aussi avoir une base commune. La signification du principe de Leibniz selon lequel deux choses ne sauraient être indiscernables, repose, selon Hegel, sur l'intériorité de toutes relations et toutes distinctions : si chaque chose « est ce qu'elle est et non pas autre chose », on ne peut pas dire, en toute rigueur que toute chose est différente. Chaque chose est ce qu'elle est et sa différence par rapport aux autres choses ne peut être appréciée que sur la base de comparaisons qui permettent, à partir de la nature intérieure des choses comparées, de conclure à leur similitude ou à leur dissimilitude. Ainsi la différence est-elle déterminée et intrinsèque et c'est en ce sens que l'on peut affirmer qu'il est impossible que deux choses sont indiscernables. La similitude est un

élément de l'identité dans la différence ; la dissimilitude est un élément de la différence dans l'identité. Ces deux termes se réfèrent donc l'un dans l'autre.

L'identité, en opposition à ce dont elle diffère, est positive, en tant qu'opposé à négative. Ces deux termes sont, une nouvelle fois, mutuellement dépendants et chacun est ce qu'il est en *opposition* avec l'autre. Ils s'excluent l'un l'autre, mais, en s'interdéfinissant, ils sont mutuellement compatibles. Ainsi, à l'instar de «*A est A*», «*A est non non-A*» se contredit-il. De plus, ces deux énoncés entrent en contradiction, puisque le premier dit que *A* n'est rien d'autre que soi-même, le second qu'il est quelque chose d'autre, en l'occurrence un opposé à non-*A*. Il y a donc contradiction à maintenir, à la fois, que *A* est *A* et qu'il n'est pas non-*A*, et il est impossible de jouir de deux qualités différentes, au même moment et dans la même perspective.

La *Science de la Logique* est un traité de la nature et de la fonction dialectique de la négation, du principe de différenciation qui distingue l'identité de la différence. Deux choses distinctes, en tant que mutuellement positive et négative, sont opposées, mais, en tant que mutuellement dépendantes, sont complémentaires et *compatibles*. Les distinctions classiques entre la contrariété et l'opposition ignorent cette dialectique et, fondées sur les notions abstraites d'identité et de différence, elles conduisent à la doctrine du carré des oppositions dans la logique classique.

Les critiques faites à Hegel semblent se fonder sur une incompréhension de ce qui constitue l'essentiel de la dialectique hégélienne. Au nom de l'inviolabilité du principe de non contradiction, les exégètes n'ont retenu que ce qui, dans l'œuvre de Hegel, s'attaque à ce principe, sans comprendre que le point de vue hégélien est moins une condamnation du principe qu'une critique de ses conséquences, notamment dans sa relation avec le principe d'identité.

La logique *pseudo-consistante* (De Glas & Jacquet, 1995, De Glas 1996, Boutot, 1994) peut être vue comme une tentative de formalisation des conceptions héraclitienne et hégélienne du logos. La négation pseudo-consistante rompt avec la négation classique en ceci qu'elle est plus *faible* que la négation classique. Plus précisément, si $\sim(p \sim p)$ est un théorème pseudo-consistant, en revanche $p \sim p \not\equiv \perp$ est non valide. Le rejet de l'*ex falso sequitur quodlibet* permet de rompre partiellement avec la non contradiction, donc avec la négation classique. Cette rupture opère sur les énoncés, non pas sur les objets du discours. Le système pseudo-consistant confère à la contradiction le statut d'outil de pensée effectif pour la représentation

scientifique des phénomènes. Sans rejeter totalement la non contradiction, mais en l'affaiblissant grâce au rejet de l'*ex falso sequitur quodlibet*, la pseudo-consistance autorise la prise en compte de contradictions locales et, partant, de rompre avec la vision aristotélicienne de l'être.

L'affaiblissement de la non-contradiction, et le rejet de l'*ex falso sequitur quodlibet*, puise sa légitimité dans le caractère en devenir des objets du discours : une assertion quelconque sur des objets en devenir ne saurait être, de façon définitive, vraie ou fausse. Plus précisément, si l'on donne à «*p* est fausse » le sens constructif, c'est-à-dire si «*p* fausse » signifie «*p* est prouvée fausse », alors «*p* vraie » signifie «*p* est non prouvée fausse », c'est-à-dire «*p* est possiblement (ou plausiblement) vraie ». Selon une telle interprétation, «*p* et non-*p* entraîne le faux » signifie que «*p* est non prouvée fausse » et « non-*p* est non prouvée fausse » sont mutuellement exclusives. Accepter que tel n'est pas le cas de toute proposition *p* vaut rejet de « $(p \rightarrow \sim p) \rightarrow \perp$ ». Un tel rejet ne relève pas du désir d'échapper à l'alternative vrai-faux mais de la nécessité de briser la symétrie entre le vrai et le faux et de rompre avec un principe d'omniscience. Outre une sémantique globale, rendant complète la syntaxe pseudo-consistante, une sémantique locale qui exploite l'idée d'états de la connaissance variables dans le temps, a pu être développée. A chaque instant, l'état de la connaissance consiste en un ensemble de faits plausibles (en ce sens que leur fausseté n'a pas été établie). Une proposition est possiblement vraie en l'état *i* si tel était le cas en tout état antérieur à *i*.

L'étude des fondements catégoriques de la pseudo-consistance montre que la théorie des topoi offre un substrat « géométrique » à la logique pseudo-consistante et permet d'envisager l'étude des liens entre logique intuitionniste et logique pseudo-consistante : l'une et l'autre peuvent être vues comme émergeant d'un substrat toposique.

3.4. Catégorisations grammaticales et lexicales

A titre d'exemple, examinons les catégories grammaticales et lexicales du langage et, plus particulièrement les catégorisations aspecto-temporelles.

La sémantique cognitive (Desclés, 1991) articule trois domaines : les domaines des primitives, le domaine des catégorisations, le domaine des champs lexicaux. Les deux premiers sont de nature représentationnelle, le dernier est accessible à une observation et une expérimentation plus directe. L'objectif général est d'évaluer la pertinence cognitive des primitives et des catégorisations sous-

jaçentes aux structurations des champs lexicaux. Il s'agit de répondre, entre autres, aux questions suivantes :

(i) Quelles sont les primitives ? Ont-elles une pertinence cognitive ? Sont-elles fondées par d'autres activités cognitives (comme la perception visuelle et la planification) ? Quelles sont les opérations définies sur les primitives et les modes de composition des schémas cognitifs ?

(ii) Certaines catégorisations reflètent-elles des mécanismes cognitifs ?

(iii) Les catégorisations et les schémas cognitifs sont-ils responsables des organisations lexicales rencontrées dans la langue ?

(iv) Existe-t-il des regroupements typologiques des prédicats ? Ces regroupements sont-ils associés aux schémas cognitifs ?

L'approche cognitive permet de mieux poser le problème en articulant au sein d'une architecture différents niveaux de représentation : représentations linguistiques opérées par les langues, représentations langagières (opérations, relations, schémas mis en œuvre par l'activité du langage), représentations cognitives articulées avec les représentations précédentes et les autres activités non langagières (perception, mémorisation, planification, ...). Le problème général est le suivant : comment les catégorisations opérées par les langues (catégorisations grammaticales et lexicales) peuvent-elles avoir un *ancrage dans les catégorisations opérées par la perception visuelle et par l'action planificatrice*, et comment cet ancrage du langagier sur le perceptif visuel permet-il l'émergence d'une prédication ? Ce problème est l'un des points de convergence fondamentaux entre les approches symboliques et les approches morphodynamiques (voir section 4).

La théorie post-frégéenne de la prédication, formalisée par le langage des prédicats du premier ordre, est fondamentalement inadéquate à l'analyse de la prédication observée dans le fonctionnement linguistique. Elle ne fournit pas, nous l'avons vu, une théorie sémantique du « sens » qui analyserait la signification des prédicats linguistiques. Elle ne permet donc pas d'articuler les opérations générales de prédications opérées par les langues avec, d'une part, les organisations syntaxiques (en noms, verbes, adjectifs, prépositions) plus directement observables et, d'autre part, les représentations cognitives. Il est, de plus, impossible d'étudier les interactions entre l'activité langagière et les activités cognitives de perception et d'action.

Par ailleurs, la réduction du prédicatif au spatial opérée par le localisme est très certainement erronée. Si la perception des positions est représentable dans un référentiel purement spatial, les mouvements spatiaux nécessitent, eux, des référentiels spatio-temporels. La perception d'un mouvement et d'un changement nécessite donc non seulement une organisation statique (repérages et orientations) de l'espace mais également une organisation cinématique (mouvements et changements) et dynamique (action) de l'espace. Les sciences cognitives nous invitent à penser ces problèmes par le biais de représentations, non pas autonomes, mais insérées dans des architectures qui articulent différents niveaux et types de représentations. Ce n'est donc pas une réduction qu'il faut viser mais

une interaction et une recherche de niveaux de compatibilités représentationnelles entre différentes activités cognitives, le langage étant l'une de ces activités (voir paragraphe 4.2).

Les primitives (les types élémentaires : entités individualisables, collectives, massives, lieux ; les opérateurs formateurs de types ; les relateurs : de repérage, de détermination topologique, de mouvement spatio-temporels, de changements d'état, ...) sont constitutives de catégorisations sémantiques très générales. Certaines de ces catégorisations ont un ancrage sur les catégorisations opérées à partir de la perception visuelle et de l'action. Une première catégorisation s'opère donc autour de l'opposition statique/cinématique/dynamique. Les primitives précédentes entrent dans des schèmes sémantico-cognitifs qui sont associés à des classes de notions différenciées entre elles par des opérations d'instanciation, d'emboîtement, ... Ainsi, par exemple, les notions prédicatives telles « entrer dans », « pénétrer dans », « aller dans », « descendre dans », ... relèvent-elles d'un même archétype cognitif qui caractérise le contrôle d'un mouvement spatio-temporel, ce dernier faisant passer de l'extérieur d'un lieu vers l'intérieur d'un lieu. Une deuxième catégorisation est celle relative à l'insertion des notions lexicales dans un discours : les catégorisations aspecto-temporelles font, par exemple, partie de cette catégorisation et se déploient à partir de la *trichotomie fondamentale* pour toute l'organisation aspecto-temporelle dans les langues : *état-processus-événement*.

La temporalité linguistique, c'est-à-dire la temporalité appréhendée au travers des langues, n'est pas une temporalité homogène, globale et linéaire où un « présent » viendrait couper un « futur » à venir, symétrique d'un « passé » réalisé. La temporalité linguistique ne doit pas, non plus, se confondre avec une temporalité phénoménologique telle qu'exprimée, par exemple, dans les *Confessions* augustinienne. En effet, la temporalité linguistique ne se confond ni avec une temporalité cosmique, ni avec une temporalité technologique, ni avec une temporalité d'un calendrier, ni avec une temporalité organisée par une chronique d'événements successifs. Elle ne s'organise donc pas selon une représentation sous forme d'un axe linéaire orienté à partir d'une origine vers le moment présent et se prolongeant dans le futur. La temporalité appréhendée par les langues met en jeu différents référentiels temporels qui peuvent être compatibles avec la structure chronologique des référentiels externes mais qui doivent en être distingués par des structurations spécifiques et originales.

La temporalité linguistique se déploie à partir d'un premier référentiel temporel local à partir duquel doivent être repérés d'autres référentiels temporels. Ce premier référentiel est le *référentiel énonciatif*, qui « prend en charge » un discours, soit de façon explicite, soit de façon plus ou moins médiatisée, soit, encore, en cherchant à

s'effacer complètement. Chaque énonciation est production d'un discours qui implique un énonciateur, lequel s'implique dans son énonciation. Certains énoncés expriment par des indices linguistiques explicites la présence de l'énonciateur ; d'autres tendent à faire oublier l'énonciateur qui a délivré cet énoncé. Il n'y a pas d'énoncé sans énonciateur.

Tout énoncé est l'expression d'un contenu qui s'inscrit dans un référentiel complexe organisé par et autour de l'énonciateur. Ce référentiel implique une disposition dialogique des interlocuteurs et l'existence d'une dimension spatiale et d'une dimension temporelle. Le référentiel temporel est donc la projection temporelle du référentiel complexe auquel se réfère l'énonciateur par le biais de son discours. Le référentiel temporel s'organise par rapport à l'énonciateur qui repère les situations relatives à son énonciation.

Dans le référentiel énonciatif, les situations exprimées par des relations prédicatives sont repérées par rapport à l'acte de parole de l'énonciateur. Cet acte de parole est un processus qui se déploie dans la temporalité externe à partir d'un début et est orienté vers une fin. Lorsqu'il est saisi dans son déploiement interne, on considère que le processus est en cours et est saisi dans son inaccomplissement, ce qui exclut la prise en compte du terme final du processus. L'intervalle de validité de l'acte de parole dépend du référentiel choisi. Dans le référentiel temporel externe, l'acte de parole est un événement qui se projette sur un intervalle temporel fermé de validité avec un début et une fin : l'intervalle de validité est un intervalle fermé qui inclut ces deux bornes. Dans le référentiel temporel énonciatif, l'acte de parole est un processus en cours qui se déploie sans qu'aucun terme ne soit pris en compte : l'intervalle de validité du processus énonciatif inclut la borne de début mais exclut toute borne de fin. Dans le référentiel externe, l'acte de parole est un événement dont l'intervalle de validité est fermé : l'instant de parole est un instant mobile ; le présent d'énonciation est instable et mouvant. Dans le référentiel énonciatif, l'acte de parole est un processus en cours, dont l'intervalle de validité est borné à gauche par un instant initial et à droite par l'« instant » d'énonciation T^0 , qui est exclu de cet intervalle.

Le référentiel énonciatif est divisé en deux parties : le réalisé et le non-réalisé. Le réalisé contient tous les instants avant T^0 , le non-réalisé contient T^0 et tous les instants après T^0 . Toutefois, le non-réalisé n'est pas symétrique du réalisé. Ce dernier est, en effet, organisé par des événements, des états et des processus qui se sont déroulés : c'est le réel de l'énonciateur. Le non-réalisé, lui, a une structuration plus

complexe puisque différentes possibilités sont envisageables. Lorsque le réalisé peut être conçu comme un demi-axe ordonné linéairement par les événements avant T^0 , le non-réalisé est structuré sous forme d'un arbre dont les branches représentent les différentes possibilités à venir. L'instant T^0 peut être considéré comme une origine fixe dans le référentiel énonciatif. Celui-ci oppose donc deux domaines qui relèvent de la modalité : le réalisé certain et le non réalisé ouvert sur les possibles. Le réalisé du référentiel énonciatif peut être orienté à partir de T^0 : d'où le temps de la reconstruction du réalisé, le temps retrospectif, le temps de la mémoire ; il peut être orienté en direction de T^0 : c'est le temps de la succession des événements dans le réalisé, le temps chronologique, le temps du vécu. L'instant T^0 ne peut être appelé « instant » d'énonciation que par abus de langage puisque l'énonciation engendre un processus qui se déploie dans le temps. De son côté, l'acte de parole n'est pas réduit à un seul instant : il demande une certaine durée pour se déployer. Lorsque le discours est proféré, l'acte de parole est en train de s'accomplir, mais n'est pas encore accompli. Cet acte de parole en cours n'est donc pas un événement ; il n'est pas non plus un état puisqu'il exprime une évolution. Toutes les phases de l'énonciation ne sont pas équivalentes entre elles : chaque phase est l'énonciation d'une parole.

Dans le référentiel narratif (ou non-actualisé), les situations exprimées par les relations prédicatives aspectualisées ne sont plus repérées par rapport à l'acte énonciatif, elles le sont de façon relative (liens anaphoriques). Les notions d'arrière-plan et de premier plan prennent alors une importance plus grande. Les repérages ne se font plus par rapport à la situation d'énonciation mais par des relations entre événements, entre événements et états, entre événements et processus. Certains marqueurs grammaticaux relèvent uniquement du référentiel narratif. Par exemple, le passé simple en français contemporain, est un indicateur absolu d'un fonctionnement narratif. D'autres marqueurs (passé composé et imparfait, par exemple) relèvent des deux référentiels mais y acquièrent des valeurs différentes.

Les structurations discursives et les valeurs sémantiques des temps morphologiques sont différents selon le référentiel dans lequel on se trouve. Dans le référentiel énonciatif, les structures rétrospective et prospective dominant. Dans une narration, repérée par rapport à l'acte énonciatif en cours, le passé composé est obligatoire et le passé simple est interdit, en français contemporain. Dans le référentiel narratif, la structure de succession domine mais des points de vue rétrospectifs et prospectifs peuvent également apparaître.

D'autres référentiels doivent, à côté des référentiels externes, énonciatifs et narratifs, être envisagés (référentiels médiatifs, référentiels des possibles, ...). La prise en compte de différents référentiels est indispensable à une analyse temporelle et aspectuelle des textes.

A cette temporalité linguistique, qui fait de l'instant la limite indivisible permettant de fonder la perpétuité du temps, il convient d'adjoindre une temporalité phénoménologique au sein de laquelle l'instant est un intervalle de temps et qui ne peut s'accomoder d'une vision atomiste du continu. Saisir la temporalité dans ces deux dimensions permet d'appréhender la trichotomie état/événement/processus suivant deux référentiels : un espace abstrait, un espace phénoménologique. Ces deux espaces sont irréductibles l'un à l'autre. A titre d'exemple, un événement, non seulement peut, dans l'espace abstrait, ne pas être ponctuel (« Il régna pendant 30 ans », « Il plut tout la journée »), mais, dans sa dimension phénoménologique, n'est jamais ponctuel dans la mesure où il fait l'objet d'une observation. Ces deux dimensions doivent néanmoins être « compatibles ». C'est ce qu'autorise le modèle locologique. On montre, par exemple, que l'événement phénoménologique est représenté comme l'ombre (au sens technique du terme) de l'événement abstrait. De plus, la modélisation locologique offre une représentation des termes de la trichotomie aspecto-temporelle dans leur déploiement temporel et dans leur dimension prédicative, laquelle autorise la prise en compte des conditions de vérité des relations prédicatives sous-jacentes aux énoncés. On montre que la *modalisation* de la vérité à laquelle procède la locologie exprime la connexion entre temporalité et spatialité dans le langage et le jugement logique. La modalisation passe donc par une *géométrisation* de la vérité. Dans la plupart des analyses logiques du langage, la temporalité et la spatialité sont présupposées sur le plan des contenus sémantiques d'un jugement, mais occultées sur celui de la structure logique. La locologie rend compte, via la géométrisation, de cette connexion et autorise une représentation des modalités (du possible et du nécessaire) du langage, sans pour autant être une logique modale (c'est-à-dire utilisant des opérateurs modaux) (De Glas & Desclés, article dans ce volume).

La catégorisation aspecto-temporelle donne des notions d'état, de processus et d'événement et des notions aspectuelles d'achèvement, d'accomplissement, de visée téléique, ... une représentation locologico-algébrique qui leur confère une signification générale, grâce à une abstraction des relations *spatio*-temporelles (de position, de mouvement, de changement) que l'on peut voir comme une projection

de l'espace spatio-temporel du référentiel dans lequel se déploient ces relations sur l'axe temporel. L'étude de ces relations est susceptible de fournir des éléments de réponse à la question de l'ancrage du langagier sur le perceptif visuel et à la question corrolaire de l'émergence d'une prédication qu'un tel ancrage autorise. Cet ancrage passe (i) par des représentations dans un référentiel spatial structuré par des relations topologiques de position, de localisation d'un objet par rapport à un lieu (à l'intérieur de, dans, hors de, sur, ...), (ii) des représentations des mouvements dans un référentiel spatio-temporel muni de relations de repérage d'un objet par rapport à des lieux (de départ, d'arrivée, ...), (iii) des représentations de situations spatio-temporelles permettant l'insertion de l'énonciateur, (iv) des représentations favorisant l'émergence d'une prédication (Desclés, 1991).

La mathématisation des concepts linguistiques doit mettre en jeu divers formalismes : topologie, locologie, logique, morphodynamique. La formalisation des phénomènes de temporalité et d'aspectualité, verbalisés au travers des catégories grammaticales du temps et de l'aspect, emprunte aux modèles topologiques et locologiques. En revanche, les diathèses sont plus facilement formalisables dans le cadre logique offert par le *I*-calcul et la logique combinatoire. Afin de relier les représentations conceptuelles et figuratives aux manifestations linguistiques, un modèle du langage doit articuler différents niveaux de formalismes, les niveaux conceptuels et figuratifs étant plutôt d'une nature topologique, les manifestations linguistiques étant plutôt logiques.

Toutefois, l'analyse ci-dessus montre qu'il n'y a pas de coupure entre le logique et le géométrique/topologique et que les modèles catégoriques et topologiques offrent un cadre relativement unifié permettant de voir le logique comme «émergeant» du géométrique, lequel s'ancre sur le perceptif grâce à la médiation offerte par le morphologique. C'est, semble-t-il, la seule façon de rompre avec le logicisme et d'envisager favorablement une intégration théorique des modèles (topologico-algébrico-) logiques et des modèles morphodynamiques.

4. MORPHODYNAMIQUE ET LOGIQUES

4.1. Morphodynamique

Les expressions de «logique du sens» et de «physique du sens», complémentaires l'une de l'autre, sont situées de part et d'autre de la

ligne de clivage entre les paradigmes symbolique et connexionniste en sciences cognitives.

L'orientation cognitiviste symbolique est incompatible avec les thèses naturalistes, puisqu'il n'existe pas de formes symboliques dans la nature. Mais, d'un autre côté, le point de vue connexionniste ne saurait suffire à fonder mathématiquement les sciences cognitives en tant que sciences naturelles car il ne permet pas à lui seul de comprendre comment des structures qualitatives intrinsèques peuvent émerger dynamiquement et de façon stable des substrats physiques. Pour élaborer une médiation entre le physique et le symbolique, on peut tenir compte du fait :

- (i) que le physique est spatio-temporellement conditionné,
- (ii) que ce conditionnement spatio-temporel de la « géno »-physique fondamentale peut être étendue aux dimensions topologiques, géométriques et dynamiques de la « phéno »-physique morphologique,
- (iii) que le symbolique constitue un niveau formel de surface par rapport aux infrastructures morphologiques.

Les thèses sous-jacentes à la morphodynamique sont que :

(i) Entre le physique et le symbolique, le morphologique offre une possibilité de médiation, permettant de dépasser le dualisme du physique et du symbolique et d'accéder à une théorie naturaliste intégrée -- moniste mais non réductionniste -- de leur unité ontologique.

(ii) Les structures morphologiques sont les produits de processus dynamiques d'organisation des substrats physiques ou mentaux. Elles émergent de ces substrats et sont phénoménologiquement dominées par les discontinuités qualitatives issues des singularités, des bifurcations, des instabilités structurelles de ces processus dynamiques.

(iii) L'information morphologique résiste à la transduction. Elle est encodée dans, et véhiculée par, les signaux lumineux et sonores, puis décodée par les transducteurs. Au cours de cette opération, l'information morphologique se reconstitue en restant en grande partie isomorphe à elle-même. Les discontinuités qualitatives se transfèrent de substrat à substrat.

La thèse centrale de la morphodynamique (Petitot, 1985, 1992) est que la naturalisation des structures et du sens est rendue possible par le développement d'une dynamique des formes : les structures sont des formes ; la forme est le phénomène de l'organisation de la matière ; par

suite, les théories dynamiques de l'auto-organisation des substrats matériels (physiques ou neuronaux) sont de théories de l'émergence de formes structurant ces substrats. Les modèles formels qui s'imposent alors sont ceux de la théorie de la stabilité structurelle des systèmes dynamiques, de leurs attracteurs et de leurs bifurcations, de la théorie des singularités et de leurs déploiements universels, de la théorie des phénomènes critiques et des phénomènes de rupture de symétrie. Ces théories fournissent des modèles pour des résultats expérimentaux qui montrent que, dans de nombreux systèmes naturels organisés à deux niveaux, le niveau «macro», émergeant à travers des comportements ordonnés et coopératifs du niveau sous-jacent, est essentiellement organisé autour des singularités des processus physiques «micro». Les singularités structurent morphologiquement les dynamiques régissant les phénomènes.

Bien que d'origine physique, le morphologique n'est pas pour autant matériel. Bien que formel, il n'est pas pour autant symbolique : il est géométriquement (topologiquement) formel, et non pas logiquement formel.

Du côté du monde, le programme de recherche, qui se dessine ainsi, a pour vocation d'étudier les processus de phénoménalisation des substrats matériels, de théoriser mathématiquement l'information morphologique qui en émerge et de comprendre comment cette information morphologique est encodée et véhiculée par les signaux lumineux et sonores. Du côté du sujet cognitif, ce programme de recherche vise à développer l'idée maîtresse émise, il y a une vingtaine d'années, par Thom (Thom, 1972) et Zeeman (Zeeman, 1977), selon laquelle une unité sémantique est modélisable par la topologie d'un attracteur d'une dynamique neuronale sous-jacente, les structures combinatoires et logico-algébriques des automatismes de la compétence symbolique devant, par conséquent, être interprétées comme des régularités émergentes stables.

Cette idée a été extensivement développée en sémio-linguistique (Thom, 1980, Petitot, 1992, 1995, Wildgen, 1982) et au sein du paradigme subsymbolique (Amit, 1989, 1995, Smolensky, 1988). Le principal apport de ces modèles est d'avoir explicité des dynamiques spécifiques qui intervenaient dans les modèles morphodynamiques généraux. Mais les concepts essentiels auxquels ces approches ont eu recours sont ceux de dynamique qualitative, de théorie des bifurcations, de théorie des singularités, de thermodynamique statistique que les modèles morphodynamiques avaient transférés, il y a plus de vingt ans, dans le domaine de la psychologie et de la linguistique.

Un système morphodynamique S est défini par la donnée (i) d'un processus interne X qui définit les états internes du système ; (ii) d'un ensemble W de paramètres de contrôle, appelé espace externe de S . Le processus interne est, en fait, un processus X_w , dépendant de la valeur w du contrôle, qui est supposé varier continûment lorsque w varie continûment. On postule, en général, que le processus interne X est un système dynamique différentiable sur un espace interne M : l'état $x(t)$, à un instant donné t , est décrit par les valeurs d'un nombre fini de paramètres internes, c'est-à-dire par un point de la variété différentielle M . Le système dynamique X est donc un champ de vecteurs différentiable sur M , associant à tout x de M un vecteur tangent $X(x)$ à M en x . Il est donc défini en termes de coordonnées locales par un système différentiel ordinaire $dx/dt = X(x)$. La solution, généralement inconnue, de cette équation différentielle est une application différentiable $f: M \times \mathbf{R} \rightarrow M$ associant à tout état initial x et à tout instant t l'état $f(x, t)$ du système. L'application f est le flot, du système : c'est l'ensemble des trajectoires.

Une bonne compréhension du comportement d'un système dynamique, des propriétés de ses singularités, de la topologie de ses attracteurs passe par l'étude de la propriété de stabilité structurelle du système : un système dynamique est structurellement stable si son comportement reste qualitativement invariant quand il est soumis à une petite perturbation dans sa structure. La définition formelle d'une telle propriété, bien qu'intuitivement simple puisqu'elle rend compte de la notion de « robustesse », fait appel à des concepts complexes de topologie différentielle.

Un sous-ensemble A de M est dit *invariant* pour f si $f(x, t) \in A$, pour tout $x \in A$ et pour tout $t \in \mathbf{R}$. L'ensemble limite de f en x est le sous-ensemble $L(x)$ de M défini par $L(x) = \{y \in M : \exists \{t_n\}, \{t'_n\} \subset \mathbf{R}, t_n \rightarrow \infty, t'_n \rightarrow -\infty, f(x, t_n) \rightarrow y\}$. Le bassin d'attraction de $A \subset M$ est le sous-ensemble $B(A)$ de M défini par $B(A) = \{x \in M : L(x) \neq \emptyset, L(x) \subset A\}$. Un sous-ensemble compact A de M est dit *stable* pour f si tout voisinage U de A contient un voisinage invariant V de A . Un compact A est appelé *attracteur* si $B(A)$ est un voisinage de A . Un attracteur stable est dit *asymptotiquement stable*. Un flot f sur $M = \mathbf{R}^n$ est linéaire si $f_t: x \mapsto f(x, t)$ est un automorphisme linéaire de M , auquel cas f est le flot d'un champ de vecteurs linéaire C , c'est-à-dire $f_t = \exp(Ct)$. Un flot linéaire f_t est dit *hyperbolique* si le spectre de f_t ne rencontre pas le cercle unité du plan complexe. De façon équivalente, un champ de

vecteurs linéaire C est hyperbolique si son spectre n'intersecte pas l'axe imaginaire. Un point m de M tel que $f(m, t) = m$ pour tout réel t est une *singularité* de f . Une singularité m d'un difféomorphisme X de M est hyperbolique si l'application tangente T_x du plan tangent à M en m dans lui-même est un automorphisme linéaire hyperbolique. Une singularité m d'un flot f est hyperbolique si pour tout $t \neq 0$, m est une singularité hyperbolique de f_t . Deux flots f et y sur M sont dits *équivalents* s'il existe un homéomorphisme h de M et un homéomorphisme croissant I de \mathbf{R} tels que, pour tout x et pour tout t , $h(f(x, t)) = y(h(x), I(t))$. Soit F l'ensemble des C^r -systèmes sur M , c'est-à-dire l'ensemble des équations différentielles $dx/dt = X(x)$ où X est r fois continûment différentiable. F peut être munie de la C^r -topologie, c'est-à-dire de la topologie induite par la métrique $d(X, Y) = \sup\{\|x_i^k(X_j - Y_j)\| : 1 = i, j = n, 0 = k = r\}$. Deux systèmes X et Y sont équivalents si leurs flots sont équivalents. Un système X est *structurellement stable* s'il existe un voisinage U de X , dans la C^r -topologie, tel que, pour tout $Y \in U$, Y est équivalent à X . En d'autres termes, un système est structurellement stable si son comportement n'est pas altéré par une petite perturbation dans sa structure. Il est clair que, pour un système linéaire, une condition nécessaire et suffisante de stabilité structurelle est que le flot correspondant soit hyperbolique. Ainsi la propriété de stabilité structurelle est-elle générique dans l'ensemble des systèmes linéaires : l'ensemble des systèmes linéaires structurellement stables est dense dans l'ensemble de systèmes linéaires. Cette propriété de généricité n'est pas valide pour les systèmes non-linéaires (Smale, 1977).

De telles approches s'imposent à l'analyse des réseaux de neurones dont les modèles s'inspirent largement des modèles morphodynamiques qui les ont précédés. Le problème de l'apprentissage dans un réseau de neurones peut, à titre d'exemple, être approché de la façon suivante. Pour une matrice donnée W de poids synaptiques, l'équation ci-dessus détermine certains attracteurs, ensembles d'états internes du réseau définis par W . L'apprentissage consiste à préspecifier des états globaux et à déterminer la matrice W_a pour laquelle ces états deviennent des attracteurs. Les algorithmes, tel l'algorithme de rétropropagation, permettant de résoudre ce problème, reviennent à étudier les dynamiques $dW/dt = h(W)$ dans l'espace externe dont les attracteurs sont les matrices W_a . Une trajectoire d'apprentissage conduit d'une dynamique X à une dynamique X' qui

n'est généralement pas du même type qualitatif ; qui, en particulier, ne possède pas les mêmes attracteurs. La déformation continue de X en X' le long de la trajectoire de la dynamique externe fait subir à la dynamique interne des changements qualitatifs. Ces phénomènes, appelés bifurcations, sont liés aux propriétés de stabilité structurelle de la dynamique interne pour certaines valeurs de la matrice de poids synaptiques. Une bonne compréhension de l'apprentissage dans un réseau de neurones passe donc par l'étude des propriétés de stabilité structurelle de la dynamique interne du réseau.

Les modèles morphodynamiques peuvent également être utilisés comme modèles de processus de catégorisation de la façon suivante. Soit un système dynamique avec ses attracteurs. L'espace interne sur lequel est défini le système dynamique est décomposé en bassins d'attraction (presque toute trajectoire du système issue d'un point de l'espace converge vers l'un des bassins) que l'on peut interpréter comme autant de catégories. Les attracteurs sont alors identifiés aux prototypes associés aux catégories. Dans de nombreux cas (par exemple, les phénomènes de perception catégorielle en phonétique), la catégorisation se manifeste en faisant varier des indices qui (comme l'indice de voisement) appartiennent à des espaces de contrôle de nature physique (acoustico-articulatoire) et non perceptive. Le changement de catégorie s'interprète alors comme un phénomène de bifurcation perceptive contrôlée par des paramètres physiques « externes ». Il prend donc place non seulement dans les espaces internes, mais aussi dans les espaces externes. Les phénomènes de bifurcation sont intimement liés aux propriétés d'instabilité structurelle du système. Les théories de la stabilité structurelle et de la bifurcation sont donc nécessaires à une bonne compréhension des processus de catégorisation. Des solutions satisfaisantes ont été élaborées dans le cas des systèmes dynamiques les plus simples, ceux pour lesquels il existe une fonction de Liapunov globale, décroissant le long des trajectoires du système. Les attracteurs sont alors des singularités du système (et les minima de la fonction de Liapunov). Dans ce cas, les problèmes de stabilité structurelle et de bifurcation sont relativement bien compris ; la théorie de Morse et la théorie de Whitney-Thom-Mather permettent de classifier les différents types de bifurcation. Pour des systèmes dynamiques plus complexes (attracteurs périodiques, quasi-périodiques, étranges), les formalismes de la mécanique statistique deviennent nécessaires.

4.2. Grammaires cognitives

Le courant morphodynamique a convergé, ces dernières années, avec d'autres courants des sciences cognitives : les grammaires cognitives (Giv'on, 1979, Jackendoff, 1983, 1987, Langacker, 1988, 1991, Talmy, 1975, 1983), l'approche morphodynamique du langage (Thom, 1972, 1980, Wildgen, 1982, Petitot, 1989), la grammaire applicative (Shaumyan, 1977, Desclés, 1990). Le courant des grammaires cognitives renoue avec une tradition européenne de l'approche des langues (que l'on songe, par exemple, aux travaux philosophiques de Cassirer qui a pris nettement appui sur des analyses linguistiques que l'on qualifierait peut-être, aujourd'hui, de «cognitives»), développée par des linguistes comme Seiler et Wildgen en Allemagne ou Guillaume et Pottier en France. Les grammaires cognitives se caractérisent par un certain nombre de traits :

(i) elles refusent le découpage simple entre morphologie, syntaxe et sémantique, pragmatique, certaines approches refusant même parfois la dichotomie entre lexique et grammaire au profit d'un continuum entre ces deux pôles ;

(ii) les *catégorisations sémantiques* deviennent centrales, d'où l'importance donnée aux conceptualisations, aux prototypes, aux degrés de typicalité et d'atypicalité ;

(iii) les schémas grammaticaux sont analysables en termes de symbolisations qui associent des manifestations phonologiques et des représentations sémantiques ;

(iv) un des objectifs est la mise en évidence de continua grammaticaux et lexicaux ;

(v) les *représentations conceptuelles* ou cognitives qui sont encodées par les langues deviennent le principal objet d'étude ;

(vi) il en résulte des relations privilégiées avec l'*anthropologie et la Gestalt* et des relations « naturelles » avec certains secteurs de la philosophie en insistant sur les processus de *métaphorisation* et sur les *analogies* opérées par la perception (la vision en particulier) et l'action ;

(vii) les représentations mises en œuvre sont, en général, non formelles, au sens où elles ne sont pas logico-computationnelles et propositionnelles.

Les grammaires cognitives proposent un certain nombre de descriptions et de modèles s'appuyant sur des hypothèses plus ou

moins explicitées par les différents auteurs qui se réclament de ce courant. Mentionnons-en quelques-unes. Les activités cognitives *ne sont pas autonomes* mais entrent en interaction les unes avec les autres. Le langage n'est donc pas considéré comme une activité cognitive autonome mais comme une activité qui interagit, d'une part, avec les activités de perception *de* l'environnement et d'action planifiée intentionnellement *sur* l'environnement et, d'autre part, avec les activités de raisonnement. Dans cette perspective, chaque langue, en tant que système de représentation, reste compatible avec la description du monde perçu, mais ne se réduit pas à en être une représentation fidèle. Plutôt que de décrire exclusivement les agencements et configurations syntaxiques, les grammaires cognitives consacrent leurs efforts à la *description de la signification des catégories grammaticales et lexicales*. Les conceptualisations sémantiques deviennent ainsi constitutives des phénomènes grammaticaux et des organisations lexicales. Les grammaires cognitives revendiquent l'utilisation de représentations *iconiques, figurales et imaginales* (fondées, entre autres, sur des analogies) de préférence aux représentations logico-symboliques purement conventionnelles. Les représentations iconiques, les schémas figuratifs jouent ainsi un rôle important dans le dispositif descriptif et théorique. Cependant, les représentations iconiques et figuratives peuvent être simplement métalinguistiques, c'est-à-dire descriptives de certains fonctionnements linguistiques, ou au contraire être inhérentes au fonctionnement langagier. Dans le premier cas, les descriptions figuratives seraient simplement commodes pour décrire économiquement certaines structurations linguistiques (sous forme de continua, par exemple) ou certaines évolutions diachroniques et, par conséquent, ces représentations imaginales auraient le même statut que les figures en physique (une force est représentée par un vecteur, le mouvement cinématique d'un mobile est représenté par une courbe dans un espace à deux ou trois dimensions). Dans le second cas, au contraire, les figures auraient une pertinence cognitive : la signification d'un symbole linguistique « est » une image (stéréotypée), une icône, une carte, une figure, une forme morphodynamique, ...

L'un des principaux problèmes de la grammaire cognitive est de trouver des outils mathématiques adéquats pour décrire les structures topologiques et dynamiques (c'est-à-dire l'information morphologique) fournies par la perception et traitée par la sémantique des langues naturelles. Le contenu de la perception étant manifestement lié à une ontologie qualitative, une phénoménologie du monde naturel, le problème revient à montrer comment le langage peut s'ancrer sur cette phénoménologie qualitative sous forme d'une description mathématique précise. Ces problèmes font partie intégrante du programme de recherche de la morphodynamique. Celui-ci admet l'existence d'un niveau conceptuel (ou cognitif « profond ») où les représentations opérées par les activités de perception, d'action et de langage deviennent sinon identiques (selon l'hypothèse forte de Langacker) du moins compatibles (selon une hypothèse plus faible mais plus plausible). Cette structure conceptuelle rend possible les transformations d'un monde physique en un monde « projeté », en un monde phénoménologique morphologiquement structuré, qui relève d'une ontologie qualitative. Le point de vue de la morphodynamique s'accorde également avec Talmy sur le fait que les classes grammaticales fermées (prépositions, auxiliaires modaux, relations grammaticales, ...) spécifient des contenus sémantiques, d'une nature très profonde, relevant d'une ontologie qualitative du monde naturel. Nous pouvons transférer à la grammaire cognitive les travaux utilisant l'approche morphodynamique. Les modèles morphodynamiques apparaissent être bien appropriés à la description mathématique des représentations sémantiques de la grammaire cognitive ainsi qu'à la modélisation des phénomènes cognitifs mis en jeu par les langues, la perception de l'environnement, par l'action sur l'environnement.

Par ailleurs, de façon à mettre en évidence les éléments nécessaires à l'engendrement des représentations symboliques assumées par les langues, il convient d'isoler les *invariants langagiers* qui transcendent les variations spécifiques (morphologiques, syntaxiques, lexicales) des diverses langues. La description sémantique des prédicats lexicaux (exprimés en général par des verbes ou par des prépositions) conduit à des représentations relationnelles, qui ne sont plus directement lexicalisées dans une langue. Ces représentations sont à la source des diverses catégorisations sémantico-cognitives comme : repérage et localisation statiques d'un objet dans un espace, relation entre un objet et un site, relation entre un objet mobile et un lieu, chemin, mouvements cinématiques spatio-temporels, changements des qualités affectant un même objet, construction ou destruction dynamiques d'objets, transferts d'objets entre les participants d'une

transaction, représentation d'actions plus ou moins contrôlées et intentionnelles, ... C'est ainsi que des primitives relationnelles de « mouvement » et de « changement », de « contrôle d'une action », de « visée téléonomique d'un but », de « capacité représentationnelle d'une action qui implique l'énonciateur lui-même », ... émergent progressivement des analyses des catégorisations (grammaticales et lexicales) opérées par les langues.

Les primitives que nous venons de mentionner ont-elles un fondement cognitif ? Reformulons la question autrement : l'activité langagière a-t-elle un ancrage catégoriel sur les catégorisations cognitives opérées par la perception et les planifications d'actions sur le monde ? Peut-on modéliser les changements de représentations, en particulier les passages des verbalisations de situations spatio-temporelles aux constructions de représentations figurales ou imagées et inversement des images aux verbalisations de ces images ? Plusieurs courants dans la linguistique contemporaine, comme celui des grammaires cognitives, tentent d'approcher ces problèmes fort anciens.

Par exemple, l'*hypothèse localiste*, dont une première formulation a été proposée par les Stoïciens, a déjà tenté d'établir des liens étroits entre les structures grammaticales et les représentations de l'espace, introduisant ainsi certaines relations motivées entre langage et monde externe, ce qui nuancait fortement le principe d'arbitrarité des signes linguistiques. Cette hypothèse a été reprise à différentes époques, notamment à l'époque moderne par Hjelmslev (1935), Anderson (1977), Miller (1985) et Petitot (1985). L'hypothèse localiste a toujours été jugée suspecte et reste, en fin de compte, peu convaincante. Nous proposons une reformulation plus abstraite de ce que nous appelons un « localisme cognitif » en :

(i) refusant de réduire les catégories grammaticales les plus fondamentales (prédication et « cas ») à des représentations purement spatiales (« localisme pur » qui nous paraît irrecevable), puisque chaque mouvement perçu consomme du temps, et que, de plus, le temps est pensé en termes spatiaux et sert à indiquer des rapports spatiaux ; les représentations les plus élémentaires opérées par les langues sont donc spatio-temporelles et non purement spatiales ;

(ii) distinguant différents niveaux de représentations ; les compatibilités et interactions entre langage, perception et action s'effectuant non pas au niveau des observables des langues (réduction effectuée par le « localisme pur ») mais à un niveau de représentations cognitives ;

(iii) insérant tous les niveaux de représentations (linguistiques, langagières, cognitives) dans une architecture cognitive générale où langage et autres activités cognitives peuvent entrer en interaction;

(iv) proposant un premier « localisme spatio-temporel » où représentations opérées par le langage et représentations opérées par la perception, perception visuelle en particulier, deviennent compatibles ; on se donne pour cela des primitives sémantico-cognitives directement liées à un modèle qualitatif du monde physique externe, perçu et organisé par les processus de perception et les modes d'actions sur ce monde ;

(v) étendant ce premier localisme spatio-temporel à l'aide de primitives plus intentionnelles, constituant ainsi un second « localisme cognitif ».

L'ancrage catégoriel du langage sur la perception visuelle et l'action (et non pas la réduction du langage à la perception et à l'action !) résulte de changements d'espaces de représentations. Tout d'abord, l'appareil percepto-cognitif construit un « modèle physique qualitatif » du monde perçu. Ainsi, les organes de perception différencient le stable de l'instable et opposent un objet saillant, mouvant et susceptible de changer de qualités, à un arrière fond stable qui sert de référentiel local. Le modèle « physique » se complique lorsqu'on y adjoint des interactions entre des « événements » singuliers (en fait des « catastrophes » au sens de la morphodynamique) et entre des objets du monde : certaines entités apparaissent et disparaissent de notre champ de visibilité, telle entité « agit » sur telle autre entité, le mouvement de tel objet « transmet » son mouvement à un autre objet, tel « événement » est responsable de l'occurrence de tel autre « événement » postérieur, tel « événement » est responsable de la situation qui en résulte ... *Les Leçons de Physique* d'Aristote sont une « description catégorielle » de ce qui *apparaît qualitativement*. Si la physique aristotélicienne n'est plus, depuis Galilée, scientifique, elle conserve néanmoins une certaine cohérence et conserve une compatibilité directe avec les phénomènes perçus. Elle établit les fondements des catégories les plus primitives des langues : entités individualisables, collectives massives, lieux, ... autant d'entités qui sont catégorisées sémantiquement par les langues. Les catégorisations lexicales différencient le stable (les entités nominales) des entités prédicatives qui servent à représenter les situations de changements effectués dans un référentiel spatio-temporel. Trois grands types de relations et d'opérations, essentiellement qualitatives, contribuent à structurer, par le biais des opérations linguistiques, les référentiels dénotatifs que les discours expriment :

(i) les *relations statiques* : localisation d'un objet ou d'un lieu par rapport à un autre lieu ; visualisation, dans son intériorité ou extériorité

ou encore par ses bords, d'un lieu, éventuellement déterminé par un objet ; attribution de qualités à une même entité, subordinations entre concepts, ingrédiences entre entités, ... ;

(ii) les *relations cinématiques* des mouvements spatio-temporels ou des changements des qualités d'un objet ou d'un lieu ;

(iii) les *relations dynamiques* d'interactions entre objets et entre situations : tel objet « cause » la modification (mouvement ou changement) qui affecte tel autre objet, ou provoque sa disparition ou son apparition, ...

Le second système n'est pas constitué selon un principe d'analogie iconique : il substitue des représentations relationnelles et symboliques aux représentations analogiques et iconiques des phénomènes. Le mouvement d'un objet n'est plus représenté par un diagramme figuratif ou par une image mais par des *relations symboliques* de repérage. Pour cela, on thématise les différentes phases d'un objet mouvant ou changeant par des repérages par rapport à des lieux perceptivement saillants : lieu de départ ou source, lieu d'arrivée ou but, éventuels lieux intermédiaires par lesquels passe l'objet au cours de son mouvement ou de son changement. Les repérages constituent un système de coordonnées qualitatives d'un objet repérable par rapport à son lieu d'origine, son ou ses lieux intermédiaires et son lieu final de repos. En tant que systèmes de représentations symboliques, les langues *encodent* les relations de repérage sous forme de *marqueurs linguistiques*. Le « pur » localisme prend ce système qualitatif de repérage comme système prototypique des valeurs sémantiques des cas et des prépositions. Il établit une *analogie iconique* entre les valeurs grammaticales et les valeurs spatiales ; ainsi, celui qui est responsable (l'agent) du mouvement d'un objet, puis, par généralisation, de l'action sur un objet, sera considéré comme un lieu initial ; l'objet intermédiaire (l'instrument) sera considéré comme un lieu intermédiaire par lequel passe l'objet en mouvement ou en action ; le lieu final constitue le lieu de repos atteint par l'objet après son mouvement ou à la suite de l'action exercée sur lui.

Le système relationnel précédent rend, entre autres, possible, l'introduction dans les représentations elles-mêmes des différentes relations pragmatiques entre l'énonciateur, son environnement (environnement des co-énonciateurs, environnement spatio-temporel) et les situations qu'il souhaite représenter. L'énonciation se ramène alors à un ensemble d'opérations qui insèrent la représentation d'une situation dans un référentiel organisé par l'énonciateur qui prend pour origine soit lui-même, soit ses co-énonciateurs, et non pas un repère objectif et absolu (comme c'était le cas dans le deuxième système). L'insertion des énonciateurs dans les représentations symboliques se

manifeste dans des catégorisations spécifiques aux langues naturelles :
personne, déixis, aspects-temps, thématisation, ...

Les relations de repérage des deux systèmes ci-dessus ne prennent pas assez en compte la dimension de plus ou moins grande intentionalité planificatrice que les langues ont tendance à grammaticaliser plus ou moins directement. Le localisme « pur » a eu le tort de vouloir « réduire » tous les schèmes prédicatifs aux seules combinaisons des relateurs de repérage. Outre les primitives topologiques (intérieurité, extérieurité, frontière et bords, ...) et un archirelateur de repérage avec ses multiples spécifications (localisation spatiale, attribution, ingrédience, possession, identification, ...), il faut se donner d'autres primitives. Du modèle physique qualitatif (premier système), on peut extraire des relateurs : mouvement spatio-temporel, changement d'états et compliquer ce système initial en lui adjoignant des relateurs plus intentionnels tels : capacité d'action sur des objets, capacité de contrôle sur des actions, capacité d'anticipation téléonomique, capacité d'un énonciateur de se représenter une action. Il a été proposé (Desclés, 1985, 1990) de prendre ces relateurs comme des primitives sémantico-cognitives combinables entre elles de façon à construire des schèmes prédicatifs susceptibles de représenter les significations des prédicats linguistiques. Chaque schème résulte d'une organisation structurée de primitives sémantico-cognitives.

A condition de poser les problèmes au travers d'une architecture cognitive mettant en jeu des niveaux de représentations et des processus d'intégration et de décomposition entre niveaux, les sciences cognitives nous offrent un cadre adéquat pour décrire les transferts de représentations, en particulier entre les représentations figurales (imagées ou analogiques) et les représentations symboliques encodées par les langues. L'architecture cognitive du langage nécessite au moins trois grands niveaux de représentations. Le premier niveau correspond à un niveau superficiel d'observation : celui des langues directement observables, c'est-à-dire des langues naturelles analysées en tenant compte des variations morpho-syntaxiques spécifiques, voir irréductibles, de chaque langue : variations par la présence de tel trait typologique caractéristique. Le second niveau de représentations vise à capter les structures les plus fondamentales du langage et à décrire les catégorisations et les opérations nécessaires qui sont mises en jeu par l'activité de langage et réalisés par les systèmes de représentations symboliques que sont les langues. Ce niveau de description est construit à la suite d'une procédure abductive qui prend appui sur les structures observables des langues. Il est également encodé, par des procédés divers, dans les différents systèmes phénotypiques. C'est à ce niveau que sont décrits les invariants opératoires les plus fondamentaux

du langage. En particulier, les structures prédicatives sont décrites formellement à ce niveau abstrait. D'autres opérations (comme détermination, thématisation, diathèse, ...) et catégorisations (comme temps et aspects, déixis, personne, modalités, ...) sont également constitutives des expressions de ce niveau. Un troisième niveau de description est nécessaire lorsque l'on veut prendre en compte les problèmes de signification, en particulier des significations des prédicats lexicaux et grammaticaux. Ce troisième niveau de représentations est constitué par les représentations cognitives engendrées par les schèmes sémantico-cognitifs qui représentent les significations des unités du deuxième niveau. Ce troisième niveau est organisé de façon à être compatible avec les représentations cognitives construites par la perception et par les schèmes d'actions planifiées. L'architecture cognitive suppose que les niveaux soient reliés entre eux par des processus descendants (ou « top-down ») onomasiologiques de « réunitarisation » (allant du cognitif aux encodages linguistiques dans une langue) ou par des processus ascendants (ou « bottom-up ») sémasiologiques de « décomposition » (allant d'une langue à la représentation cognitive sous-jacente). C'est dans cette architecture que l'hypothèse de la compilation généralisée acquiert ainsi toute sa pertinence.

Il convient donc de disposer d'un modèle qui permette de définir un cadre général où, d'une part, les relations entre langues, langage et cognition puissent être discutées et où les problèmes de mathématisation des concepts cognitifs de base et les traitements informatiques par changements de niveaux trouvent une place. C'est ce à quoi s'attache la *grammaire applicative et cognitive* (GAC) qui peut être caractérisée par les objectifs suivants (Desclés, 1990, 1991) :

(i) Le modèle se veut interactionniste au sens où le langage n'est pas une activité autonome et indépendante des autres activités cognitives ; plus précisément, les représentations et les catégorisations opérées par les langues sont ancrées (non pas réductibles) aux représentations et catégorisations opérées par la perception de l'environnement et par l'action qu'un énonciateur peut exercer sur son environnement. Les primitives sémantico-cognitives de mouvement, de changement de contrôle, ... sont constitutives des opérations et schèmes cognitifs développés par l'activité de langage.

(ii) La GAC étend le modèle de la grammaire applicative de Shaumyan (1987) qui a déjà proposé deux niveaux de description : le niveau du langage génotype et les niveaux des langues phénotypes. La GAC reprend explicitement ces deux niveaux, mais y adjoint toutefois

d'autres opérations comme les opérations d'énonciation, et y ajoute une couche supplémentaire, le niveau des représentations sémantico-cognitives, qui rend possible l'analyse des significations à l'aide de schèmes. La GAC reprend également le formalisme applicatif utilisé par Shaumyan qui fait appel à la logique combinatoire de Curry. Ce formalisme applicatif repose sur l'opération de base qu'est l'application d'opérateurs à des opérands et sur des opérateurs particuliers, appelés combinateurs, qui permettent de composer entre eux des opérateurs. Toutes les unités, aux différents niveaux, sont alors considérées comme des opérateurs ou des opérands de différents types. En effet, on considère, selon les niveaux, des types syntaxiques au niveau du phénotype, des types fonctionnels au niveau du génotype et des types sémantico-cognitifs au niveau des représentations cognitives. Etant donnés les liens étroits entre la logique combinatoire, le *I*-calcul et la programmation fonctionnelle, les analyses linguistiques opérées dans le cadre de la GAC, sont réalisées effectivement dans un langage de programmation fonctionnelle. La proximité des formalismes de la logique combinatoire et du *I*-calcul de Church met, en outre, en évidence l'étroite parenté entre : théorie des langues analysées dans le modèle de la GAC ; problèmes logiques analysés dans le cadre de la logique combinatoire ou du *I*-calcul ; programmation fonctionnelle ; représentations cognitives exprimées par des *I*-expressions ; théorie des catégories cartésiennes fermées et théorie des topoi.

(iii) Les concepts topologiques et locologiques étant nécessaires à la mathématisation de certains concepts linguistiques (relatifs au temps, à l'espace, aux modalités ...), la GAC doit les intégrer sous forme d'opérateurs spécifiques constitutifs des opérations, des schèmes grammaticaux et lexicaux et des représentations cognitives. Les relations avec les approches morphodynamiques du langage sont explorées au niveau cognitif en recherchant des transferts entre les schèmes logico-symboliques construits par l'activité de langage et les schèmes de nature plus géométrique construits par les activités de perception et d'action.

(iv) Les relations entre les niveaux sont décrites sous forme de processus explicites de « décompositions » ou de « réunitarisations » à l'intérieur du formalisme fonctionnel typé qu'est la logique combinatoire ; elles sont programmées dans un langage fonctionnel. En particulier, les combinateurs sont des opérateurs qui permettent de réunitariser une unité d'un niveau d'analyse à partir d'unités déterminées à un autre niveau d'analyse. Le modèle contient donc des modèles plus particuliers (grammaire catégorielle combinatoire

applicative) qui permettent de construire, *effectivement*, à partir des analyses syntaxiques opérées dans le phénotype, les représentations applicatives du langage génotype puis les représentations sémantico-cognitives sous forme de schèmes exprimables par des *I*-expressions particulières. Pour tenir compte des indices contextuels qui interviennent dans la construction des représentations cognitives à partir des textes, il est cependant nécessaire d'introduire, dans le programme général de construction, des procédures d'exploration contextuelles qui participent à la levée de l'indétermination des marqueurs linguistiques.

L'approche morphodynamique, en raison notamment de la puissance des outils mathématiques qui la fondent, est attrayante et prometteuse. Elle souffre néanmoins de certaines limites. D'une part, le recours massif aux outils de la topologie différentielle nous place dans un cadre hautement *non constructif*. D'autre part, le morphologique est topologiquement et géométriquement formel mais *non logiquement formel*. Comment, dès lors, des formes géométriques peuvent-elles être manipulées à des niveaux cognitifs supérieurs, descriptibles en termes logiques ?

Une voie de recherche consisterait à élaborer une théorie constructive des systèmes dynamiques, une *morphodynamique constructive*, dont la contepartie logique serait une *logique intensionnelle*, ayant une interprétation géométrique. La pertinence d'un tel projet se fonde sur l'idée qu'il est possible d'appréhender les dimensions respectivement logico-symbolique et morphodynamique de la cognition comme deux versants d'un même paradigme. Il convient donc de « constructiviser » la théorie des systèmes dynamiques, afin de redéfinir les concepts de la théorie des singularités, de la théorie de la stabilité structurelle et de la théorie des bifurcations dans un cadre constructif.

Ce projet d'intégration théorique vise donc à l'élaboration d'un paradigme dans lequel logique intensionnelle et morphodynamique constructive seraient les deux versants d'une même théorie ; théorie scientifique qui seule serait susceptible de dépasser ce « hiatus infranchissable entre le logique et le morphologique » (Thom), c'est-à-dire de relier les structures formelles du niveau symbolique qui, d'une façon ou d'une autre, sont des formes *logiques* aux dynamiques qui régissent le niveau *physique*.

Rien ne s'oppose, dans le principe, à un tel projet, c'est-à-dire à une « nouvelle alliance » pour *l'appréhension formelle du sens*.

Bibliographie

- Abdi, H. (1985) Représentations arborées de l'information verbatim. In J. Wittwer (ed.) *Psycholinguistique textuelle*, Bulletin de psychologie (numéro spécial).
- Abdi H., Barthélemy J.-P. & Luong X. (1984) Tree Representations of Associative Structure in Semantic and Episodic Memory. In E. Degreef & J. Van Buggenhaut (eds.) *Trends in Mathematical Psychology*, Amsterdam, Elsevier, North-Holland.
- Abdi H., Barthélemy J.-P. & Luong X. (1989) Représentations arborées des catégories naturelles, *Cumfid* 16, 157-178.
- Abraham M. & Desclés J.-P. (1992) Interaction Between Lexicon and Image : Linguistic Specification of Animation, *XIVth International Conference on Computational Linguistics*, Vol. 3, 1043-1047.
- Amari Sh.-I. (1990) Formation of Cortical Cognitive Map by Self-Organization. In E.L. Schwartz (ed.) *Computational Neuroscience*, Cambridge, The MIT Press.
- Amit D. (1989) *Modeling Brain Function*, Cambridge University Press.
- Amit D. *et al* (1995) The Hebbian Paradigm Revisited, *Behavioral and Brain Sciences*, 18, 4.
- Andler D. (1988) Progrès en situation d'incertitude, *Le Débat* 47, 5-25.
- Andler D. (éd.) (1992) *Introduction aux sciences cognitives*, Gallimard, Paris.
- Aristote (1946-50) *Organon*, trad. fr. J. Tricot, Paris, Vrin.
- Aristote (1953) *La métaphysique*, trad. fr. J. Tricot, Paris, Vrin.
- Arnold V.I. (1985) *Catastrophe Theory*, Springer-Verlag.
- Arnold V.I., Varchenko A. & Goussein-Zade S. (1985) *Singularities of Differential Maps*, Boston, Birkhäuser.
- Attneave F. (1950) Dimensions of Similarity, *American Journal of Psychology* 63, 516-566.
- Attneave F. (1957) Transfert of Experience with a Class-Schema to Identification-Learning of Patterns and Shapes, *Journal of Experimental Psychology* 54, 81-108.
- Aubenque P. (1962) *Le problème de l'être chez Aristote*, Paris, Presses Universitaires de France.
- Auroux S. (1993) *La logique des idées*, Montréal, Bellarmin - Vrin.
- Auroux S. (1994) L'hypothèse de l'histoire et la sous-détermination grammaticale, *Langages* 114, 25-40.

- Bachimont B. (1992) *Le contrôle dans les systèmes à base de connaissances ; contribution à l'épistémologie de l'intelligence artificielle*, Paris, Hermès.
- Barthélemy J.-P. (1984) About the Asymmetries of Similarity Judgments, An Ordinal Point of View. In E. Degreef & J. Van Buggenhaut (eds.) *Trends in Mathematical Psychology*, Amsterdam, Elsevier, North-Holland.
- Barthélemy J.-P. (1993) Similitudes, arbres et typicalité. In D. Dubois (éd.) *Sémantique et cognition*, Paris, Editions du C.N.R.S.
- Barthélemy J.-P. & Guénoche A. (1991) *Trees and Proximity Representations*, Chicester, J. Wiley.
- Barwise J. (ed.) (1977) *Handbook of Mathematical Logic*, Amsterdam, North-Holland.
- Benveniste E. (1966) *Problèmes de linguistique générale*, Paris, Gallimard.
- Boldini P. (1994) Morphismes et catégories : une lecture formelle de Piaget, *Intellectica*.
- Boldini P. (1994) Structuration cognitive et logique intrinsèque, *Mathématique, Informatique et Sciences Humaines*.
- Boldini P. (1995) *Contributions de la théorie des catégories à la représentation des connaissances*, Thèse, Université Rennes 1.
- Boldini P. & Barthélemy J.-P. (1994) Representativity as Emergence, *7th International Conference on Systems Research, Informatics and Cybernetics*, Baden-Baden.
- Boole G. (1854) *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*, New York, Dover Publications.
- Boudot M. (1972) *Logique inductive et probabilité*, Paris, Armand Colin.
- Boutot A.-S. (1994) *Vers une représentation logico-géométrique de la contradiction*, Mémoire de DEA de Sciences Cognitives, Université Pierre et Marie Curie.
- Brouwer L.E.J. (1975) *Collected Works*, Vol. 1, Amsterdam, North-Holland.
- Carnap R. (1947) *Meaning and Necessity*, University of Chicago Press.
- Caroll J.D. & Arabie P. (1980) Multidimensional Scaling, *Annual Review of Psychology* 31, 229-240.
- Casati R., Smith B. & White G. (eds.) (1993) *International Wittgenstein Symposium*, Vienna, Verlag-Hölder-Pichler-Tempsky.
- Cassirer E. (1972) *La philosophie des formes symboliques. I. Le langage*, Paris, Editions de Minuit.
- Chambreuil M. & Pariente J.-C. (1990) *Langue naturelle et logique. La sémantique intensionnelle de R. Montague*, Berne, Peter Lang.
- Chenciner A (1985) Singularités des fonctions différentiables, *Encyclopaedia Universalis*.
- Changeux J.-P. (1980) *L'homme neuronal*, Paris, Fayard.

- Chomsky N. (1964) Current Issues in Linguistic Theory. In J.A. Fodor & J.J. Katz (eds.) *The Structure of Language. Readings in the Philosophy of Language*, Englewood Cliffs, Prentice-Hall, 51-118.
- Chomsky N. (1970) *Le langage et la pensée*, Paris, Payot.
- Chomsky N. (1971) *Aspects de la théorie syntaxique*, Paris, Le Seuil.
- Chomsky N. (1980) *Essais sur la forme et le sens*, Paris, Le Seuil.
- Chomsky N. (1987) *La nouvelle syntaxe*, Paris, Le Seuil.
- Chomsky N. (1988) *Language and Problems of Knowledge*, Cambridge (Mass.), The MIT Press.
- Churchland P.M. (1979) Eliminative Materialism and Propositional Attitudes, *J. Philosophy* 78, 2, 67-90.
- Churchland P.M. (1984) Some Reductive Strategies in Cognitive Neurobiology, *Mind*, 279-309.
- Churchland P.S. (1986) *Neurophilosophy*, Cambridge (Mass.), The MIT Press.
- Cohen P. (1962) *Set Theory and the Continuum Hypothesis*, New York, Benjamin.
- Collins A. & Quillian M.R (1969) Retrieval Time from Semantic Memory, *Journal of Verbal Learning & Verbal Behavior* 8, 240-247.
- da Costa N.C.A. (1974) On the Theory of Inconsistent Formal Systems, *Notre Dame J. Math. Logic* 11, 497-510.
- Culioli A. (1973) Sur quelques contradictions en linguistique, *Communications* 20, 83-91.
- Culioli A. (1991) *Pour une linguistique de l'énonciation. Opérations et représentations*. Tome 1, Paris, Ophrys.
- van Dalen D. (1978) An Interpretation of Intuitionistic Analysis, *Ann. Math. Logic*, 13, 1-43.
- Davidson D. (1984) *Inquiries into Truth and Interpretation*, New York, Oxford University Press.
- Davidson D. (1992) Mental Causation. In A. Mele & J. Heil (eds.) *Thinking Causes*, Oxford University Press.
- De Glas M. (1990) Locological Spaces : Knowledge Representation in an Intensional Setting, *Cognitiva* 90, Madrid.
- De Glas M. (1991) Intuitionism and Knowledge Representation. In M. De Glas & D. Gabbay (eds.) *Proceedings of WOCFAI 91*, Angkor, Paris.
- De Glas M. (1992) A Local Intensional Logic, *International Conference on Algebraic Methods and their Computer Science Applications*, Polish Academy of Sciences, Warsaw.
- De Glas M. (1993) An intuitionistic Basis for Qualitative Reasoning, *QUARDET*, Barcelone.

- De Glas M. (1993) Pensée Logico-mathématique et Intelligence Artificielle. In O. Houdé & D. Miéville (eds., avec la participation de J.-P. Changeux, M. De Glas, P. Engel & J.-B. Grize) *Pensée Logico-Mathématique. Nouveaux Objets Interdisciplinaires*, PUF.
- De Glas M. (1996) Pseudoconsistent Logic. Towards a Formal Framework for Handling Contradictions, *Studia Logica, Special Issue dedicated to the Memory of Helena Rasiowa*, à paraître.
- De Glas M. & Jacquet J.-Ph. (1995) Pseudoconsistent Logic, *Proceedings of WOCFAI 95*, Paris, Angkor.
- De Glas M., Jacquet C. & Thiénot C. (1996) Implication does not imply Inclusion, *Logic Colloquium*, San Sebastian.
- Dennett D. (1969) *Content and Consciousness*, London, Routledge.
- Dennett D. (1987) *The Intentional Stance*, Cambridge (Mass.), The MIT Press.
- Desclés J.-P. (1985) Représentation des connaissances : archétypes cognitifs, schèmes conceptuels, schémas grammaticaux, *Actes sémiotiques, Documents VII*, Editions du CNRS, 69-70.
- Desclés J.-P. (1987) Réseaux sémantiques : la nature logique et linguistique des relateurs, *Langages*, 87, 57-78.
- Desclés J.-P. (1990a) *Langages applicatifs, langues naturelles et cognition*, Paris, Hermès.
- Desclés J.-P. (1990b) State, Event, Process, and Topology, *General Linguistics*, 29, 3, 159-200.
- Desclés J.-P. (1990c) Les représentations intermédiaires, *Revue de Synthèse*, 4ème Série, 1-2, 33-56.
- Desclés J.-P. (1991) La prédication opérée par les langues, *Langage*, 103, 83-96.
- Desclés J.-P. (1993) Dialogue sur les prototypes et la typicalité. In M. Denis & G. Sabah (éds.) *Modèles et concepts pour la science cognitive. Hommage à Jean-François Le Ny*, Presses Universitaires de Grenoble.
- Desclés J.-P. (1996) Cognition, compilation et langage. In G. Chazal & M.N. Terrasse (éds.) *Philosophie du langage et informatique*, Paris, Hermès.
- Desclés J.-P. & Kanellos I. (1991) Approximation et typicalité : une approche formelle. In D. Dubois (éd.) *Sémantique et cognition, catégories, prototypes, typicalité*, Editions du CNRS, 225-244.
- Dretske F. (1988) *Explaining Behavior*, Cambridge (Mass.), The MIT Press.
- Dreyfus H. (1984) *Intelligence artificielle, mythes et réalité*, Paris, Flammarion.
- Dubois D. (dir.) (1993) *Sémantique et cognition. Catégories, prototypes, typicalité*, Paris, Editions du CNRS.

- Dummett M. (1977) *Elements of Intuitionism*, Oxford, Oxford University Press.
- Dumont J.-P. (dir.) (1994) *Les Présocratiques*, La Pléiade.
- Duquenne V. (1986) Conceptual Implications Between Attributes and Some Representation Properties for Finite Lattices. In B. Ganter, R. Wille & K.E. Wolff (eds.) *Beiträge zur Begriffsanalyse*.
- Edelman G.M. (1989) *The Remembered Present : A Biological Theory of Consciousness*, New York, Basic Books.
- Engel P. (1989) *La norme du vrai, philosophie de la logique*, Paris, Gallimard.
- Engel P. (1992) *Etats d'esprit, questions de philosophie de l'esprit*, Aix-en-Provence, Alinea.
- Enjalbert P. (1989) Notes préliminaires à une théorie opérationnelle du sens, *Intellectica* 8, 109-159.
- Enjalbert P. & Stefani J.B. (1990) Savoir qui, quand, où,... Une approche centrée sur les notions de question et de contexte de résolution de problèmes, *4ème colloque de l'ARC*, Paris.
- Ehresmann A.C. & Vanbremeersch J.-P. (1987) Hierarchical Evolutive Systems : A Mathematical Model for Complex Systems, *Bulletin of Mathematical Biology*, Vol. 49, N°1, 13-50.
- Ehresmann A.C. & Vanbremeersch J.-P. (1991) Un modèle pour des systèmes évolutifs avec mémoire, basé sur la théorie des catégories, *Revue Internationale de Systémique*, Vol. 5, N°1, 5-25.
- Feferman S. (1988) Hilbert's Program Relativized : Proof-Theoretical and Foundational Reductions, *The Journal of Symbolic Logic*, 53, 2, 364-384.
- Feferman S. (1989) Infinity in Mathematics : Is Cantor Necessary ?, *Philosophical Topics*, XVII, 2, 23-45.
- Fodor J. (1975) *The Language of Thought*, Cambridge (Mass.), The MIT Press.
- Fodor J. (1981) *Representations. Philosophical Essays on the Foundations of Cognitive Science*, Cambridge (Mass.), The MIT Press.
- Fodor J. (1983) *The Modularity of Mind*, Cambridge (Mass.), The MIT Press.
- Fodor J. (1990) *A Theory of Content*, Cambridge (Mass.), The MIT Press.
- Fodor J. & Pylyshyn Z. (1988) Connectionism and Cognitive Architecture : A Critical Analysis, *Cognition*, 28, 1/2, 3-72.
- Fourman M.P. (1977) The Logic of Topoi. In J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam, North-Holland.
- Friendly M. (1979) Methods for Finding Graphic Representations of Associative Memory Structures. In C.R. Puff (ed.) *Memory Organization and Structure*, New York, Academic Press.
- Galton F. (1879) Composite Portraits, Made by Combining Those of Many Different Persons into a Single Resultant Figure, *Journal of the Anthropological Institute* 8, 132-144.

- Garfinkel H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, New Jersey, Prentice Hall.
- Gauthier Y. (1991) *De la logique interne*, Paris, Vrin.
- Genter D. (1981) Verb Semantic Structures in Memory for Sentences : Evidence for Componential Representation, *Cognitive Psychology* 16, 341-370.
- Gödel K. (1938) The Consistency of the Axiom of Choice and the Continuum Hypothesis, *Proc. Natl. Acad. Sci.* 24, 556-557.
- Gödel K. (1940) *The Consistency of the Axiom of Choice and the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Annals of Mathematical Studies, Volume 3, Princeton, Princeton University Press.
- Goldblatt R.I. (1981) Grothendieck Topology as Geometric Modality, *Zeitschr. für Math. Logik und Grundlagen der Math.*, Bd. 27, 495-529.
- Goldblatt R.I. (1984) *Topoi. The Categorical Analysis of Logic*, Amsterdam, North-Holland.
- Granger G.-G. (1994) *Formes, opérations, objets*, Paris, Vrin.
- Grossberg S. (1988) *Neural Networks and Natural Intelligence*, Cambridge (Mass.), The MIT Press.
- Grothendieck A. (1971) *Revêtements étales et groupe fondamental*, Lecture Notes in Mathematics 224, Berlin, Springer-Verlag.
- Harris E. (1987) *Lire la logique de Hegel. Commentaires de la logique de Hegel*, Lausanne, L'Age d'Homme.
- Hart E. (1986) Does the Brain Compute ?, *Behavioral and Brain Sciences* 9, 98-99.
- Henley, N.M. (1969) A Psychological Study of Semantics of Animal Terms, *Journal of Verbal Learning and Verbal Behavior* 8, 176-184.
- Heyting A. (1966) *Intuitionism*, Amsterdam, North-Holland.
- Hintikka J. (1967) Individuals, Possible Worlds and Epistemic Logic, *Nous* 1, 33-62.
- Hintikka J. (1969) *Models for Modalities*, Reidel.
- Hintikka J. (1991) *L'intentionnalité et les mondes possibles*, traduction française par N. Lavand, Presses Universitaires de Lille.
- Hirsch M. (1989) Convergent Activation Dynamics in Continuous Time Networks, *Neural Networks*, 2, 331-349.
- Hjelmslev L. (1972) *La catégorie des cas*, Munich, Wilhem Fink Verlag.
- Houdé O. & Miéville D. (eds.) (1993) *Pensée logico-mathématique. Nouveaux objets interdisciplinaires*, Paris, Presses Universitaires de France.
- Husserl E. (1950) *Idées directrices pour une phénoménologie*, Paris, Gallimard.
- Husserl E. (1959-1963) *Recherches logiques*, t. 1-3, Paris, Presses Universitaires de France.

- Jackendoff R. (1983) *Semantics and Cognition*, Cambridge (Mass.), The MIT Press.
- Jackendoff R. (1987) *Consciousness and the Computational Mind*, Cambridge (Mass.), The MIT Press.
- Jacquet J.-Ph. (1995) *Logique pseudo-consistante*, Thèse d'Université, Université Pierre et Marie Curie.
- Jacob P. (1992) Le problème du corps et de l'esprit. Essai sur les forces et les faiblesses du fonctionnalisme. In D. Andler (éd.) *Introduction aux sciences cognitives*, Paris, Gallimard.
- Jeannière A. (1959) *La pensée d'Héraclite d'Ephèse et la vision présocratique du monde*, Paris, Presses Universitaires de France.
- Kant J.-D. (1996) *Modélisation et mise en œuvre de processus cognitifs de catégorisation à l'aide d'un réseau connexionniste*, Thèse, Université de Rennes 1.
- Keil F.C. (1979) *Semantic and Conceptual Development*, Cambridge (Mass.), Harvard University Press.
- Kleene S.C. & Vesley R. (1965) *The Foundations of Intuitionistic Mathematics*, Amsterdam, North-Holland.
- Kleiber G. (1990) *La sémantique du prototype. Catégories et sens lexical*, Paris, Presses Universitaires de France.
- Könderink J. (1984) The Structure of Images, *Biological Cybernetics*, 50, 363-370.
- Kripke S.A. (1965) Semantical Analysis of Intuitionistic Logic. In J.N. Crossley & M.A.E. Dummett (eds.) *Formal Systems and Recursive Functions*, Amsterdam, North-Holland, 92-130.
- Kripke S.A. (1980) *Naming and Necessity*, Oxford, Blackwell.
- Lakoff G. (1987) *Women, Fire, and Other Dangerous Things? What Categories Reveal about the Mind*, Chicago, University of Chicago Press.
- Lakoff G. & Johnson M. (1980) *Metaphors We Live By*, tr. fr. *Les métaphores vives*, Paris, Les Editions de Minuit.
- Lambek J. & Scott P.J. (1986) *Introduction to Higher Order Categorical Logic*, Cambridge, Cambridge University Press.
- Langacker R. (1987) *Foundations of Cognitive Grammar*, Vol. 1, Stanford University Press.
- Langacker R. (1988) An Overview of Cognitive Grammar. In B. Rudzka-Ostyn (ed.) *Topics in Cognitive Linguistics*, Amsterdam, John Benjamins.
- Langacker R. (1990) *Concept, Image, and Symbol : The Cognitive Basis of Grammar*, Berlin, Mouton de Gruyter.
- Langacker R. (1991) *Foundations of Cognitive Grammar*, Vol. 2, Stanford University Press.
- Largeault J. (1992) *L'intuitionisme, Que sais-je ?*, Paris, Presses Universitaires de France.

- Largeault J. (éd) (1992) *Intuitionisme et théorie de la démonstration*, Paris, Vrin.
- Lawvere F.W., Maurer C. & Wraith G.C. (eds) (1975) *Model Theory and Topoi*, Lecture Notes in Mathematics, Vol. 445, Berlin, Springer-Verlag.
- Le Ny J.-F. (1989) *Science cognitive et compréhension du langage*, Paris, Presses Universitaires de France.
- Le Ny J.-F. (1989) Catégories, *Encyclopédie philosophique universelle*, t. 1, Paris, Presses Universitaires de France.
- Mac Lane S. & Moerdijk I. (1992) *Sheaves in Geometry and Logic. A First Introduction to Topos Theory*, Berlin, Springer-Verlag.
- Marr D. (1982) *Vision*, San Francisco, Freeman.
- Mervis C.B. & Rosch E. (1981) Categorization of Natural Objects, *Annual Review of Psychology* 32, 89-115.
- Miéville D. (1984) Logique naturelle et méréologie. In J.-B. Grize (éd.) *Sémiologie du raisonnement*, Peter Lang, 209-242.
- Miller G.A. (1978) Semantic Relations among Words. In Halle, Bresnan & Miller (eds.) *Linguistic Theory and Psychological Reality*, Cambridge (Mass.), The MIT Press.
- Miller G.A. & Johnson-Laird P.N. (1976) *Language and Perception*, Cambridge University Press.
- Millet G.A. (1969) A Psychological Method to Investigate Verbal Concepts, *Journal of Mathematical Psychology* 32, 169-191.
- Newell A. (1990) *Unified Theories of Cognition*, Cambridge (Mass.), The Harvard University Press.
- Osherson D. et al (eds) (1990) *An invitation to Cognitive Science*, Cambridge (Mass.), The MIT Press.
- Pacherie E. (1993) *Naturaliser l'intentionnalité. Essai de philosophie et de psychologie*, Paris, Presses Universitaires de France.
- Pacherie E. (éd.) (1996) Fonctionnalismes, *Intellectica* n° 21.
- Piatelli-Palmarini M. (éd.) (1979) *Théories du langage, théories de l'apprentissage, le débat entre Jean Piaget et Noam Chomsky*, Paris, Le Seuil.
- Petitot J. (1989) Hypothèse localiste, modèles morphodynamiques et théories cognitives : remarques sur une note de 1975, *Semiotica*, 77, 65-119.
- Petitot J. (éd.) (1990) Sciences cognitives : quelques aspects problématiques, *Revue de Synthèse* IV, 1-2.
- Petitot J. (1991a) Syntaxe topologique et grammaire cognitive, *Langage*, 103, 97-128.
- Petitot J. (1991b) Why Connectionism Is Such a Good Thing. A Criticism of Fodor's and Pylyshyn's criticism of Smolensky, *Philosophica*, 47, 1, 49-79.
- Petitot J. (1992) *Physique du sens. De la théorie des singularités aux structures sémio-narratives*, Paris, Editions du CNRS.

- Petitot J. (éd.) (1994) Linguistique cognitive et modèles dynamiques, *Sémiotiques*, 6-7.
- Petitot J. (éd.) (1995a) Sciences cognitives et phénoménologie, *Archives de Philosophie*, 58.
- Petitot J. (1995b) Sheaf Mereology and Husserl's Morphological Ontology, *International Journal of Human Computer Studies*, 43, 741-763.
- Petitot J. (1995c) Pour un platonisme transcendantal. In M. Panza & J.M. Salanskis (éds.) *L'objectivité mathématique. Platonisme et structures formelles*, Paris, Masson.
- Piaget J. (1978) *La formation du symbole chez l'enfant*, Delachaux & Niestlé.
- Piaget J., Henriques G. & Ascher E. (1990) *Morphismes et Catégories*, Delachaux et Niestlé.
- Posner M.L. & Keele S.W. (1968) On the Genesis of Abstract Ideas, *Journal of Experimental Psychology* 77, 353-363.
- Posner M.L. & Keele S.W. (1970) Retention of Abstract Ideas, *Journal of Experimental Psychology* 83, 304-308.
- Pottier B. (1993) *Sémantique générale*, Paris, Presses Universitaires de France
- Pruzansky S., Tversky A. & Carrol J.D. (1982) Spatial Versus Tree Representation of Proximity Data, *Psychometrika* 47, 3-24.
- Putnam H. (1975) *Philosophical Papers II*, Cambridge, Cambridge University Press.
- Putnam H. (1988) *Representation and Reality*, Cambridge (Mass.), The MIT Press.
- Pylyshyn Z. (1984) *Computation and Cognition*, Cambridge (Mass.), The MIT Press.
- Quetelet J. (1835) *Sur l'homme et le développement de ses facultés ou essai de physique sociale*, Paris.
- Quine W.V.O. (1960) *Word and Object*, Cambridge (Mass.), The MIT Press.
- Rastier F. (1981) *Sémantique et recherches cognitives*, Paris, Presses Universitaires de France.
- Restle F. (1959) A Metric and an Ordering on Sets, *Psychometrika* 24, 207-230.
- Riguet (1948) Relations binaires, fermetures, correspondances de Galois, *Bull. Soc. Math. France* 76, 114-155.
- Rips L.J., Shoben E.J. & Smith E.E. (1973) Semantic Distance and the Verification of Semantic Relations, *Journal of Verbal Learning & Verbal Behavior* 12, 1-20.
- Rivenc F. & de Rouilhan Ph. (éds) (1992) *Logique et fondements des mathématiques. Anthologie (1850-1914)*, Paris, Payot.
- Robinson A. (1974) *Non-Standard Analysis*, Amsterdam, North-Holland.
- Rorty R. (1972) Indeterminacy of Translation and Truth, *Synthèse* 23, 443-462.

- Rosch E. (1975) Cognitive Representation of Semantic Categories, *J. Exp. Psych.* 104, 192-233.
- Rosch E. & Lloyd B. (eds) (1978) *Cognition and Categorization*, Hillsdale, L. Erlbaum.
- Rumelhart D. & McLelland J. (eds) (1986) *Parallel Distributed Processing*, Cambridge (Mass.), The MIT Press.
- Ryle G. (1949) *The Concept of Mind*, Hutchinson.
- Salanskis J.-M. (1991) *L'herméneutique formelle*, Paris, Editions du CNRS.
- Salanskis J.-M. & Sinaceur H. (éds.) (1992) *Le labyrinthe du continu*, Paris, Springer-Verlag France.
- Scott D.S. (1967) A Proof of the Independence of the Continuum Hypotheses, *Math. Syst. Theory* 1, 89-111.
- Shaumyan S. (1977) *Applicational Grammar as a Semantic Theory of Natural Language*, Chicago, Chicago University Press.
- Shaumyan S. (1987) *A Semantic Theory of Natural Language*, Blommington, Indiana University Press.
- Searle J. (1983) *Intentionality*, Cambridge University Press.
- Shepard R.N. (1974) Representation of Structure in Similarity Data : Problems and Prospects, *Psychometrika* 39, 373-421.
- Shepard R.N. (1982) Multidimensionnal Scaling, Tree-Fitting and Clustering, *Science* 210, 390-398.
- Shepard R.N., Romney A.K. & Nerlove S.B. (eds.) (1972) *Multidimensionnal Scaling* (2 vols.), New York, Academic Press.
- Shoben E.J. (1976) The Verification of Semantic Relations in the Same-Difference Paradigm : An Assymetry in Semantic Memory, *Journal of Verbal Learning & Verbal Behavior* 15, 365-381.
- Simon H., 1969. *The Sciences of the Artificial*, Cambridge (Mass.), The MIT Press.
- Simpson S.G. (1988) Partial Realizations of Hilbert's program, *The Journal of Symbolic Logic*, 53, 2, 349-363.
- Smale S. (1967) Differentiable Dynamical Systems, *Bull. Amer. Math. Soc.*, 73, 747-817.
- Smolensky P. (1988) On the Proper Treatment of Connectionism, *The Behavioral and Brain Sciences*, 11, 1-23.
- Stich S. (1978) Beliefs and Subdoxastic States, *Philosophy of Science*, 45, 499-518.
- Stich S. (1983) *From Folk Psychology to Cognitive Science : The Case Against Belief*, Cambridge (Mass.), The MIT Press.
- Talmy L. (1983) How Language Structures Space. In H. Pick & L. Acredolo (eds) *Spatial Orientation : Theory, Research, and Applications*, New York, Plenum Press.

- Talmy L. (1985) Force Dynamics in Language and Thought, *Papers from the Parasession on Causatives and Agentivity at Twenty-First Regional Meeting of Chicago Linguistic Society*, 293-337.
- Talmy L. (1988) The Relation of Grammar to Cognition. In B. Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, Amsterdam, John Benjamins.
- Talmy L. (1990) Semantics and Syntax of Motion. In J. Kimball (ed.) *Syntax and Semantics*, vol. 4, New York, Academic Press, 181-238.
- Tarski A. (1938) *Logics, Semantics, Metamathematics*, Oxford, Clarendon Press.
- Ténin A. (1996) *Catégorisation et systèmes symboliques. Approches différentielle des traitements, des matériels et des tâches*, Thèse d'université, Université Paris 5.
- Thom R. (1972) *Stabilité structurelle et morphogénèse*, Paris, Benjamin.
- Thom R. (1980) *Modèles mathématiques de la morphogénèse*, Paris, Bourgois.
- Troelstra A. & van Dalen D. (1988) *Constructivism in Mathematics*, Amsterdam, North-Holland.
- Turing A. (1950) Computing Machinery and Intelligence, *Mind*, 59, 433-460.
- Tversky A. (1991) Features of Similarity, *Psychological Review* 84, 327-352.
- Vakarelov D. (1991) Logical Analysis of Positive and Negative Similarity Relations in Property Systems. In M. De Glas & D.M. Gabbay (eds.), *WOCFAI'91*, 491-499.
- Vandeloise C. (1986) *L'espace en français*, Paris, Edition du Seuil.
- Vignaux G. (1991) *Les sciences cognitives. Une introduction*, Paris, La Découverte.
- Visetti Y.-M. (1990) Modèles connexionnistes et représentations structurées. In D. Memmi & Y.-M. Visetti (éds) *Modèles connexionnistes, Intellectica*, 9-10, 167-212.
- Weyl H. (1949) *Philosophy of Mathematics and Natural Science*, Princeton, Princeton University Press.
- Wildgen W. (1982) *Catastrophe Theoretic Semantics*, Amsterdam, Benjamin.
- Wildgen W. (1986) Portée et limites d'une application de la théorie des catastrophes en linguistique. In J. Petitot (éd.) *Logos et théorie des Catastrophes*, Genève, Editions Patino.
- Wittgenstein L. (1953) *Philosophische Untersuchungen*, Oxford, Blackwell.
- Zeeman C. (1977) *Catastrophe Theory : Selected Papers 1972-1977*, New York, Addison-Wesley.