L'agent rationnel abstrait, objet de l'IA ?

Denis Berthier

INSTITUT NATIONAL DES TÉLÉCOMMUNICATIONS
9 rue Charles Fourier, 91011 Evry Cédex

RÉSUMÉ⊡ *Nous montrons qu'à l'intérieur du strict paradigme du système de symboles concrets, les évolutions de l'intelligence artificielle, sous la poussée combinée des avancées conceptuelles, de la pratique du développement de systèmes experts, et de préoccupations méthodologiques, convergent vers une conception nouvelle de l'objet de l'IA - celle d'un agent rationnel abstrait en situation dans une organisation. Il en résulte un décalage du centre de gravité de l'IA, et de certains de ses concepts centraux, comme la connaissance, du champ mental vers le champ socioculturel.*

ABSTRACT⊡ *We show that, within the strict physical symbol system paradigm, conceptual advances in artificial intelligence, together with the practice of expert systems development and methodological issues, jointly lead AI to a new vision of its object - that of a rational abstract agent, situated in an organization. As a result the focus of the field and of some of its main concepts, such as knowledge, shifts from the mental towards the socio-cultural domain.*

MOTS-CLÉS⊡ *Méthodologie, épistémologie, agent rationnel, modèle, représentation*

KEY-WORDS⊡ *Methodology, epistemology, rational agent, model, representation*

1. Introduction

Nous rattachons la distinction classique de Newell entre les niveaux des savoirs (knowledge level) et des symboles (symbol level) à une distinction générale, analysée en préambule, entre un modèle et ses représentations, et nous étudions une ligne d'évolution de l'intelligence artificielle (IA) sous la poussée combinée de trois facteurs⊡
- certaines avancées conceptuelles et théoriques, telles que formalisées dans l'école de Newell, et concrétisées dans le système SOAR,
- les contraintes résultant de la pratique du développement de systèmes experts,
- les préoccupations méthodologiques plus récentes, dont nous prenons KADS comme exemple représentatif, pour focaliser la discussion.

L'évolution en question ici est celle qui s'est produite à l'intérieur du paradigme représentationnel et computationnel de l'intelligence (la connaissance, la cognition, etc). Partant de la notion fondatrice de système de symboles concrets, passant par l'assertion d'un niveau des symboles (où règnent les questions de représentation et manipulation informatique des connaissances), puis par la mise à jour d'un niveau des savoirs (où apparaît la question de la modélisation), elle aboutit aux méthodologies de développement de systèmes à base de connaissances (où se précise ce qui est concerné par la modélisation).

Nous ne prétendons pas rendre ainsi compte de toute l'évolution de l'IA. En particulier, après le renouveau de la recherche sur les réseaux de neurones, une autre ligne pourrait être identifiée, qui conduirait à mettre en valeur l'idée d'une émergence du symbolique dans un système de neurones formels. Une autre encore, s'appuyant sur certaines avancées de la robotique ([BRO 91]), sur l'étude des aspects développementaux ou sur des considérations plus philosophiques, pourrait insister sur le rôle des interactions avec l'environnement et conduire à mettre en cause la notion même de représentation, plus radicalement qu'elle ne peut l'être dans les réseaux de neurones. Mais nous considérons que la ligne ici étudiée est représentative de l'évolution qu'a connu ce qu'on peut considérer comme le noyau dur de l'IA, à l'intérieur du strict paradigme du système de symboles concrets. Ce que nous appelons IA par la suite est cette partie de l'informatique qui se situe dans ce paradigme, qui reste dominante en termes d'applications, mais qui n'est certainement pas le sens le plus général que l'on puisse donner à ce terme.

Nous montrons que les évolutions décrites convergent vers une conception de l'objet de cette discipline et de certains de ses concepts très différents de ce qu'elle a pu être dans sa période héroïque. En particulier, nous suggérons que□
- l'objet de l'intelligence artificielle, tel qu'il résulte de ces considérations, n'est ni l'intelligence, ni la cognition, ni la connaissance (avec tout le flou qui entoure ces notions), mais précisément la modélisation constructive d'agents rationnels abstraits en situation dans des organisations, et la structuration, la formalisation et l'informatisation des savoirs nécessaires à leur fonctionnement,
- le savoir ne peut être conçu indépendamment de sa mise en perspective dans le cadre d'une organisation sociale, et de sa subordination à ses objectifs,
- le centre de gravité de l'IA, qui se situait jusqu'alors dans le champ du mental, s'en trouve décalé vers le champ socioculturel.

Nous analysons enfin dans ce cadre les trois grands moments du traitement du savoir en IA□capitalisation, opérationnalisation, interprétation.

2. Préliminaires sur les modèles et leurs représentations

2.1 La notion de modèle

Notre ambition ici est d'examiner (et non de définir) un concept fondamental de l'épistémologie, qui est en pratique généralement considéré comme allant de soi et rarement questionné.

Convenons de dire qu'un modèle porte sur un *objet* - au sens abstrait le plus général du mot objet. L'"objet" d'un modèle peut être un système physique, réel ou imaginaire, un processus, un ensemble de concepts ou de notions interconnectés. Modéliser, c'est créer une relation asymétrique entre deux termes, dont le premier est considéré comme mieux formalisé que le deuxième, dans une certaine perspective.

Un modèle est une formalisation de certains aspects de son objet. C'est un point de vue particulier sur son objet, choisi dans un but déterminé, qui n'épuise pas en principe tous les aspects possibles de cet objet. A un modèle est donc indissolublement associé un *domaine de validité*, qu'il est important de connaître pour ne pas appliquer inconsidérément le modèle au-delà. (Le domaine de validité peut s'avérer plus large que celui résultant de la fonction qu'un modèle peut avoir été initialement destiné à remplir). Aussi les erreurs suivantes constituent-elles les trois "péchés capitaux" de la modélisation, qui sont à la source de nombreux malentendus□ confondre un modèle et son objet, choisir un modèle inadapté au point de vue que l'on veut privilégier, outrepasser sans discernement les limites du domaine de validité d'un modèle.

Parfois, le domaine de validité peut recouvrir tellement d'aspects connus de l'objet que l'on a tendance à oublier d'autres aspects possibles de cet objet, simplement parce qu'ils n'entrent pas dans le modèle. Parfois aussi, il peut arriver qu'un modèle soit tellement prévalent culturellement qu'il devienne difficile de distinguer le modèle de son objet. Parfois encore, l'objet est tellement complexe que les seuls modèles accessibles au raisonnement font par nécessité des hypothèses hyper-simplificatrices.

Plus radicalement, le lien de dépendance entre un modèle et son objet peut être en quelque sorte inversé (par rapport à une conception courante qui voudrait que le "monde" préexiste avec des découpages qu'il n'y aurait qu'à dévoiler)□ l'objet peut ne pas préexister au modèle, ne devenir pensable qu'avec l'apparition d'un modèle. Un exemple saisissant est celui de certaines particules élémentaires, comme les quarks, dont l'apparition dans le champ de la pensée des physiciens résulte de l'existence d'un modèle mathématique de calcul (lui-même s'insérant dans un modèle plus vaste). Cela ne signifie pas pour autant que la part de réalité que le modèle a permis d'isoler soit épuisée par le modèle. Pour nous, la plupart des termes généraux utilisés en IA (l'intelligence, les connaissances, etc) sont de ces notions vagues qui n'acquièrent quelque précision qu'au travers des modèles permettant de les penser.

La démarche de modélisation est un instrument universel de la pensée scientifique. Un modèle est l'équivalent dans le domaine de la pensée scientifique de ce qu'est un référentiel en physique. Un événement physique ne peut être décrit précisément que par rapport à un certain référentiel (système de coordonnées), et les lois physiques générales, qui sont des lois d'invariance par changement de référentiel, trouvent leur expression concrète dans chaque référentiel particulier. De même, un modèle constitue un référentiel nécessaire de la pensée. Et, de même qu'il peut ne pas exister de référentiel unique couvrant l'ensemble de l'univers (en relativité générale), il peut ne pas exister de modèle couvrant tous les aspects de son objet. Plusieurs modèles peuvent s'appliquer à un "même" objet. Dans ce cas, dire que l'objet est le "même" peut constituer une affirmation très forte sur l'"existence" d'un "objet" commun à ces modèles - tout comme le principe de relativité, générale ou galiléenne, constitue une affirmation forte d'existence d'une réalité physique objective

(les lois d'invariance par changement de référentiel étant considérées comme expression mathématique de l'indépendance par rapport à l'observateur).

Un modèle est descriptif, mais ne porte pas nécessairement en soi (sauf à titre virtuel) tous les modes opératoires susceptibles de l'exploiter, même d'ailleurs s'il est de nature rigoureusement mathématique. Ce caractère non totalement opératoire peut provenir de deux origines, que nous ne ferons qu'illustrer par un exemple issus de l'univers de la modélisation mathématique classique. Considérons un modèle constitué par un système d'équations différentielles et/ou aux dérivées partielles, et supposons que l'on puisse prouver théoriquement que le système a une solution et une seule. Nous sommes donc dans la meilleure situation théorique possible, et le modèle s'avère déjà ainsi être un "bon" modèle et apporter une contribution fondamentale à l'étude de son objet. Néanmoins, premièrement, une preuve théorique ne constitue pas nécessairement une méthode constructive, algorithmique, de calcul (songer aux démonstrations par l'absurde), et deuxièmement, même si une telle méthode existe en théorie, elle peut être inutilisable en pratique (par exemple, à cause de problèmes d'instabilité due aux erreurs d'arrondi).

2.2 Représentations d'un modèle

Nous n'aborderons pas le problème de la représentation en général, nous intéressant exclusivement aux représentations formalisées de modèles. Nous excluons ainsi de nos considérations l'étude de la relation directe qui peut exister entre un "objet" et une représentation quelconque (par exemple une photo).

Partons néanmoins d'une conception courante□ représenter X par Y, c'est considérer la présence de Y comme substitut valable à la présence de X. Dès lors, la question qui se pose est□ d'où un modèle peut-il être absent, pour qu'une représentation puisse s'y substituer en ce lieu? Nous avons vu qu'un modèle est en partie absent du champ opératoire. C'est donc dans ce champ qu'il est heuristiquement raisonnable d'exprimer en quoi une représentation peut s'y substituer.

Nous définirons une représentation d'un modèle comme une expression de ce modèle qui introduit des éléments dépourvus de sens du strict point de vue de la modélisation, mais qui supportent des modes opératoires supplémentaires, compatibles avec le modèle, et destinés à faciliter le traitement pragmatique des questions qui ont conduit à construire le modèle. Cette définition s'accorde avec l'usage du terme en IA dans l'expression "représentation des connaissances" (voir section 4.3), ou en algèbre dans l'expression "représentation linéaire d'un groupe" (où un groupe abstrait est représenté comme un groupe d'isomorphismes d'un espace vectoriel, et éventuellement de matrices inversibles). C'est dans la mesure où ces éléments étrangers au modèle et ces modes opératoires ne sont pas impliqués de manière nécessaire par le modèle qu'un modèle n'est pas, statutairement, (ou pas seulement) une représentation particulière.

Cette définition, évidemment très restrictive par rapport à la notion générale de représentation, autorise cependant un modèle à avoir des représentations extrêmement

variées, basées sur des modes opératoires très différents, comme le montrent certains des exemples suivants. Certaines représentations peuvent même être dérivées les unes des autres par affinement des modes opératoires.

Tout comme un modèle peut apparaître antérieurement à son objet, une représentation peut apparaître avant l'explicitation du modèle qu'elle veut représenter, le modèle émergeant alors comme une abstraction de la représentation. Mais, une représentation étant donnée, il y a en général ambiguïté sur un possible modèle sous-jacent, en particulier sur ce qui doit être considéré comme significatif. Aussi le "péché capital" en la matière est-il d'utiliser des représentations sans avoir clairement spécifié le modèle qu'on entend ainsi représenter. Chacun sait l'impact de ce problème sur les réflexions épistémologiques des spécialistes de la représentation des connaissances ([Woo 75], [Bob 75], [Fin 79], [Bra 79] peuvent être lus après coup comme la quête de modèles génériques sous-jacents aux systèmes de représentation des connaissances de l'époque).

2.3 Exemples de modèles et de représentations possibles

Nous donnons quelques exemples de modèles tellement courants qu'on a tendance à les prendre pour la réalité et qu'on les emploie sans même y penser comme modèles. Pour illustrer la différence entre modèle et représentation, et pour montrer que des représentations d'un même modèle peuvent être de diverses natures, nous donnons simultanément des représentations possibles de ces modèles; là aussi peut exister une forte tendance à confondre les niveaux, en prenant une représentation pour le modèle.

2.3.1 L'espace physique est *modélisé* mathématiquement par un espace euclidien de dimension 3. Il a fallu attendre la découverte d'autres types d'espaces mathématiques pour comprendre qu'il s'agit là d'un modèle particulier de l'espace physique - du modèle euclidien, précisément. (La relativité générale nous a maintenant habitués à considérer d'autres modèles). Cette modélisation a une forte cohérence□ les transformations géométriques rigides sont modélisées par des applications linéaires, la composition de transformations correspond à la composition des applications associées, etc. Ce modèle permet de nombreux calculs relevant de l'algèbre linéaire.

Cet espace euclidien et ces applications linéaires peuvent elles-mêmes être *représentées* algébriquement par des triplets et des matrices de nombres réels, et les opérations du modèle peuvent être exprimées en termes d'opérations algébriques sur ces représentations; à ce stade, la représentation a introduit un élément étranger au modèle, à savoir une base particulière de l'espace euclidien. Ces représentations algébriques peuvent à leur tour recevoir diverses représentations informatiques, selon le langage employé, les types d'objets définissables dans ce langage, les impératifs de conception liés à l'utilisation prévue pour ces objets.

2.3.2 Les forces et les vitesses sont *modélisées* mathématiquement comme éléments d'un espace vectoriel.

Ces vecteurs peuvent être *représentés* algébriquement, comme précédemment, par des triplets de nombres réels, la composition de forces ou de vitesses étant alors calculée, composante par composante, par des sommes algébriques.

Ces vecteurs peuvent également être *représentés*, géométriquement, par des petites flèches. A cette représentation, on peut appliquer la règle du parallélogramme pour construire avec une règle et un compas la résultante de deux forces ou de deux vitesses.

Autrement dit, nous avons deux représentations, algébrique et géométrique, compatibles avec le modèle vectoriel, mais supportant des modes opératoires fondamentalement différents. Ces modes sont évidemment cohérents, mais il est intéressant de remarquer que la cohérence n'est visible qu'à travers la référence à un même modèle mathématique. La cohérence en l'absence de cette référence pourrait d'ailleurs subir quelques accrocs si l'on se préoccupe des erreurs d'arrondi d'un côté, des imprécisions de tracé de l'autre.

2.3.3 Qu'entend-on en physique par le "modèle planétaire de l'atome de Bohr"? L'image apparaît d'électrons tournant autour du noyau comme les planètes tournent autour du soleil. Mais cette *analogie* ne vient à constituer un *modèle* que dans la mesure où elle est considérée comme allusion au modèle newtonien de la gravitation. (Chacun sait les difficultés de transposition de ce modèle à l'atome). Ainsi en est-il de beaucoup de "modèles" physiques, ou "modéliser X par Y" signifie souvent "modéliser X par le modèle habituel de Y".

2.4 Modèles fonctionnels, modèles de fonctionnement et représentations

Dans le cas où l'objet de la modélisation est un système (inter-)actif, celle-là peut porter sur deux aspects différents du système⊡
- son comportement observable, tel que perçu par un observateur extérieur,
- les processus internes du système produisant ce comportement observable.

Dans le premier cas, on parlera de *modèle fonctionnel (ou modèle de compétence*), entendant que l'on ne s'intéresse qu'à l'ensemble abstrait des fonctions d'entrée-sortie du système, indépendamment de la manière dont elles sont produites, c'est-à-dire à la spécification fonctionnelle du système.
Dans le second cas, on parlera de *modèle de fonctionnement (ou modèle de performance*, selon un anglicisme malheureux mais consacré dans les milieux linguistiques et informatiques).

Dans l'exemple d'un analyseur syntaxique (objet de la modélisation), la spécification du langage accepté en entrée et des arbres syntaxiques correspondants en sortie constitue un modèle fonctionnel, la spécification de l'algorithme d'analyse syntaxique, qui permet de construire ces arbres, constitue un modèle de fonctionnement.

Bien qu'on attende souvent d'un modèle de fonctionnement qu'il soit opératoire, ce n'est pas toujours le cas⊡le comportement interne d'un système peut être décrit par des équations différentielles du type de celles évoquées à la première section. Si, par

exemple, il était possible de modéliser un cerveau par des milliards d'équations différentielles représentant le comportement individuel de ses neurones, il est probable qu'on aurait quelque difficulté à intégrer ce système d'équations.

La distinction entre modèle fonctionnel et modèle de fonctionnement doit être maintenue très clairement dans le cas où l'objet est un système dynamique S1□ un modèle fonctionnel de S1 ne décrit que le comportement dynamique observable de S1, sans se préoccuper de son état interne, non observable (à supposer déjà que la notion d'état interne ait un sens clair pour S1 - quel est par exemple l'état interne d'un système intelligent?). Ce modèle de S1 peut avoir pour représentation un système dynamique S2 défini explicitement par son fonctionnement interne, et produisant le même comportement observable que S1. (Les spécialistes de théorie mathématique des systèmes disent que S1 et S2 sont observationnellement équivalents). Il est tout-à-fait crucial pour la suite de notre propos de bien comprendre que S2 n'a aucune raison a priori de constituer un modèle de fonctionnement de S1□ il peut en effet exister de nombreux systèmes observationnellement équivalents à S1, et très différents structurellement les uns des autres.

Il est assez clair que la logique mathématique ne constitue pas un modèle cognitif du raisonnement en action, mais un modèle abstrait de l'ensemble des déductions légitimes. C'est donc un modèle fonctionnel du raisonnement (d'un raisonneur idéal, d'ailleurs), non un modèle de fonctionnement de ce raisonneur. Par rapport à ce modèle, tout algorithme de calcul déductif bâti sur une axiomatisation particulière, et a fortiori tout système informatique le mettant en œuvre, est une représentation de ce modèle fonctionnel. La conclusion du paragraphe précédent revient à dire que c'est a priori *seulement* une représentation du modèle fonctionnel, et non un modèle de fonctionnement du raisonneur. [En logique, l'utilisation technique du mot "modèle" dans la "théorie des modèles" n'a pas de rapport direct avec le sujet discuté ici].

3. Le paradigme du système de symboles concrets

3.1 L'hypothèse du système de symboles concrets

L'hypothèse du système de symboles concrets (the physical symbol system hypothesis) est une formulation tardive, par Newell (1976, *in* [NEW 80]), de l'hypothèse de base à l'origine de l'IA, restée longtemps implicite, qui peut se résumer ainsi□*une condition nécessaire et suffisante pour qu'un système physique puisse manifester une forme générale d'intelligence est qu'il soit un système universel de manipulation formelle de symboles concrets*. Par "concret", on entend deux choses□d'une part, le système est physiquement réalisable, par exemple sous forme de programme informatique; d'autre part, les symboles en question n'ont pas de sens par eux-mêmes, ils ne sont que des objets de manipulation formelle, de calcul. Remarquons que l'hypothèse concerne la *possibilité* de l'intelligence, elle ne dit rien sur la manifestation effective de comportements intelligents, ni sur l'éventuelle nécessité de structures organisationnelles plus élaborées construites sur ce système de symboles.

Newell et Simon, en 1976, considèrent que cette notion de système de symboles concrets a été la contribution la plus importante de l'IA à l'étude de l'*intelligence*. Comme le souligne Newell, cette notion de symbole comme objet de calcul est radicalement distincte des divers sens accordés à ce mot dans la littérature, la philosophie ou les sciences humaines classiques. Néanmoins, nous pensons qu'elle est en fait totalement cohérente avec la vision de l'école de pensée structuraliste, et en particulier la théorie du signifiant. Ainsi Deleuze écrit-il, à propos du concept central de structure : "Il s'agit d'une combinatoire portant sur des éléments formels qui n'ont par eux-mêmes ni forme, ni représentation, ni contenu, ni réalité empirique donnée, ni modèle fonctionnel hypothétique, ni intelligibilité derrière les apparences" ([DEL 73], p. 303). Bien que les deux courants se soient largement ignorés (voir cependant [ECO 68]), il n'est pas inutile de percevoir l'IA comme un accomplissement formel du structuralisme, lequel peut lui offrir en retour une base d'insertion culturelle plus large que son cercle de spécialistes. (Pour une introduction au structuralisme, voir [DUC 68], [ECO 68] ou [DEL 73]).

Les systèmes de symboles concrets concernés par l'hypothèse de Newell sont dits "universels", c'est-à-dire qu'ils sont supposés capables de produire toute fonction d'entrée-sortie calculable. Formellement, un système de symboles concrets, tel que défini par Newell, est donc simplement une machine de Turing, et ce que nous appellerons *le premier volet de la thèse de Newell* peut s'énoncer ainsi : *pour qu'un système physique possède potentiellement une forme générale d'intelligence, il faut et il suffit qu'il soit (équivalent à) une machine de Turing*. Plus brutalement, la thèse se réduit à : *un ordinateur digital possède potentiellement les capacités nécessaires à la manifestation d'une forme générale d'intelligence*. Sous cette formulation, le présupposé est transparent : une condition nécessaire et suffisante à la manifestation potentielle de l'intelligence est la capacité calculatoire universelle. *La thèse est ainsi le fondement d'une vision computationnelle de l'intelligence,* vision qui aura des répercussions dans les sciences cognitives, loin au-delà de l'IA.

3.2 La hiérarchie des niveaux informatiques et le niveau des symboles

Si toutefois l'on devait s'arrêter là, la notion de système de symboles concrets serait totalement inutile, puisque non spécifique, et puisque la thèse pourrait se formuler sans elle, en termes de machine de Turing. La notion de système de symboles concrets est une notion non formelle, de même que la notion de machine. Toutes deux, indépendamment, trouvent une expression formelle dans la notion de fonction calculable. Mais les équivalences affirmées par Church et Newell sont des équivalences abstraites. Elles ne disent rien des aspects pragmatiques des machines ou systèmes de symboles concernés, et ne garantissent pas que l'on atteigne ainsi le niveau le plus adéquat de description du fonctionnement d'une machine ou d'un système intelligent.

En ce qui concerne les machines, la hiérarchie classique des niveaux informatiques (voir par exemple [GAN 93]) et l'invention de langages de niveaux de plus en plus élevés, représentant des paradigmes de programmation différents, constituent la preuve pratique que l'universalité d'une machine laisse la place à de multiples points de vue complémentaires sur son fonctionnement. En particulier, il est possible de présenter une machine universelle sous forme d'un système de manipulation de

symboles concrets (Lisp, par exemple, constitue la spécification abstraite d'une telle machine); une telle présentation constitue, dans la hiérarchie des niveaux informatiques, un *niveau des symboles*, effectivement implémentable au-dessus des niveaux classiques. Dès lors, *le deuxième volet de la thèse de Newell* peut s'énoncer ainsi□ *le niveau des symboles est pragmatiquement approprié à l'expression d'une forme générale d'intelligence.*

Ultérieurement, Newell complétera l'hypothèse du système de symboles concrets, relative à la potentialité de manifestation de l'intelligence, par une thèse beaucoup plus spécifique□ l'architecture du système Soar, développé dans son équipe ([LAI 86], [LAI 87]), constituerait un modèle général de l'intelligence. Nous ne pousserons pas ici l'analyse de ces développements dans leurs aspects cognitifs, qui nous éloigneraient de notre propos. Mais Soar, système de manipulation de symboles, étant une machine universelle, est par là-même un système de symboles concrets. Soar constitue un exemple intéressant, à un tout autre niveau structurel que le langage Lisp sur lequel il est implémenté, illustrant le vaste champ de possibilités que laissent ouvertes les notions de système de symboles concrets ou de niveau des symboles - et du même coup le manque relatif de spécificité du deuxième volet de la thèse de Newell.

4. Une ligne d'évolution de l'intelligence artificielle

Cette section a pour objectif de décrire les évolutions dans la manière d'aborder certaines questions fondamentales de l'IA, à l'intérieur du paradigme du système de symboles concrets. Ne pouvant faire ici un état de l'art détaillé de ces évolutions, nous nous contenterons de tracer quelques vecteurs tangents à la ligne que nous voulons suggérer.

4.1 De la recherche de mécanismes généraux de raisonnement à la primauté des connaissances

Dans la première phase du développement de l'IA, l'accent est mis sur l'intelligence, conçue comme reposant essentiellement sur des mécanismes généraux de raisonnement, et le paradigme dominant peut s'analyser a posteriori comme construit sur le postulat suivant□ l'objet de l'IA est l'intelligence, comme capacité universelle de résolution de problèmes, et tout problème d'IA peut se formuler comme l'exploration d'un espace de recherche, pour atteindre un état but à partir d'un état origine. Un espace de recherche est constitué d'un ensemble d'"états", représentant chacun ce qui est connu du système à un instant donné, et de transitions entre ces états, représentant chacune un pas possible dans le raisonnement.

Ce postulat se traduit simplement en termes mathématiques□ tout problème d'IA est la recherche, dans un graphe orienté, d'un chemin entre deux nœuds. (Cela est quelque peu caricatural, car les choix de représentation liés à la construction du graphe sont ainsi passés sous silence). En pratique, les problèmes d'IA étant (par définition, pourrait-on dire) complexes, le graphe à explorer est très gros et les algorithmes classiques se heurtent au problème de l'explosion combinatoire (leur complexité étant exponentielle par rapport à la taille du graphe). Le problème central de l'IA était ainsi de déterminer des algorithmes efficaces d'exploration de graphes.

Le résultat de la conception de l'IA comme exploration d'un espace de recherche est bien connu□ à une période d'euphorie, pendant laquelle des résultats spectaculaires pour l'époque étaient atteints, comme la démonstration automatique des premiers théorèmes des Principia Mathematica de Russell et Whitehead, a succédé une certaine déception, liée aux problèmes d'explosion combinatoire. Pour contourner ces problèmes, on a commencé par vouloir guider l'exploration du graphe par des heuristiques. Celles-ci sont d'abord apparues sous forme de méthodes générales d'exploration (algorithmes A* et alpha-béta), puis comme liées au problème à traiter, sous forme de connaissances spécifiques.

Ainsi est apparue progressivement la découverte fondamentale de l'IA, à nos yeux deuxième pierre de fondation de la discipline□ *l'intelligence ne résulte pas (principalement) de méthodes générales de raisonnement, mais de la disponibilité de connaissances spécifiques (et massives).*

4.2 Connaissances et représentation des connaissances

Une première conséquence de la découverte de la nécessité fondamentale de disposer de connaissances spécifiques est un changement de l'axe de recherche prioritaire, qui devient dès lors la représentation des connaissances. De nombreux formalismes apparaissent, de plus haut niveau structurel que le langage Lisp, les uns basés directement sur la logique des propositions ou des prédicats, les autres sur des approches plus ou moins formelles□ réseaux sémantiques, frames, règles de production, objets, graphes conceptuels.

La définition générale suivante laisse la place à une très large diversité de formalismes. Un *système de représentation de connaissances* est un ensemble de structures de données et de procédures opérant sur ces structures de manière cohérente avec l'interprétation que l'on veut en faire en termes de connaissances, et en particulier des déductions que l'on veut tirer des connaissances représentées. En ce sens, la représentation des connaissances exige des procédures de manipulation plus complexes que la représentation, par exemple, de vecteurs. Un tel système, qui se situe au niveau des symboles dans la hiérarchie informatique, constitue un exemple de système de symboles concrets.

La première phase de recherche sur les systèmes de représentation des connaissances est marquée par une ambiguïté fondamentale. De la lecture de nombreux travaux ressort l'impression que ces systèmes prétendent représenter des réalités naturelles "internes", qu'il suffirait d'observer avec un regard neutre et de traduire dans le système. (L'exemple le plus caricatural est la figure "a cat sitting on a mat" dans le livre, pourtant tardif, de Sowa ([SOW 84], p. 70), avec l'entonnoir alambiqué représentant le système perceptif et cognitif qui produit le graphe conceptuel approprié à la perception d'un chat assis sur un tapis).

Nous voyons dans cette *précipitation*, de la découverte du rôle des connaissances vers la recherche sur la représentation des connaissances, la manifestation d'un présupposé philosophique resté plus ou moins implicite dans le paradigme du système de symboles concrets - présupposé jamais remis en cause dans la communauté IA (au sens large) jusqu'à une époque récente, même s'il n'aurait pas

nécessairement été consciemment accepté tel quel□le monde perçu est représentation du réel, les connaissances constituent un stock de représentations, l'intelligence est manipulation de représentations.

Le deuxième volet de la thèse de Newell consiste alors à postuler (comme les structuralistes) que ces représentations sont discrètes et que les manipulations effectuées sur elles sont formelles, a-signifiantes par elles-mêmes - ce qui autorise à les traduire sous forme informatique (premier volet). Le paradigme du système de symboles concrets apparaît ainsi comme une spécification à implications techniques d'un paradigme philosophique général.

Dans ces conditions, quoi d'étonnant à la prétention de simplement extraire les connaissances de la tête d'un expert pour les coder dans un système informatique approprié; quoi d'étonnant au sentiment d'accéder en direct à une réalité "interne"? Quoi de plus tentant que d'ignorer, ou rejeter, l'idée de modélisation, omniprésente dans les autres sciences, à partir du moment où ces représentations internes tiennent implicitement la place d'une espèce de *modèle naturel* du réel?

4.3 Représentation et modélisation, ou niveau des symboles et niveau des savoirs

Mais de telles prétentions n'ont pu résister à un examen critique. Toute la littérature sur les problèmes épistémologiques liés à la représentation des connaissances ([WOO 75], [BOB 75], [FIN 79], [BRA 79], etc) et la multiplicité même des systèmes développés, que nous ne pouvons détailler ici, font échec à ce point de vue. Encore moins que les quarks, l'intelligence, les connaissances, la mémoire, la cognition, ne peuvent être considérés comme des réalités naturelles qu'il suffirait d'observer avec un regard neutre□ce sont des constructions intellectuelles, dépendantes du cadre théorique qui seul permet de leur donner sens, et historiquement situées. Le regard participe à la création de son objet.

Dès lors, se pose la question□ *qu'est-ce qui est significatif dans les divers systèmes informatiques de représentation des connaissances, et qu'est-ce qui est accessoire de représentation? Autrement dit, quel est le modèle sous-jacent à ces représentations*; en particulier, visent-elles des modèles fonctionnels ou de fonctionnement? Faute de se poser ces questions, il était inévitable que l'on débouche sur une "accablante diversité" des approches des différentes équipes travaillant sur la représentation des connaissances.

Faisant ce constat en 1980, Newell remarqua qu'il faudrait dépasser les recherches sur les systèmes de représentation des connaissances, relevant du niveau informatique des symboles, pour s'attacher aux connaissances elles-mêmes. L'introduction par Brachman ([BRA 79]) de principes de structuration des réseaux sémantiques à un niveau épistémologique où pouvait être définie une sémantique précise des liens, avait permis de commencer à dégager une approche de la représentation des connaissances plus conceptuelle, moins liée à leur implémentation directe dans un réseau, mais dont les choix épistémologiques restaient néanmoins très liés aux systèmes existants. Newell va plus loin, et postule *un niveau indépendant de toute représentation, le niveau des savoirs (knowledge level), au-dessus du niveau des symboles* ([NEW 82]).

Le niveau des savoirs est concerné par la modélisation du comportement observable du système, c'est-à-dire du comportement tel qu'il peut être perçu, compris et prévu par un observateur extérieur; il ne s'intéresse pas à la manière dont le fonctionnement interne produit ce comportement observable. *Ce qui est visé à ce niveau est donc un modèle fonctionnel du système, et non un modèle de fonctionnement*. Ainsi, le niveau des savoirs ne vise pas à la modélisation cognitive d'un expert du domaine, contrairement aux présupposés implicites des pratiques alors dominantes d'extraction des connaissances.

À ce niveau, *le système est modélisé comme un agent abstrait* descriptible en termes de buts, d'actions et de connaissances, et régi par le principe de rationalité suivant⊡ l'agent utilise ses connaissances pour déterminer quelles actions entreprendre pour atteindre ses buts. Les connaissances spécifiées à ce niveau sont celles nécessaires à la réalisation de ces buts.

Cette distinction entre niveau des savoirs et niveau des symboles est à l'origine d'une évolution considérable en IA (qui a pris une dizaine d'années pour se concrétiser dans la pratique)⊡ d'une part, sont proclamées l'activité de modélisation, et la différence entre modélisation et représentation; d'autre part, ces deux niveaux deviennent la justification d'une démarche méthodologique qui vise à séparer la spécification des connaissances nécessaires au traitement d'un problème de leur représentation dans un système particulier. Le niveau des savoirs devient une spécification de ce que, au niveau des symboles, on doit être capable de réaliser à l'aide du système de représentation. L'approche antérieure du développement d'un système expert est désormais comprise comme le court-circuitage du niveau des savoirs, résultant de la confusion entre un modèle et ses représentations.

Cependant, l'histoire ne s'arrête pas là. Les sections qui suivent concernent des évolutions complémentaires, indiquant la nécessité⊡
- d'introduire, d'une manière ou d'une autre, des principes de *structuration globale* des connaissances d'un système,
- de préciser les modèles constituant cet agent, jusque là très abstrait, au niveau des savoirs.

4.4 Des "atomes de connaissance" aux "sources de connaissances"

La découverte de la primauté de la connaissance s'est aussi traduite par l'apparition des "systèmes experts". Le credo à l'origine des systèmes experts, qui est resté la conception dominante de la première génération, est le principe de séparation des connaissances, en général exprimées sous forme de règles, d'avec les procédures, constituant le moteur d'inférence, qui les exploitent.

Dans cette première génération, il était considéré comme important que toutes les règles soient "à plat" dans la base de connaissances, laquelle ne devait pas être structurée. Cela allait de pair avec le principe de séparation et avec l'idée que chaque règle constitue en soi un atome de connaissance et que ces atomes sont indépendants les uns des autres. Cette conception est parfaitement cohérente avec l'idée que les connaissances sont déjà là (dans la tête de l'expert) et qu'il suffit de les extraire et les traduire une à une dans le langage de représentation. Elle est aussi parfaitement

cohérente avec la notion d'espace de recherche□ chaque instanciation de règle définit une transition permettant de passer d'une étape du raisonnement à la suivante, et toutes sont sur un même plan, comme les arcs d'un graphe; le moteur d'inférence indépendant des règles correspond à un algorithme de parcours de graphe, indépendant du graphe. Cela convenait aussi parfaitement aux marchands de miracles, qui étaient légion dans les années 80, et qui y trouvaient un argument commercial imparable□ tout le monde pouvait sans aucune formation utiliser leurs produits.

Face au dogme, la pratique s'accommodait d'indulgences, comme l'utilisation des particularités du moteur d'inférence pour coder implicitement des connaissances relatives au contrôle du raisonnement. Les utilisateurs d'OPS5, par exemple, sont en général experts des astuces de contrôle ([FOR 77], [BRO 86]). Ce qui avait pour effet de rendre la base de connaissances et le fonctionnement résultant du système à peu près totalement incompréhensibles pour les non initiés. (Ces remarques sur les principes ne doivent pas faire oublier les réalisations remarquables accomplies avec ces systèmes).

La structuration des bases de connaissances est d'abord apparue comme une nécessité pratique pour faire face à l'anarchie résultant de la mise à plat des règles. Elle s'est traduite par l'apparition d'architectures plus complexes que les premiers moteurs d'inférence, architectures qui sont d'un niveau structurel plus élevé que des systèmes comme OPS5, et a fortiori que le langage Lisp. Nous n'en donnerons que deux exemples□
- les architectures de "tableau noir" (blackboard) ([HAY 83], [HAY 85]) mettent en œuvre des "sources de connaissances" (knowledge sources, KS) multiples, chacune, spécialisée dans la résolution d'un problème ou d'une partie de problème, pouvant être constituée par un ensemble de règles, et qui communiquent toutes par l'intermédiaire du tableau noir, sorte de mémoire partagée assurant le contrôle de la coopération entre sources de connaissances;
- l'architecture du système Soar ([NEW 80b], [LAI 86], [LAI 87]), sur laquelle nous reviendrons.

Dans ces architectures, les connaissances ne sont plus codées à plat, mais sont au contraire fortement structurées en termes de "sources de connaissances" ou d'"espaces de problèmes", et l'idée d'atomicité des connaissances fait place à celle de modularité, caractérisée par un niveau de granularité plus gros. Si ces systèmes sont bien encore au niveau des symboles, cette structuration macroscopique des connaissances ne saurait être totalement étrangère au niveau des savoirs.

4.5 De l'espace de recherche aux espaces de problème

L'introduction d'un niveau des savoirs institutionnalise la séparation entre les connaissances et les procédures susceptibles de les exploiter, puisque, par définition, ce niveau ignore ces procédures. Mais en même temps, la distinction entre niveaux des savoirs et des symboles libère ce dernier du souci de maintenir la séparation, garantie par ailleurs.

Ainsi, l'architecture du système Soar ([NEW 80b], [LAI 86], [LAI 87]) suppose-t-elle une structuration bien précise au niveau des symboles. Elle est basée sur l'idée qu'il faut remplacer la notion d'un espace de recherche par celle de plusieurs *espaces*

de problème. La tâche du système est conçue comme une hiérarchie de buts et de méthodes. Chaque méthode est décrite par un espace de problème, lequel est défini par un état initial, un état désiré et des connaissances sous forme d'un ensemble d'opérateurs de transition (en ce sens, c'est un espace de recherche). L'originalité de l'architecture réside dans la manière dont la coordination entre ces différents espaces est assurée par une structure de contrôle universelle, mettant en œuvre un principe de moindre choix. Contrairement aux autres systèmes, Soar n'a pas de phase de résolution de conflit, phase qui conduit ceux-là à faire des choix arbitraires dans les cas où plusieurs règles sont activables simultanément. Quand une impasse se produit dans un espace de problème, c'est-à-dire quand les connaissances disponibles dans cet espace sont insuffisantes pour choisir univoquement l'action suivante, ou quand aucune action n'est possible, un sous-but est automatiquement généré, et un autre espace de problème est sélectionné, en fonction du type d'impasse et des espaces de problème disponibles (il existe un certain nombre d'espaces généraux, qui définissent des actions par défaut au cas où il n'y a pas de connaissances spécifiques). Ce principe de fonctionnement (universal subgoaling) est suffisamment général pour produire, selon les connaissances disponibles, tous les algorithmes d'exploration d'un graphe constituant les "méthodes faibles" de recherche.

Ultérieurement, Newell sera conduit à promouvoir l'architecture de Soar en un "modèle computationnel des espaces de problèmes", positionné *entre* le niveau des savoirs et le niveau des symboles. La signification profonde de cette évolution, liée à la vision du système comme agent, sera analysée à la section 6.

4.6 L'âge des méthodologies

De la nécessité de structurer les connaissances, il n'y avait qu'un pas à franchir pour reconnaître la nécessité de structurer aussi la démarche de développement. L'émergence de préoccupations méthodologiques, puis l'élaboration des premières méthodes de développement inspirées du génie logiciel, dans les années 80, ont conduit les praticiens des systèmes à base de connaissances à reconnaître que s'imposait une conception nouvelle de l'objet d'un tel système□ aux prétentions initiales aussi floues qu'exorbitantes, s'est substituée l'idée que *l'objet d'un système expert est la modélisation d'un comportement observable en vue de la résolution d'un certain type de problèmes bien déterminés dans un certain domaine, dans des contextes eux aussi bien déterminés. Ainsi, les connaissances nécessaires doivent-elles être spécifiées, non dans l'absolu, dans une perspective de modélisation cognitive d'un expert humain ou d'une partie de ses activités, mais par rapport aux objectifs précis du système en cours de développement*. Ces idées sont l'extrapolation concrète de la notion de niveau des savoirs et de la vision du système comme agent introduites par Newell une dizaine d'années plus tôt. Elles sont aussi la conséquence logique de l'idée même de modélisation, inséparable de celle des objectifs assignés aux modèles recherchés.

Plusieurs méthodologies ont été élaborées sur ces bases. Nous prendrons ici l'exemple de KADS, devenue une espèce de standard européen de fait. L'exemple particulier choisi a d'ailleurs peu d'importance pour notre propos, qui est d'illustrer une tendance générale, et d'analyser les modèles apparaissant au niveau des savoirs.

Le premier principe fondamental de KADS est que le développement d'un système à base de connaissances est essentiellement une activité de modélisation et qu'il doit être guidé par la construction d'un certain nombre de *modèles*. Chaque modèle exprime certaines caractéristiques du système à développer, spécifiques à un aspect de ce système, et faisant abstraction d'autres aspects; d'une façon très cartésienne, le problème est ainsi divisé en sous-problèmes plus simples. Dans notre terminologie, les modèles au sens de KADS sont des modèles fonctionnels de leur objet.

Les modèles prescrits par KADS sont représentés dans la figure 1, issue de [SCH 93], et constituent une expression détaillée du second principe fondamental de KADS□dans la pure orthodoxie du génie logiciel, le développement d'un système à base de connaissances doit être guidé prioritairement par les besoins de l'organisation dans laquelle le système devra fonctionner et par la structure des tâches à réaliser, puis par les modèles de connaissance à mettre en jeu, avant toute référence à un mode particulier de représentation ou d'implémentation. A ce schéma, nous avons par anticipation superposé une répartition en trois niveaux□ de l'organisation, des savoirs, des symboles, sur lesquels nous reviendrons ultérieurement.

Nous donnons une description sommaire des différents modèles, d'après [SCH 93], limitée aux aspects nécessaires pour la suite.

Au niveau de l'organisation, les trois modèles (de l'organisation, de l'application, des tâches) définissent les buts de la construction du système□

- le modèle de l'organisation (organizational model) décrit l'environnement socio-organisationnel dans lequel le système devra fonctionner,
- le modèle de l'application (application model) définit le problème que le système est censé résoudre dans l'organisation, la fonction du système dans cette organisation, et les contraintes externes imposées au système,
- le modèle des tâches (task model) décrit comment la fonction du système, définie dans le modèle de l'application, sera réalisée grâce à un découpage en un certain nombre de tâches à effectuer, et comment ces tâches seront réparties entre le système et les différents acteurs de l'organisation (humains ou autres systèmes). Il comporte donc un choix parmi les différentes manières possibles de réaliser la fonction du système, ainsi qu'un choix d'organisation du travail.

Au niveau des savoirs, les trois modèles suivants (de coopération, d'expertise et conceptuel) concernent la modélisation des connaissances dont le système devra disposer pour réaliser les tâches définies dans le modèle des tâches. Ces modèles se situent sur le plan du comportement observable du système dans l'accomplissement de ses tâches, et ne s'intéressent pas aux processus internes qui peuvent produire ce comportement□

- le modèle de coopération (model of cooperation) définit les tâches qui requièrent une coopération entre le système et les autres acteurs,
- le modèle d'expertise (model of expertise) est la partie la plus spécifique à un système à base de connaissances. C'est une spécification fonctionnelle, sous forme d'une catégorisation exhaustive, des connaissances nécessaires à la résolution des tâches qui sont assignées au système,
- le modèle conceptuel (conceptual model) est la réunion des modèles d'expertise et de coopération.

Au niveau des symboles, le modèle de conception (design model) spécifie les choix de conception□ méthodes de représentation et de manipulation des connaissances, que le système doit utiliser pour implémenter le modèle conceptuel, dans un paradigme particulier de représentation. Il ne suppose pas le choix d'un outil particulier d'implémentation de ce paradigme.

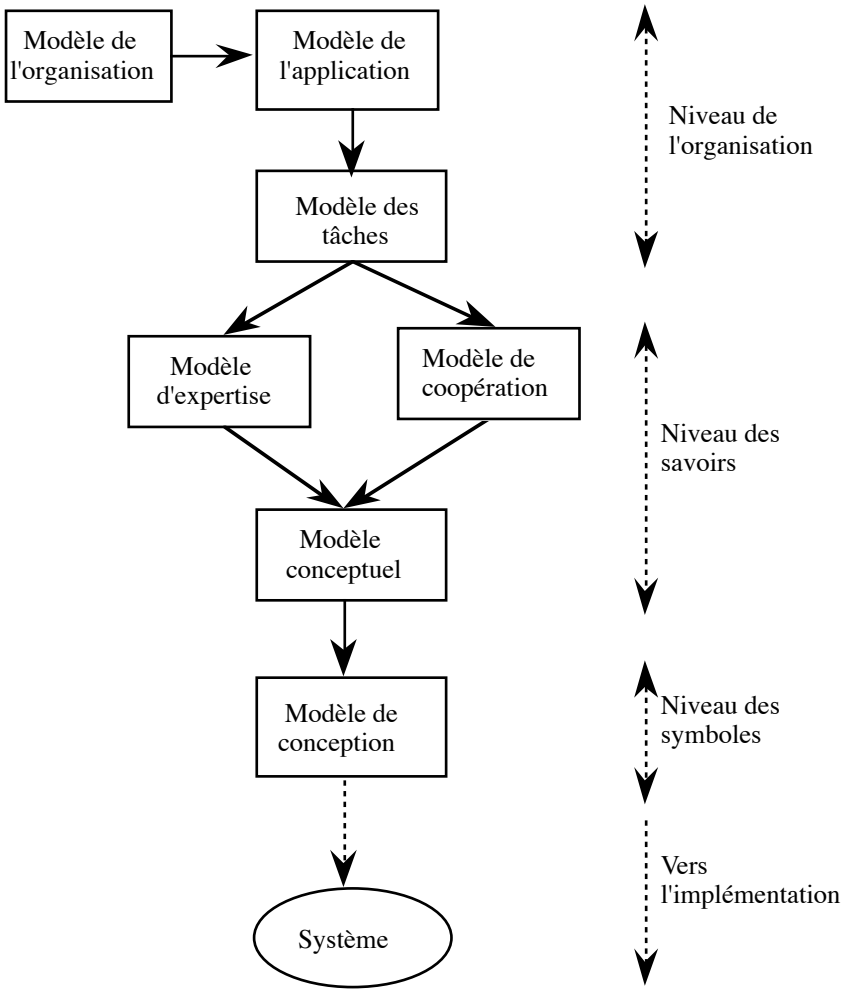

Figure 1: les modèles dans KADS, répartis en trois niveaux

5. L'objet de l'intelligence artificielle

5.1 Retour sur les niveaux de description d'un système intelligent

Le découpage en trois niveaux que nous avons superposé à KADS donne un éclairage nouveau sur la hiérarchie des modèles de KADS. Il nous permet en effet d'établir un lien avec des questions classiques de gestion des ressources humaines□
- *au niveau de l'organisation*, les trois modèles de KADS constituent *une description de poste* du système, considéré comme agent dans l'organisation,
- *au niveau des savoirs*, les trois modèles suivants constituent *une description des compétences nécessaires pour tenir ce poste*,
- au niveau des symboles, le modèle de conception constitue la description de la manière dont un artefact qui tiendrait ce poste peut être conçu.

Le second principe de KADS, dont nous pensons exprimer ainsi l'essence, peut être paraphrasé d'une manière assez triviale□ avant d'"embaucher" un artefact intelligent, il est préférable de décrire son poste, ainsi que les compétences nécessaires pour le tenir.

La fidélité à l'esprit de la méthododologie KADS nous semble garantie. C'est pour nous l'apport implicite essentiel de KADS que d'avoir traduit une démarche classique de génie logiciel en une série de modèles interprétables dans ces termes - même si l'agent de KADS peut nous paraître maintenant quelque peu étriqué.

La fidélité de cette formulation, donc de KADS, à la définition que donne Newell du niveau des savoirs, est plus discutable□d'un côté, le modèle que Newell place au niveau des savoirs est bien un modèle fonctionnel; d'un autre côté, rien chez Newell ne vient contraindre la notion d'agent rationnel, sinon un principe de rationalité passablement éthéré - en particulier, rien n'apparaît de la nécessité d'interagir avec l'environnement. Aussi est-il clair que nous avons, à la suite de KADS, complété la notion d'agent que Newell place à ce niveau, transformant une vision purement mentaliste du système comme agent en la vision à dimension plus sociale d'un agent en situation dans une organisation. Mais *lui donner une dimension sociale n'est-il pas la seule manière de prendre au sérieux la vision du système comme agent*?

5.2 L'agent rationnel abstrait, objet de l'intelligence artificielle

Ce qui ressort ainsi de notre étude, c'est que *la notion centrale qui permet de mettre en évidence une convergence des évolutions analysées, et qu'on peut donc considérer comme l'objet de l'IA, n'est plus ni l'intelligence, ni la cognition, ni la connaissance, mais c'est la notion d'un agent rationnel abstrait, en situation dans une organisation, et en interaction avec d'autres agents (humains ou non) - et, c'est clairement sous-entendu, mais nous ne devons laisser aucune ambiguïté là-dessus, physiquement implémentable*.

Conviendrait-il alors de transposer nos réflexions dans le cadre d'un paradigme général de conception et programmation par agents (qui aurait en outre l'avantage pratique de rassembler les trois aspects classiques de la programmation□ données, procédures, contrôle - là où l'approche objet ne réunit que les deux premiers)? L'intelligence artificielle distribuée (IAD) s'engage dans cette voie, à travers l'étude des systèmes multi-agents, en mettant l'accent sur la communication et la

coopération entre agents, et donc sur les aspects sociaux des agents. Mais l'IA ne saurait se réduire purement et simplement à un paradigme général de conception par agents, tombant ainsi du Charybde du cognitivisme au Scylla de l'interactionnisme□ sa spécificité irréductible reste dans le traitement explicite par ces agents de "concepts formels", et ce d'une manière fondamentalement utile du point de vue de leur description conceptuelle. (Les exemples où c'est effectivement utile abondent. Quand ce n'est pas utile, cela signifie qu'il n'y a pas lieu de recourir à des méthodes d'IA.)

Un agent rationnel abstrait est défini par ses "concepts formels" et ses lois de comportement et d'interaction. Par "concepts formels" nous entendons des formalisations de concepts mentaux généraux tels que buts, connaissances, croyances, hypothèses, etc, et de concepts spécifiques à un domaine, dont la sémantique est définie par un modèle conceptuel. Les lois de comportement définissent ses comportements spontanés observables (par exemple les déductions légitimes) et les lois d'interaction définissent les principes de son dialogue avec d'autres agents (par exemple la manière de traiter des requêtes d'information, ou de négocier certaines collaborations).

Selon cette perspective, les problèmes centraux de l'IA concerneraient désormais□
- du côté de la théorie□ la définition précise de certains concepts mentaux formels généraux; la multiplication des publications visant à la modélisation logique dans les revues ou congrès d'IA nous semble révélatrice d'une tendance dans ce sens,
- du côté des méthodes et des applications□ les principes de spécification du comportement observable attendu d'un agent (la description détaillée de son poste), de formalisation des savoirs de l'organisation nécessaires à l'obtention de ce comportement (la description de ses compétences), et d'opérationnalisation de ces savoirs afin d'obtenir effectivement le comportement souhaité; là, c'est évidemment l'épanouissement des travaux sur la méthodologie qui est significatif,
- du côté des outils informatiques généraux□ les systèmes de représentation et manipulation des connaissances, et , en amont, les outils qui correspondent à des paradigmes généraux d'opérationnalisation des savoirs; dans cette catégorie, on peut citer $(ML)^2$, Model-K, OMOS, etc ([Sch 93], [For 93]).

5.3 Problèmes de chevauchement

Dans cette section, nous formulons en termes d'agent rationnel abstrait deux problèmes plus ou moins connus.

5.3.1 Aspects mentaux et aspects sociaux des agents rationnels abstraits

Peut-on distinguer deux types de tâches, et séparer aussi radicalement que le fait KADS le modèle de coopération du modèle d'expertise - c'est-à-dire les aspects "sociaux" de l'agent de ses aspects "mentaux" (question posée au niveau des modèles fonctionnels)? Plus grande est la nécessité d'interactivité du système avec d'autres agents, plus grande sera la nécessité de pourvoir le système d'états mentaux formels complexes, et plus cette séparation risque d'être difficile.

C'est là qu'apparaît une limitation fondamentale de KADS, comme de toute méthodologie existante de génie logiciel; cette limitation correspond au caractère artisanal que garde la conception. Certaines exigences, comme la capacité à expliquer

ses raisonnements, ou à apprendre par expérience, sont de nature générale, et seraient très difficilement spécifiables au niveau du modèle conceptuel□ il faudrait pour cela un découpage trop fin des tâches de coopération. Elles apparaissent donc comme contraintes au moment de la conception, contraintes pour lesquelles la notion de conception préservant la structure (KADS) n'est d'aucun secours. (Rajouter le modèle intermédiaire des espaces de problème, analysé plus loin, n'apporte rien non plus en matière d'explication). La grande difficulté de ces contraintes est d'abord liée à leur chevauchement des aspects mentaux et sociaux. Une approche de l'interaction entre ces aspects, est proposée dans [SHO 93], où Shoham définit un langage de programmation combinant des primitives qui correspondent à des états mentaux formels, et des primitives de communication entre agents, qui reposent sur la théorie des actes de parole de Searle ([SEA 68]).

5.3.2 Chevauchement des niveaux des savoirs et des symboles

Une deuxième difficulté apparaît dans les deux exemples ci-dessus, liée à un autre chevauchement, à notre avis beaucoup plus grave conceptuellement□ celui des niveaux des savoirs et des symboles. Soar, par exemple, réalise bien de l'apprentissage au niveau des symboles, mais ne produit rien au niveau des savoirs - ce qui rend les règles apprises incompréhensibles. Du côté de l'explication, la nécessité d'introduire dans le système de diagnostic médical MYCIN des connaissances "de support" pour en faire le système d'enseignement GUIDON ([CLA 87]), ou encore l'approche de Swartout de construction de systèmes experts capables d'expliquer leur raisonnement ([SWA 83]) peuvent être compris comme un besoin d'accès du système au niveau des savoirs.

Devons-nous y voir l'indication qu'il faudra un jour dépasser la vision purement inférentielle d'une base de connaissances pour lui adjoindre d'autres principes fondamentaux de structuration, conduisant par exemple à superposer à l'actuelle strate "inférentielle" une strate "navigationnelle" contenant les savoirs sous une forme humainement communicable, et reflétant leur structure hypertextuelle (cf 5.4.2)? La difficulté de définir les principes de coopération de deux structures aussi hétérogènes est évidemment considérable. Néanmoins, [SWA 83] et [CLA 87] constituent de petits pas dans cette direction.

Une telle extension aurait des répercussions intéressantes sur la notion d'agent rationnel abstrait□ au lieu d'une "simple" opérationnalisation des savoirs par un programmeur externe, elle obligerait à *modéliser explicitement certains rapports de l'agent au savoir*. La portée d'une telle extension peut se mesurer à son impact sur la méthodologie, où elle correspondrait à la nécessité de formaliser, même partiellement, la démarche d'opérationnalisation des savoirs, et ce de manière à conserver dans la strate inférentielle des références exploitables vers la strate navigationnelle, allant ainsi au-delà de langages ou systèmes orientés vers l'opérationnalisation, comme $(ML)^2$, Model-K, OMOS. La difficulté peut, elle, se mesurer en considérant qu'un rapport de l'agent à son savoir signifie pour cet agent une certaine forme de réflexivité ou de conscience de soi - aspect qui reste aujourd'hui largement hors de portée pratique des approches formalistes-symboliques.

5.4 Évolution de la conception du savoir

De même que nous avons déduit de l'évolution de l'IA ce qu'est implicitement devenu son objet, nous pouvons tenter de déduire comment s'est modifiée la notion (jamais clairement explicitée) de connaissance au cours de la transformation radicale qu'ont subies les pratiques, passant de l'*extraction et codage* des connaissances à la *modélisation constructive* de domaines de savoir.

5.4.1 L'organisation des savoirs et les savoirs dans l'organisation

Nous avons vu que, pour KADS, la structuration des savoirs est essentiellement contrainte par l'organisation dans laquelle le système devra fonctionner et la répartition des tâches entre le système et les autres acteurs de l'organisation - ce qui se traduit par l'apparition d'un niveau de l'organisation au-dessus du niveau des savoirs. (KADS se démarque en cela de Newell, pour qui les principes de structuration du niveau des savoirs semblent être de nature purement cognitive).

Il en résulte clairement qu'il ne saurait y avoir, pour KADS, de vérité universelle de l'organisation des savoirs, mais uniquement des vérités locales de l'articulation des savoirs dans les organisations. Le fondement de cette articulation est chaque fois la répartition des tâches, par laquelle chaque organisation définit une perspective sur les savoirs et un découpage en fonction de ses propres objectifs. (Il y a d'ailleurs une certaine ironie dans la constatation qu'admettre ces vérités locales, qui donnent à chaque organisation un immense pouvoir sur ses agents, s'oppose en retour à la tentation du pouvoir absolu que donnerait une vérité universelle).

Il n'est pas indifférent, culturellement, de remarquer que la conception du savoir qui résulte implicitement de cette évolution est en opposition radicale⊡
- d'une part avec une vision mentaliste, naturaliste, de la connaissance, qu'il suffirait d'aller extraire de l'expert, (vision très largement dépassée aujourd'hui),
- d'autre part, et plus significativement, avec la vision hégélienne d'un savoir absolu naturellement articulé en domaines, vision qui conserve une place importante dans les sciences, sous la forme d'un arbre généalogique des savoirs.

Une conséquence des différents modèles posés par KADS est que *l'articulation des savoirs passe nécessairement par les médias de communication* (langage, représentations graphiques,..., tout type de système conventionnel de communication). *Le savoir relève donc plus du champ socioculturel que du domaine mental*. Bien sûr, les médias de communication s'appuient en définitive sur des structures cognitives innées; mais ils ne sauraient s'y réduire, étant essentiellement des constructions socioculturelles.

De manière tout-à-fait fondamentale, *le savoir est formulé (et non pas mis en formules) dans les modes socialisés de communication*. Aussi, plutôt que de positionner le savoir comme une entité abstraite, inanalysable, l'institution d'un niveau des savoirs encourage à *identifier le savoir au réseau de ses formulations effectives*. Ce pas est loin d'être anodin en pratique, puisqu'il ouvre la voie à l'étude du savoir en tant que tel par les méthodes structuralistes d'analyse de corpus.

5.4.2 Aspects modulaires et hypertextuels du savoir

Que l'articulation des savoirs soit dans chaque cas contrainte par l'organisation concernée ne s'oppose pas à l'énonciation de *principes génériques de structuration locale*. Nous en formulerons deux, qui se déduisent aussi des pratiques constatées.

5.4.2.1 *Le savoir est fondamentalement modulaire*. Et la "taille" des modules n'est pas celle d'un "atome de connaissance", d'une règle individuelle, mais plutôt celle d'une micro-théorie, d'une "source de connaissances", d'un "espace de problème". *Bien que fondamentale, la modularité du savoir n'a pas d'expression privilégiée naturelle*□ nous avons vu que tout découpage des savoirs doit être subordonné à des principes et objectifs extérieurs aux savoirs eux-mêmes, définis au niveau de l'organisation.

5.4.2.2 *Le savoir, tel qu'il apparaît dans le paradigme du système de symboles concrets, est fondamentalement de nature hypertextuelle*. Nous entendons par là□
- que le savoir forme un réseau d'associations, c'est-à-dire de liens orientés et typés entre noeuds; nous ne préjugeons pas du contenu de ces nœuds, qui peuvent être de nature textuelle, sonore, visuelle (statique ou dynamique, schématique ou picturale); ce premier point est donc assez trivial,
- qu'il existe une sorte d'étiquetage de chaque noeud non textuel par un noeud textuel□dans la mesure où, dans le paradigme du système de symboles concrets, le savoir est formulé selon des modes socialisés de communication, il est difficile d'imaginer du savoir qui échapperait complètement au domaine du verbe. (C'est peut-être une limitation de ce paradigme vis-à-vis du savoir-faire. Néanmoins, cette idée d'une association entre éléments textuels et éléments non textuels apparaît aussi dans la sémantique conceptuelle de Jackendoff ([JAC 83]), comme nécessaire à la construction des structures sémantiques de certains concepts). C'est pourquoi nous employons le terme "hypertextuel" (et pas seulement "en réseau"), dont il convient à notre avis de dégager un sens structurel indépendant des techniques informatiques qui lui ont donné vie. (Le terme "textuel" serait inadéquat, la linéarité du texte n'étant pas nécessaire à la formulation du savoir).

Par ailleurs, les noeuds sont typés, et le nombre de types est a priori fini, ou tout au moins finiment engendré (c'est-à-dire à partir d'un nombre fini de types de base, à l'aide d'un nombre fini de règles de création). Cette condition de finitude constitue peut-être la caractéristique essentielle du savoir□ *le savoir est formulé selon des modes répertoriés, associés à des modes de communication*. Non qu'il faille considérer que les types sont immuables□ à des types liés aux structures cognitives innées s'ajoutent des types liés à la culture, le livre en étant l'exemple canonique.

Un livre (pas au sens de l'objet matériel, mais d'un contenu pouvant présenter diverses instances physiques) constitue un noeud du réseau du savoir. Il possède un certain nombre de types de liens vers d'autres noeuds, qu'on trouve dans la bibliographie, dans les mentions faites à d'autres livres, ou d'autres types de noeuds (films, etc). Il est aussi la cible de certains liens□ je sais, par exemple, que je trouverai telle information, que j'ai oubliée, dans ce livre-là. Mais ce noeud du savoir qu'est un livre possède lui-même une structure interne hypertextuelle extrêmement riche, basée sur un certain nombre de types de liens tout-à-fait explicites: table des matières, pagination, glossaire, index, graphes de dépendances entre chapitres ou sections, notes de bas de page, annotations, colophons, renvois à d'autres chapitres ou pages, figures, tableaux, illustrations, résumés, appendices, exercices, rappels, annexes, etc. Cet exemple donne une idée de la complexité de l'organisation du

savoir□ ce qui apparaît ici comme noeud peut apparaître ailleurs comme un sous-réseau fortement structuré. Encore n'avons-nous pas pris en compte les noeuds implicites de nature sémantique, beaucoup plus difficiles à énumérer□ liaisons de paraphrasage, de dépendance conceptuelle, de subsomption, d'illustration, etc.

6. Capitaliser, opérationnaliser, interpréter

Nous analysons dans cette section les trois grands moments du traitement du savoir en intelligence artificielle.

6.1 La capitalisation du savoir

6.1.1 Formes de la capitalisation

Admettant que l'organisation des savoirs soit de nature hypertextuelle, sommes-nous pour autant justifiés à identifier la capitalisation du savoir avec la construction d'un hypertexte, au sens informatique du terme, comme certains environnements de développement de systèmes experts le suggèrent? Ce raccourci cadrerait mal avec la réalité historique, qui fait apparaître le livre comme support privilégié de la capitalisation. Aussi avons-nous pris soin de distinguer pour la notion d'hypertexte un sens conceptuel d'organisation, indépendant des outils informatiques de même nom.

Par ailleurs, on ne peut parler de capitalisation sans évoquer la question de l'*accès* au capital de savoir. De ce point de vue, la technologie informatique hypertexte peut constituer un apport significatif. Si l'on considère qu'une bonne capitalisation du savoir doit permettre de le faire fructifier au maximum, et donc en faciliter l'accès, la forme hypertextuelle présente des avantages - si ce type d'outil devient d'usage courant. Mais la capitalisation du savoir ne saurait se réduire à l'utilisation de telle ou telle technique.

6.1.2 Vers un niveau sociétal?

Nous l'avons vu, *la construction d'un système expert passe par la socialisation préalable des savoirs nécessaires à son fonctionnement*. Aussi, l'intelligence artificielle constitue-t-elle une formidable entreprise de production et capitalisation de savoir, qui touche à tous les domaines de l'activité humaine. Très nombreux sont d'ailleurs les témoignages indiquant que l'un des principaux résultats de la construction d'un système expert a été une meilleure compréhension du domaine et une formalisation de ses concepts et méthodes. Cette entreprise met en jeu des volumes impressionnants de savoir. Après l'informatisation des entreprises (c'est-à-dire de leur gestion et de leurs chaînes de production), le prochain défi de la compétition pourrait bien être l'axiomatisation et l'informatisation de leurs savoirs - tâche dont chacun appréciera l'impact culturel et social.

Dans ce contexte, normalisation, interopérabilité, réutilisabilité, évolutivité, etc, deviennent des problèmes clés. Une organisation se situe dans un contexte sociétal global (de plus en plus mondial). Une méthodologie prenant explicitement en compte les questions ci-dessus se devrait donc de rajouter un niveau sociétal au-

dessus du niveau de l'organisation. A ce niveau, nous situons les diverses normes régissant le domaine de l'application, et, de manière plus hypothétique, certaines ontologies globales relatives à des connaissances non spécifiques au domaine - connaissances de sens commun, par exemple, qu'il peut être fastidieux de spécifier chaque fois au niveau de l'organisation. Nous plaçons à ce niveau le projet CYC de Lenat ([GUH 89], [LEN 90]).

Nous situons également à ce niveau les recherches en sémantique conceptuelle, champ récent de la linguistique, issu de travaux de Gruber et Talmy ([GRU 65], [TAL 78], [TAL 80], [TAL 83], [TAL 88]). Gruber avait émis l'hypothèse que la sémantique des langues naturelles peut s'analyser selon divers champs sémantiques (spatial, temporel, possession, existence, etc), et que les éléments de base constituant la structure du champ spatial suffisent à décrire métaphoriquement les autres champs. Les analyses linguistiques minutieuses, lexicales et grammaticales, de Jackendoff, Levin, Pinker, Pustejowski, etc ([JAC 83], [JAC 87], [PIN 89], [JAC 90], [LEV 91]) tendent à confirmer cette hypothèse et à lui donner des expressions précises, sous forme, pour la première fois, de principes pour construire une représentation informatisable de la sémantique des concepts. Ces représentations multi-linguales ont prouvé leur intérêt, étant à la base d'un système d'aide à la traduction ([DOR 93]). Aussi peut-on espérer qu'elles contribueront à réduire l'arbitraire de la représentation qui caractérisait les premiers réseaux sémantiques.

Situant ces travaux à ce niveau sociétal, c'est-à-dire au niveau où peut se mesurer leur efficacité, nous ne prenons pas parti en ce qui concerne leurs interprétations philosophiques⊡ invariants cognitifs innés sous-jacents ayant des relents kantiens ([PIN 89], p. 372), ou expressions de l'universalité mathématique des catastrophes élémentaires ([PET 92], p. 67 et 293).

6.2 L'opérationnalisation des savoirs

Capitaliser des savoirs n'implique pas qu'on puisse les utiliser tels quels. Les savoirs n'étant (théoriquement) pas liés à un mode particulier d'utilisation, toute utilisation est a priori légitime et virtuellement possible; *l'opérationnalisation consiste à* transformer cette virtualité en comportements réels, c'est-à-dire, plus précisément, à *produire une représentation opératoire d'un modèle fonctionnel.*

Les savoirs, comme des partitions musicales, demandent à être interprétés. Chez l'homme, l'appareil cognitif est le seul interprète possible des savoirs, ce qui le rend subjectivement universel pour chacun, par défaut de concurrence. En intelligence artificielle, *pour transformer la spécification abstraite au niveau des savoirs en un système implémentable, l'ingénieur de la connaissance doit définir son interprétation. Il lui appartient de déterminer un paradigme d'interprétation informatique des savoirs assez puissant pour mener un ensemble d'inférences suffisant par rapport aux besoins spécifiés pour le système qu'il développe*.

6.2.1 L'opérationnalisation selon Newell et KADS

Les approches de KADS et de Newell concernant l'opérationnalisation des savoirs sont apparemment compatibles⊡
- KADS recommande une démarche de conception préservant la structure⊡ le modèle de conception (niveau des symboles) doit être un affinement du modèle conceptuel;

on entend par là qu'un processus d'affinement agit sur le modèle conceptuel; mais, tenant compte de l'existence de boucles de rétro-action dans le cycle de vie, il faut se garder d'en déduire qu'il constituerait l'unique processus de transformation appliqué à partir d'un modèle conceptuel figé; au sens strict, c'est seulement à la fin du développement, que le modèle de conception doit être, structuralement, un affinement du modèle conceptuel; *l'opérationnalisation ne concerne donc pas que le niveau des symboles, elle agit de manière cohérente sur l'ensemble des niveaux*,
- l'école de Newell introduit le modèle computationnel des espaces de problèmes, qui représente une manière particulière de préserver la structure.

On peut trouver une complémentarité formelle entre ce dernier modèle et la hiérarchie des niveaux de KADS□ il vient remplir un vide entre le modèle conceptuel et le modèle de conception; à l'inverse, les modèles de KADS que nous avons regroupés dans les niveaux de l'organisation et des savoirs, et qui sont des modèles fonctionnels, peuvent constituer un étage amont par rapport au modèle de fonctionnement des espaces de problèmes. Mesurer la portée pratique de cette complémentarité demanderait toutefois une expérimentation prolongée, qui fait défaut aujourd'hui, les approches les plus avancées de l'opérationnalisation, comme $(ML)^2$, Model-K, OMOS, etc, ne passant pas par ce modèle des espaces de problème.

La cohérence globale de chacune des deux approches repose sur l'hypothèse, informulée, qu'on peut structurer de manière compatible les niveaux des savoirs et des symboles. Pour KADS, cela se traduit par le postulat d'un modèle fonctionnel général (la spécification des connaissances, en quatre strates, se fait en gros dans le modèle conceptuel de KL-ONE, plus de la logique temporelle), susceptible d'être affiné dans chaque application en un modèle de fonctionnement. Dans la mesure où KADS reste muette sur les principes concrets de passage de la spécification à la conception (des modèles des tâches et conceptuel au modèle de conception), ces tautologies ne vont guère plus loin que leurs homologues en génie logiciel. Pour Newell, la compatibilité supposée entre les niveaux se traduit par le postulat d'un modèle général de fonctionnement, le modèle des espaces de problème, supposé être suffisant pour la reformulation des modèles fonctionnels de chaque application. Dans la mesure où ce modèle est une machine universelle abstraite, et puisqu'un agent rationnel abstrait est physiquement implémentable, son modèle fonctionnel est théoriquement traduisible dans ce modèle; avec quelle facilité est une autre question.

6.2.2 Vers un modèle paradigmatique de fonctionnement d'agent rationnel abstrait?

Ainsi, la promotion de l'architecture de Soar en un modèle computationnel des espaces de problème ([NEW 91]) peut être comprise comme le choix d'un modèle paradigmatique abstrait de fonctionnement d'une machine, dans lequel il est théoriquement possible de représenter et d'exécuter les modèles fonctionnels d'agents rationnels définis au niveau des savoirs. (Que ce modèle constitue aussi pour Newell une théorie de l'architecture cognitive humaine est une considération que nous excluons de notre propos). Dans cette interprétation, on est en toute rigueur au sommet du niveau des symboles, mais pas au-dessus - et la proposition n'est qu'une variante de la thèse de Newell.

Placer ce modèle abstrait à un niveau intermédiaire entre les niveaux des savoirs et des symboles en fait une thèse beaucoup plus ambitieuse□ le modèle computationnel des espaces de problème se veut un modèle de fonctionnement

universel, physiquement implémentable, d'agent rationnel abstrait. Notons bien que cette thèse est orthogonale aux affirmations concernant son interprétation cognitiviste. Notons aussi que ce modèle général ne suppose pas que l'ensemble des connaissances de l'agent soit clos par déduction logique (ce qui en ferait un modèle très restrictif), il ne fait strictement aucune hypothèse sur les inférences susceptibles d'être effectuées. Les seules propriétés universelles de l'agent sont les suivantes: quand il a un but, il essaie de l'atteindre en utilisant les connaissances dont il dispose, ce qui le conduit à créer des sous-buts, et quand il n'arrive pas à atteindre un sous-but dans un certain contexte, il passe dans un autre contexte approprié à ce sous-but. C'est, si l'on veut, très rationnel, cartésien, idéaliste; mais l'objectif n'est pas la modélisation cognitive. Une question reste ouverte□ ce modèle, qui est certes universel au sens de la thèse de Church, est-il si universel (au sens courant du terme) qu'il le prétend en tant que modèle de fonctionnement d'un agent?

6.3 Interprétation de la base de connaissances

Nous avons vu que la construction d'une base de connaissances se fait par capitalisation, puis opérationnalisation, des savoirs. Le niveau des savoirs apparaît ainsi comme médiateur entre des connaissances informulées et leurs représentations. Mais la médiation se fait aussi en sens inverse, depuis les chaînes de symboles a-signifiantes reposant dans la base de connaissances et celles qui sont engendrées par inférence, vers les connaissances, lors de l'interprétation par un humain du comportement du système (le mot "interprétation" est pris dans cette section dans son acception habituelle, contrairement à la section précédente). En effet, si l'on s'accorde à donner un sens aux actions du système, ce ne peut être que par référence à une théorie de son comportement et non par une compréhension directe, *im*médiate de ses manipulations de concepts mentaux formels (même si une interprétation "intuitive" de ceux-ci peut faciliter une compréhension plus rigoureuse). Ce modèle du comportement auquel on se réfère, c'est bien sûr le modèle fonctionnel du système en tant qu'agent rationnel abstrait, spécifié au niveau des savoirs.

L'accord auquel s'appliquaient ces remarques est un accord scientifique entre spécialistes sur la compréhension du fonctionnement du système, un accord qui peut servir de base à la validation de son comportement. Il existe un autre aspect de l'interprétation, celle que peut faire un utilisateur non averti, qui conduit à un autre type d'accord, basé sur un penchant partagé à l'anthropocentrisme, consistant à projeter sur le système son propre modèle de fonctionnement, et revenant à prêter à un artefact des processus cognitifs internes qu'il n'a pas nécessairement, mais qui sont compatibles avec son comportement externe. Nous ne pouvons pas aborder ici l'analyse des bases de ce penchant, forme particulière d'une activité inconsciente, générale et permanente, de remplissage perceptif et cognitif, par laquelle nous construisons notre monde. Remarquons seulement qu'il consiste techniquement à projeter sur un modèle fonctionnel un modèle de fonctionnement observationnellement équivalent.

Plus grande sera la nécessité d'interactivité du système, plus forte aussi sera probablement la tendance à faire de l'exploitation de ce penchant anthropocentrique l'un des objectifs de l'organisation, pour des raisons évidentes d'efficacité. Mettre à profit ce penchant pour concevoir des systèmes qui non seulement répondent aux

objectifs de l'organisation, mais donnent en outre l'illusion d'être réellement intelligents, relève de l'ergonomie cognitive.

6.4 Organisation, savoir, représentation

La figure 2 ci-dessous situe les uns par rapport aux autres les concepts analysés dans cet article. On remarquera que nous situons les connaissances individuelles sur le même plan que la base de connaissances, au niveau de savoirs opérationnalisés. La mention d'un "niveau mental / niveau des symboles" est cependant destinée à écarter tout présupposé sur la nature symbolique de la cognition humaine.

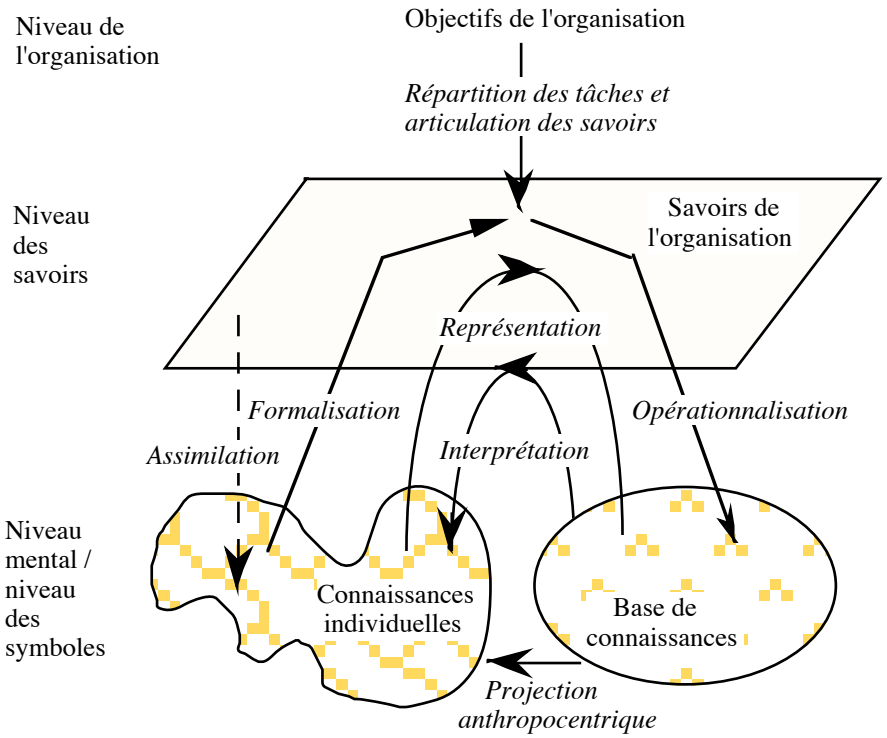

Figure 2: Relations entre les trois niveaux

7. Conclusion

Guidés par une distinction générale entre un modèle et ses représentations, nous avons survolé l'évolution de l'intelligence artificielle qui s'est produite à l'intérieur du paradigme du système de symboles concrets, sous la poussée combinée des développements conceptuels, applicatifs et méthodologiques. Nous en avons tiré des conséquences de natures diverses⊡

- l'objet de l'IA nous apparaît désormais comme la modélisation d'agents rationnels abstraits en situation dans des organisations, avec leurs aspects "mentaux" et "sociaux"; mais l'agent rationnel abstrait ne doit pas être conçu comme un concept figé□l'activité de modélisation est en IA constitutive de son objet,
- la notion de savoir qui joue un rôle concret en IA est celle d'un savoir formulé dans le cadre d'une organisation sociale et subordonné à l'accomplissement de tâches précises (et non celle de connaissances idéales logées dans la tête d'experts); cela nous a conduit à identifier le savoir au réseau concret de ses formulations effectives, et à y déceler une structure hypertextuelle,
- le centre de gravité de l'IA s'en trouve décalé du champ mental vers le champ socioculturel; ou encore, l'IA , pour autant que ses productions soient utilisées comme modèles métaphoriques de la cognition, promeut désormais une vision de la cognition et de la connaissance qui déborde du niveau mentaliste individuel, auquel elle se cantonnait initialement, pour s'étendre au niveau socioculturel, où s'élaborent les modèles d'interprétation du monde.

Nous avons constaté que l'IA en est venue tardivement à se reconnaître comme une activité de modélisation, renonçant ainsi au mirage d'un accès direct à des représentations internes qui constitueraient des espèces de modèles naturels du réel. Quoiqu'apparaissant aujourd'hui comme très naïves, ces prétentions initiales étaient selon nous une conséquence naturelle, à défaut de légitime, d'un a priori philosophique représentationnel inanalysé, et projeté trop brutalement sur un niveau technique non conforme à sa généralité. Mais ce qui nous semble légitime, au-delà de cette naïveté des débuts, c'est l'intention proprement scientifique de décrire les choses "telles qu'elles sont", objectivement, ou encore le rejet de l'arbitraire et du subjectif.

Or, de ce point de vue, la ligne d'évolution de l'IA décrite ici pourra sembler profondément frustrante si elle est comprise comme la promotion du niveau des savoirs et de la spécification fonctionnelle abstraite d'agents rationnels au détriment du niveau des symboles. Nous avons vu en effet que la définition du poste et des compétences de l'agent était totalement subordonnée aux objectifs spécifiques de l'organisation; l'objectivité à ce niveau ne peut donc être que celle de la bonne modélisation de l'organisation (modélisation dont on sait par ailleurs qu'elle induit des changements dans l'organisation, donc qu'elle est partiellement constitutive), loin du degré de généralité scientifique que visait l'IA.

Promouvoir un niveau au détriment d'un autre nous semblerait cependant une erreur d'interprétation et de KADS et de la notion d'agent rationnel abstrait□
- quoique les modèles de KADS soient ordonnés conceptuellement, cet ordre ne se traduit pas en un ordre temporel linéaire; ignorer le cycle de vie, analogue à celui de n'importe quel développement logiciel, aurait l'effet désastreux de creuser artificiellement un fossé infranchissable entre les savoirs et leur représentation ([BAC 94], qui ne mentionne pas une seule fois le cycle de vie, nous semble tomber dans ce piège, p. 314),
- l'agent rationnel abstrait est physiquement implémentable; il ne se réduit pas à sa spécification au niveau des savoirs, laquelle ne constitue un modèle que virtuellement□tant que le cycle n'est pas achevé, tant qu'il n'y a pas de système informatique concret, il n'y a pas d'objet à ce modèle.

L'IA construit progressivement un ordre (au sens de la classification des espèces) d'objets informatiques interprétables, dans leur contexte d'opération, comme des

agents rationnels. Ces objets sont construits sur des systèmes de symboles formels et les manipulent de manière qui les rend interprétables, toujours dans leur contexte, en termes de traitement de connaissances. Rien ne s'oppose à ce que cet ordre se compose de genres et d'espèces différents, distinguables soit par leur architecture globale (l'une pouvant être le modèle computationnel des espaces de problèmes) soit par la structure locale (briques représentationnelles de base) de leurs systèmes de symboles formels.

C'est, à notre sens, de par son insertion au niveau sociétal que cet ordre peut conquérir une certaine forme d'objectivité, dont le paragraphe 6.1.2 donne une idée - ce qui va à l'encontre de l'idée parfois admise sans plus de procès que l'IA s'enracinerait nécessairement dans le cognitivisme psychologique.

Remerciements

L'auteur remercie les referees pour leurs nombreuses remarques sur la première version de cet article.

Bibliographie

[AGR 82] Agre P.: Interview with Allen Newell, *Artificial Intelligence*, Vol 18, pp 87-127, 1982.

[BAC 94] Bachimont B.: *Le contrôle dans les systèmes à base de connaissances*, 2ème édition, Hermès, Paris 1994.

[BER 93] Berthier D.: Évolution et objet de l'IA, rapport interne de recherche N° 930901, Institut National des Télécommunications, 1993.

[BOB 75] Bobrow D. & Collins A., eds.: *Representation and Understanding*, Academic Press, New York, 1975.

[BRA 79] Brachman R.: On the Epistemological Status of Semantic Networks, in [Fin 79].

[BRA 85] Brachman R. & Schmolze J.: An Overview of the KL-ONE Knowledge Representation System, *Cognitive Science* Vol 9, N° 2, pp 171-216, April 1985.

[BRO 91] Brooks R.: Intelligence Without Representation, *Artificial Intelligence*, Vol 47, N° 1-3, 1991.

[BRO 86] Brownston L., Farrell L., Kant E. & Martin N.: *Programming Expert Systems in OPS5: an Introduction to Rule-Based Programming*, Addison Wesley, Reading, MA, 1986.

[CAR 73] Carbonnel J. & Collins A.: Natural Semantics in Artificial Intelligence, *Proceedings of the Third International Joint Conference on Artificial Intelligence*, 1973.

[CLA 83] Clancey W.: The Epistemology of a Rule-Based Expert System - a Framework for Explanation, *Artificial Intelligence*, Vol 20, pp 215-251, 1983.

[CLA 85] Clancey W.: Heuristic Classification, *Artificial Intelligence Journal*, Vol 27, 1985.

[CLA 87] Clancey W.: *Knowledge Based Tutoring□The Guidon Program,* The MIT Press, Cambridge, Mass., 1987.

[DAV 90] Davis E.: *Representations of Commonsense Knowledge*, Morgan Kaufmann Pub., 1990.

[DAV 93] David J.M., Krivine J.P. & Simmons R., eds. : *Second Generation Expert Systems*, Springer, 1993.

[DEL 73] Deleuze G.: A quoi reconnaît-on le structuralisme?, in Châtelet F., ed., *Histoire de la Philosophie*, Vol 8, Hachette, Paris 1973.

[DOR 93] Dorr B.: *Machine Translation□ A View from the Lexicon*, MIT Press, Cambridge, Mass., 1993.

[DUC 68] Ducrot O. et al.: *Qu'est-ce que le structuralisme?*, Editions du Seuil, Paris, 1968.

[ECO 68] Eco U.: *La Structure Absente: Introduction à la Recherche Sémiotique*, Mercure de France, Paris, 1968.

[FAH 79] Fahlman S.: *NETL: A System for Representing and Using Real-World Knowledge*, The MIT Press, Cambridge, Mass., 1979.

[FIN 79] Findler N., ed.: *Associative Networks: Representation and Use of Knowledge by Computer*, Academic Press, New York, 1979.

[FOR 77] Forgy C. & Mc Dermott J.: OPS, a Domain-Independent Production System Language, *Proceedings of the Fifth Internationnal Joint Conference on Artificial Intelligence,* Cambridge, Ma, 1977.

[FOR 93] Ford K. & Bradshaw J.: *Knowledge Acquisition as Modeling*, Wiley, New York, 1993. Book form of*International Journal of Intelligent Systems*, Special issues Vol 8, N° 1 & 2.

[GAN 93] Ganascia J.G.: *L'Intelligence Artificielle*, Flammarion, Paris, 1993.

[GRU 65] Gruber J.: *Studies in Lexical Relations*, Doctoral Dissertation, reprinted in *Lexical Structures in Syntax and Semantics*, North Holland, Amsterdam, 1976.

[GUH 89] Guha R. & Lenat D.: Cyc: a Midterm Report, *AI Magazine,* Vol. 11, N° 3, Fall 1989.

[HAY 77] Hayes P.: In Defense of Logic, *Proc. Fifth International Joint Conference on Artificial Intelligence*, Computer Science Dep., Carnegie-Mellon Univ., Pittsburgh, PA, 1977.

[HAY 83] Hayes-Roth B.: The Blackboard Architecture: a General Framework for Problem Solving?, Technical Report, HPP83-30, Stanford University, 1983.

[HAY 85] Hayes-Roth B.: A Blackboard Architecture for Control, *Artificial Intelligence*, Vol 26, N° 2, pp 251-321, 1985.

[HIC 89] Hickman F., Killin J., Land L., Mulhall T., Porter D. & Taylor R.: *Analysis for Knowledge-Based Systems: a Practical Guide to the KADS Methodology*, Ellis Horwood / John Wiley, 1989.

[JAC 83] Jackendoff R.: *Semantics and Cognition*, The MIT Press, Cambridge, Mass., 1983.

[JAC 87] Jackendoff R.: *Consciousness and the Computational Mind*, The MIT Press, Cambridge, Mass., 1987.

[JAC 90] Jackendoff R.: *Semantic Structures*, The MIT Press, Cambridge, Mass., 1990.

[LAI 86] Laird J., Rosenbloom P. & Newell A.: *Universal Subgoaling and Chunking: the Automatic Generation and Learning of Goal Hierarchies,* Kluwer Academic Pub., 1986.

[LAI 87] Laird J., Newell A. & Rosenbloom P.: Soar: an Architecture for General Intelligence, *Artificial Intelligence,* Vol 33, pp 1-64, 1987.

[LEN 90] Lenat D. & Guha R.: *Building Large Knowledge Based Systems: Representation and Inference in the CYC Project*, Addison Wesley Pub., 1989.

[LEV 91] Levin B. & Pinker S.: *Lexical and Conceptual Semantics*, Blackwell, Cambridge, Mass., 1991.

[MEY 88] Meyer B.: Object-Oriented Software Engineering, Prentice-Hall, Englewood Cliffs, N.J., 1988.

[NAG 92] Nagle T., Nagle J., Gerholz L. & Eklund P.: *Conceptual Structures: Current Research and Practice,* Ellis Horwood, Chichester, England, 1992.

[NEW 80a] Newell A.: Physical Symbol Systems, *Cognitive Science*, Vol 4, pp 135-183, 1980.

[NEW 80b] Newell A.: Reasoning, Problem Solving and Decision Processes : the Problem Space as a Fundamental Category, in R.Nickerson,ed., *Attention and performance* VIII, Lawrence Erlbaum, Hillsdale, NJ, 1980.

[NEW 82] Newell A.: The Knowledge Level, *Artificial Intelligence*, Vol 59, pp 87-127, 1982.

[NEW 90] Newell A.: *Unified Theories of Cognition*, Harvard Univ. Press, Cambridge, Mass., 1990.

[NEW 72] Newell A. & Simon H.: *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, N.J., 1972.

[NEW 91] Newell A., Yost G., Laird J., Rosembloom P. & Altmann E. : Formulating the Problem Space Computational Model, in Rashid R., ed., *Carnegie Mellon Computer Science : a 25-Year Commemorative*, Addison Wesley / ACM Press, Reading, MA, 1991.

[PET 92] Petitot-Cocorda J.: *Physique du Sens*, Editions du CNRS, Paris, 1992.

[PIN 89] Pinker S.: *Learnability and Cognition*, MIT Press, Cambridge, Mass., 1989.

[QUI 68] Quillian M.: Semantic Memory, in Minsky M., ed., *Semantic Information Processing,* MIT Press, Cambridge, Mass., 1968.

[RUM 72] Rumelhart D., Lindsay P., Norman D.: A Process Model for Long-Term Memory, in Tulving E. & Donaldson W., eds., *Organisation of Memory,* Academic Press, New York, 1972.

[SCH 75] Schank R.: *Conceptual Information Processing*, American Elsevier, New York, 1975.

[SCH 86] Schank R.: *Explanation Patterns: Understanding Mechanically and Creatively,* Erlbaum, Hillsdale, N.J., 1986.

[SCH 93] Schreiber G., Wielinga B. & Breuker J.: *KADS: A Principled Approach to Knowledge-Based System Development*, Academic Press, 1993.

[SEA 68] Searle J.: *Speech Acts*, Cambridge Univ. Press, 1968.

[SHA 71] Shapiro S.: A Net Structure for Information Storage, Deduction and Retrieval, *Proceedings of the Second Internationnal Joint Conference on Artificial Intelligence,* 1971.

[SHO 93] Shoham Y. : Agent Oriented Programming, *Artificial Intelligence,* Vol 60, N° 1, 1993.

[SIM 73] Simmons R.: Semantic Networks: their Computation and Use for Understanding English, in Schank R. & Colby K., eds., *Computer Models of Thought and Language,* Freeman, San Francisco, Ca., 1973.

[SMI 93] Smith J. & Johnson T. : A Stratified Approach to Specifying, Designing, and Building Knowledge Systems, *IEEE Expert*, June 1993.

[SOW 84] Sowa J.: *Conceptual Structures: Information Processing in Mind and Machine*, Addison Wesley Pub., 1984.

[SOW 91] Sowa J., ed.: *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, Morgan Kaufmann Pub., 1991.

[SWA 83] Swartout W. : XPLAIN, a System for Creating and Explaining Expert Consulting Systems, *Artificial Intelligence,* Vol 21, N° 3, pp 285-325, 1983.

[TAL 78] Talmy L.: The Relation of Grammar to Cognition - a Synopsis, in Waltz D., ed., *Theoretical issues in language Processing 2,* ACM, New York, 1978.

[TAL 80] Talmy L.: Lexicalization Patterns: Semantic Structure in Lexical Forms, in Shopen T. et al., eds., *Language Typology and Syntactic Description,* Vol 3, Cambridge University Press, New York, 1980.

[TAL 83] Talmy L.: How Language Structures Space, in Pick H. et Acredolo L., eds., *Spatial Orientation: Theory, Research, and Application*, Plenum, New York, 1983.

[TAL 88] Talmy L.: Force Dynamics in Language and Thought, *Cognitive Science,* Vol 12, pp 49-100, 1988.

[WOO 75] Woods W.: What's in a Link: Foundations for Semantic Networks, in [Bob 75].