

**Yves-Marie VISETTI**

## **DES SYSTÈMES EXPERTS AUX SYSTÈMES À BASE DE CONNAISSANCES : À LA RECHERCHE D'UN NOUVEAU SCHÉMA RÉGULATEUR**

**Résumé.** L'article décrit trois sortes de contraintes auxquelles sont soumis les systèmes à base de connaissances : contraintes de la modélisation cognitive, de la reconstruction logique, de la communicabilité et de l'interaction en situation réelle. Les conflits de priorité qui en résultent sont tout particulièrement soulignés. Simultanément, on rappelle quelques-unes des critiques les plus couramment adressées au paradigme dominant de l'IA, celui des systèmes symboliques classiques. On donne ensuite les grandes lignes d'un "nouveau" schéma régulateur, au plan épistémologique, pour la conception et l'utilisation de ces systèmes, qui sont alors présentés comme des aides à la modélisation.

**Abstract.** We describe here three kinds of constraints to which knowledge-based systems are submitted : constraints of cognitive modelling, of logical reconstruction, of communicability and interaction in a real situation. Priority conflicts between these constraints are specially analyzed. Simultaneously, the paper examines in a critical way the main paradigm of AI, that of classical symbol systems. We next give the broad lines of a "new" regulating scheme, at the epistemological level, for designing and using these systems, which are then presented as aids to modelling.

Les systèmes experts sont donc pour ainsi dire en voie de disparition. Lentement et sûrement, ils cèdent progressivement la place aux systèmes à base de connaissances. Certains départements industriels modifient leur dénomination en conséquence, et il en va ou ira de même pour les programmes de recherche, les cours d'université ou les conférences spécialisées... Après une quinzaine d'années d'usage triomphal, le label de *système expert* pourrait bien tomber en désuétude, avec l'accord plus ou moins tacite de ceux qui l'ont défendu jusqu'ici et l'emploient à présent avec des connotations restrictives.

Le fait est qu'il flotte désormais autour de ce vocable comme un parfum de désuétude, de naïveté, de rusticité des temps pionniers<sup>1</sup>. Mais que veut-on dire maintenant, en termes de sémantique différentielle, lorsqu'on promeut à sa place l'appellation de *système à base de connaissances* ? Veut-on suggérer qu'il faudrait faire appel à des *connaissances* différentes

---

<sup>1</sup> La revue de l'Association Américaine d'Intelligence Artificielle, *AI Magazine*, a publié en 1989 (vol. 10, nos 1 et 2) la retranscription d'un débat de 1985 (organisé dans le cadre de l'IJCAI à Los Angeles) sur le thème *Les systèmes experts : jusqu'où peuvent-ils aller ?* Les participants en étaient T. Winograd, S. Dreyfus, B. Smith et R. Davis. Vue de 1991, l'affaire ne devrait plus guère nous émouvoir. Et pourtant... Winograd et S. Dreyfus y résumaient certaines des thèses que l'on a pu lire en détail un an plus tard (1986) dans leurs livres respectifs *Understanding Computers and Cognition* (écrit avec F. Flores) et *Mind over Machine* (œuvre commune de Stuart et Hubert Dreyfus). B. Smith esquissait une problématique de la modélisation et de ses limites, qui n'est pas sans points communs avec celle qui sera avancée ici-même en conclusion. Quant à R. Davis, jusque là peu suspect de connivence avec les deux premiers intervenants, il est plaisant de le voir s'en prendre à l'appellation de *système expert* (“des systèmes qui ne sont pas experts”), pour en rejeter malicieusement la responsabilité sur les “gens de Stanford” (également crédités par lui de l'expression calamiteuse d'*Intelligence Artificielle*). La revue prend bien soin de préciser, probablement à l'intention de ceux qui voudraient faire mine d'avoir déjà entendu tout cela : “Bien que quatre années ou presque aient passé depuis que ce débat a eu lieu, les questions soulevées et les points en discussion ne nous paraissent pas moins pertinents aujourd'hui”. *Indeed...* Voir aussi, plus récemment, le débat de Lenat et Feigenbaum avec B. Smith dans le numéro spécial de *Artificial Intelligence* consacré aux fondements de l'IA (vol. 47, 1-3, janvier 91).

de celles que les experts reconstruisaient, avec l'aide des cognitivistes, à l'époque déjà reculée des systèmes experts de première et seconde générations ? Ou bien que les structures et processus reconstruits n'ont plus qu'un rapport très lointain avec les véritables *connaissances* expertes qui en constituent la *base* ? Ou encore que les *connaissances*, quelles qu'elles soient, ne sont qu'une *base* parmi d'autres également nécessaires à la production d'un comportement complexe ? Ou enfin, quitte à jouer sur les mots, que le système ne dispose pas à lui seul des moyens de comprendre une situation, et qu'il ne peut donc offrir qu'une *base* pour une compréhension véritable ? Sans être nécessairement praticiens ou théoriciens dans ce domaine emblématique de l'Intelligence Artificielle qu'est la modélisation des raisonnements, nous n'en sommes pas moins les observateurs ou les acteurs de ce glissement terminologique. Même si sa fonction initiale, purement épisodique, était de signaler l'arrivée sur le marché de systèmes plus diversifiés que ceux de la première génération, il nous semble maintenant que sa valeur principale, durable celle-ci, est d'*escamoter la référence directe à l'expertise* : signe probable d'un changement d'attitude vis-à-vis des possibilités de la simulation, et plus généralement symptôme d'une modification des croyances quant à la portée de l'Intelligence Artificielle<sup>2</sup>.

---

<sup>2</sup> D'une certaine façon, le concept de système à base de connaissances n'est qu'une notion vague, générique, qui pourrait s'appliquer à tous les secteurs de l'IA. On la prendra ici dans un sens plus spécifique, pour désigner des systèmes qui, tels les systèmes experts : (a) contribuent à la construction de *raisonnements* d'une relative technicité (complexité, cohérence) ; (b) y parviennent par des moyens explicables dans un style "déclaratif" et "symbolico-inférentiel" (leur utilité provient de ce type d'explicabilité) ; (c) tirent leur principale justification de leur dimension applicative (utilitarisme de principe). Les environnements de simulation, les outils de modélisation mathématique ou physique, et même les bases de données relationnelles, les systèmes d'indexation ou certaines interfaces en langue naturelle, ne sont donc pas *en eux-mêmes* des systèmes à base de connaissances dans le sens restreint du terme (mais ils en constituent éventuellement des composants essentiels). On reviendra sur le schéma de base des SBDC dans la dernière section de cet article, et notamment pour y intégrer les modalités visuelles, iconiques, de la compréhension.

Tout progrès – ou changement de perspective – apporté à la conception ou à la réalisation des systèmes (autrefois) experts devrait tôt ou tard avoir des effets sur l'IA tout entière. Cette évidence ne tient pas seulement à l'interdépendance des quatre domaines traditionnels de la modélisation cognitive (perception, langage, raisonnement, action) ; elle tient également au nombre et à l'importance relative des applications, qui se développent pour la plupart dans une remarquable continuité de méthode avec la recherche théorique. Ainsi les problèmes rencontrés et la considérable expérience accumulée sont intégralement rapportables aux schèmes généraux de modélisation qui les ont rendus possibles. Cette sorte de privilège épistémologique contraste avec ce que l'on peut observer ailleurs, par exemple dans le champ des industries de la langue, qui ne peuvent à l'heure actuelle être considérées au même titre comme la branche appliquée d'une spécialité de l'IA.

De plus, c'est bien à l'interface entre IA et psychologie qu'ont été élaborés un grand nombre des concepts fondamentaux de la modélisation cognitive – nous reviendrons sur ce point dans la section suivante – et ces concepts sont dans l'ensemble connus des spécialistes des systèmes à base de connaissances, quand ils ne sont pas tout simplement inventés par eux. De ce fait, un examen de leurs outils, méthodes, et résultats, pourrait bien nous donner, plus encore qu'une indication globale de l'évolution de l'IA, une évaluation intéressante du crédit attribuable aux grands paradigmes explicatifs des sciences cognitives, que l'on peut considérer comme *effectivement impliqués dans la réalisation et l'emploi* des ci-devant systèmes experts. C'est à quoi cet article veut aussi contribuer, d'une façon qui ne se prétend évidemment pas exhaustive, ni plus impartiale ou définitive que d'autres qui peuvent lui être opposées. Il s'agit en somme de présenter dans ses grandes lignes un cadre de discussion dans lequel il soit possible d'estimer la portée et la généralité potentielle des systèmes à base de connaissances, de reconnaître les évolutions probables de leurs schémas de conception, d'en tirer les enseignements généralisables à l'IA dans son

ensemble, voire d'observer les similitudes ou différences les plus significatives vis à vis d'autres courants remarquables des sciences de la cognition naturelle ou artificielle<sup>3</sup>.

Très schématiquement, donc, et avec une certaine dose d'arbitraire, nous présenterons le problème de la construction automatique de raisonnements comme relevant de trois ordres de contraintes (ou de ressources, si l'on préfère) :

- les contraintes de la modélisation cognitive,
- les contraintes de la reconstruction logique,
- les contraintes de la communicabilité et de l'interaction en situation réelle, avec l'expert comme avec l'utilisateur.

Nous examinerons tour à tour ces différentes perspectives (dont aucune n'épuise à elle seule le domaine, et encore moins l'IA entière), en soulignant à la fois leur nécessité et leurs insuffisances, leur dépendance mutuelle et leurs incompatibilités – avant que d'en tirer des conclusions, nécessairement partielles, partiales et provisoires, comme nous venons de le dire. Le ton général de cet article pourra sembler par trop “négatif”, mais son but n'est pas de fermer ou dissimuler des options : on veut au contraire rappeler qu'il en existe, et que les présupposés, les hypothèses, les méthodes peuvent et doivent toujours être explicités, confrontés, réévalués. Les choix nécessaires ne procèdent pas de situations à sens unique, mais au contraire multiples, qui font que la décision reste toujours quelque peu un pari.

---

<sup>3</sup> Travail qui requiert probablement le format d'un livre pour être mené à bien, et offrir une base suffisante à l'expression de (dés)accords intéressants. Une étude inachevée, aux conclusions trop rapidement formulées comme celle qui est présentée ici, ne peut que provoquer de violentes frustrations chez un lecteur délicat, à juste titre mécontent qu'on ne se soit pas autocensuré avec toute la vigueur dont il se sait capable. Qu'on nous pardonne d'avoir passé outre.

### *I. Les contraintes de la modélisation cognitive.*

Écartons d'abord une possible interprétation erronée de ce titre. Les chercheurs qui croient à la fidélité psychologique des traductions artificielles de l'expertise sont aujourd'hui une minorité – en IA plus encore qu'en psychologie, peut-être. Parmi ceux qui considèrent par ailleurs l'Intelligence Artificielle comme une sorte de “science des conditions les plus générales de l'intelligence”, bien peu iraient jusqu'à penser qu'il vaut mieux, pour construire ou déduire cette prétendue “intelligence générale”, éviter toute distinction entre intelligence humaine et artificielle<sup>4</sup>. Enfin – et c'est là le point décisif – la perspective “artificialiste”, qui domine à juste titre dès qu'il s'agit de traitement de l'information, réduit les recherches cognitives à l'état de simples indices, extrêmement utiles mais non contraignants, s'il s'agit de concevoir des intelligences artificielles de substitution. C'est donc par ce biais, et par ce biais seulement, que se fait sentir ici la profonde nécessité de la modélisation cognitive, que ce soit pour définir les structures internes des systèmes ou bien améliorer l'ergonomie des interfaces (soulignons toutefois que ce statut purement instrumental, utilitaire, serait peut-être problématique s'il s'agissait de traitement automatique des langues *naturelles*). L'originalité de la situation présente (ou imminente) dans le domaine des systèmes à base de connaissances réside précisément dans l'association d'un artificialisme ouvertement déclaré avec les tendances de pointe du génie cognitif ou linguistique. On est donc loin, en principe, de postuler une quelconque harmonie préétablie entre l'objectivité psychologique et l'objectivité informatique, comme celle que suggérait naguère la métaphore publicitaire des “gisements de connaissance” : évolution fort positive qui nous éloigne de l'époque où l'on pouvait découvrir, réunis sous ce type de slogans, une modélisation cognitive trop souvent indigente (méconnaissant les

---

<sup>4</sup> C'est pourtant la position constante de Newell, qu'il a soutenue encore récemment dans un article collectif (Rosenbloom et al., 1991, p. 292).

difficultés et le sens du transfert de connaissances) combinée à un artificialisme sournois (s'en remettant par force à la spontanéité des utilisateurs pour déterminer la véritable utilité des systèmes construits).

Le problème se pose alors de trouver un cadre général de description où recueillir les indices cognitifs pertinents avant de les transférer dans l'univers informatique. Ce cadre, comme on sait, est à l'heure actuelle principalement défini par le paradigme “classique” des sciences cognitives. Il en existe d'ailleurs bien des variantes, qui s'opposent sur des points fondamentaux, et notamment sur la pertinence des programmes de recherche de l'Intelligence artificielle (on évoquera ici, sans les exposer, les différences entre les conceptions actuelles de Fodor et celles exprimées avec continuité ces vingt dernières années par Newell et Simon<sup>5</sup>). La version relativement vague qui en a été majoritairement retenue par les chercheurs en IA nous représente les processus cognitifs, qu'ils soient naturels ou artificiels, comme fondés sur la manipulation *purement syntaxique* des symboles physiquement instanciés d'un “langage de la pensée”, très exactement conçu comme un langage formel. Ces symboles, en tant qu'entités physiques, jouent un rôle causal précis dans l'interaction du système ou de l'organisme avec son environnement<sup>6</sup>. Ils constituent les

---

<sup>5</sup> Thèmes de comparaison possibles (que nous ne développerons pas ici) : types de modularisme, formes de rationalisme (innéiste ou empiriste), primauté des formats propositionnels ou applicatifs, relations entre le système symbolique physique et l'intentionnalité, exprimabilité des contenus sémantiques profonds à partir de notions de sens commun, etc... On sait que les “modularistes” comme D. Marr ou J. Fodor ont estimé prématuré de chercher à comprendre, comme on le fait en IA, les processus cognitifs centraux, éminemment “holistiques”, qui seraient en charge de la construction des buts et des croyances. Seuls certains processus périphériques (correspondant aux problèmes de type 1 chez Marr, ou aux modules cloisonnés de Fodor) seraient pour le moment accessibles à l'investigation, analysables en primitives, puis modélisables en termes d'un calcul. Les problèmes de classification, d'heuristique, de *design*, de diagnostic relèvent évidemment des processus centraux, et seraient donc intouchables.

<sup>6</sup> Cette assimilation des systèmes et des organismes est une tendance irréprouvable des approches fonctionnalistes de la cognition. La Nature pratiquerait donc une sorte

supports nécessaires et suffisants de l'intentionnalité du système, responsable de la visée des objets et des situations comme de la détermination des buts. Pour une bonne part d'entre eux, également, ils sont intuitivement interprétables à la façon d'une notion de sens commun, voire assimilables au contenu d'un terme de la langue naturelle, à la condition que ce terme soit correctement appréhendé dans un usage local bien déterminé. La frontière entre langage interne et langue externe se laisserait donc franchir sans trop de difficultés – du moins peut-on l'espérer. Dans la mesure où toute performance “cognitive” est en principe rapportable au premier de ces deux langages, on sera tout aussi facilement porté à confondre sous la même notion de *connaissance* – et en un certain sens, on *doit* confondre – la possession de schémas corporels ou de savoir-faire opératoires à la systématité plus ou moins affirmée (comme marcher, danser, calculer, raisonner, parler...), et la mémorisation d'une représentation explicite<sup>7</sup> qui fait, en quelque sorte, fonction de *théorie* pour ces *pratiques*.

Au moins dans une première approche, les systèmes experts relèvent tous de ce paradigme – par définition même, pourrait-on dire<sup>8</sup>. Quel que soit le rôle actuel de la modélisation cognitive, ils sont encore (et resteront probablement) marqués par des structures d'inspiration psychologique, comme celles construites dans les trente dernières années par un des courants principaux de l'IA de cette période, celui de la *simulation cognitive*. Le schéma général d'instanciation de ces systèmes symboliques

---

d'ingénierie des connaissances. A titre d'exemple relativement récent, on a pu entendre un spécialiste fort connu de la représentation des connaissances, et notamment de leur représentation logique, terminer un exposé sur ces questions en mentionnant sans crier gare : “les connaissances qui nous sont nécessaires pour marcher” (colloque sur les sciences cognitives, Cerisy-la-Salle, 1987).

<sup>7</sup> Il s'agit ici d'une représentation interne ; mais comme on la dit “explicite”, elle s'externalise quelque peu...

<sup>8</sup> Seules quelques minuscules réalisations connexionnistes de systèmes de productions “jouets” font pour le moment exception : cf. par exemple Touretzky and Hinton, Dolan and Smolensky (1988).

“stricts” (que l'on suppose représenter le niveau fonctionnel de la cognition, susceptible en tant que tel de concilier le naturel et l'artificiel) est alors, comme on sait, le suivant :

(i) Les “connaissances” sont généralement exprimées sous forme de règles de production et de schémas hiérarchisés (offrant des possibilités de déclarations de contraintes, de déclenchements opportunistes de procédures, de raisonnement par défaut ou de raisonnements hypothétiques avec un certain suivi de la cohérence). Ces différentes notions sont définissables dans un cadre unique, celui des langages *objets*, si bien qu'une règle, par exemple, est un objet comme un autre ; le système dispose donc en principe de la possibilité de réifier ses propres processus pour raisonner sur les structures ainsi obtenues ; un raisonnement apparaît ainsi, à la limite, comme la construction ou la transformation d'un objet “générique” réalisant un certain agencement d'objets de base (objets du domaine ou objets-tâches, selon les cas, tels que les présentent Chandrasekaran, McDermott, et bien d'autres<sup>9</sup>).

(ii) La modularité est un principe impératif, et les unités significatives “minimales” sont généralement de petite taille (prémises ou conclusions d'une règle, par exemple, dépassent rarement la dizaine). Seule l'inférence déploie des structures de grande dimension, qu'une modification locale peut perturber massivement. Ces structures sont en général oubliées après retrait des conclusions obtenues lors d'une session, si bien que le fonctionnement du système ne contribue pas à le réorganiser sur le long terme.

(iii) Les objets qui sont évidemment, à un certain niveau d'abstraction, tous de la même nature, se répartissent en types suivant plusieurs axes de diversification liés à la modélisation du domaine (structures, fonctions, lois) ou à la modélisation du raisonnement (tâches,

---

<sup>9</sup> Cf. par exemple l'article de synthèse de L. Steels dans *AI Magazine* (été 90), le numéro de cette même revue consacré au *design* (hiver 1990), et la contribution de J.-P. Krivine et J.-M. David à ce volume.

méthodes, stratégies), chaque objet s'inscrivant dans plusieurs hiérarchies, recevant des modalités temporelles ou hypothétiques, etc.<sup>10</sup>. Cependant, quelle que soit la sémantique visée et la complexité obtenue, on peut dire que le *pattern* générique sous-tendant intuitivement toutes ces structures dans l'esprit du modélisateur reste une sorte de *graphe*, dont les noeuds seraient constitués par des schémas discrets et finis (c'est-à-dire par d'autres graphes à une échelle inférieure, permettant la récursion). La

---

<sup>10</sup> De plus, les objets peuvent subir de la part du modélisateur, à différents “niveaux” de la représentation, des transformations du type *compilation*, qui n'ont pas d'autre motif que l'efficacité informatique, et ne représentent donc pas une réélaboration des données sous d'autres modalités cognitives. Cette compilation ne doit pas être confondue avec celle, de nature psychologique, qui est censée représenter la formation progressive d'une expertise (selon J. Pitrat, par exemple). J.-P. Desclés, de son côté, utilise le terme de *compilation* en deux sens différents au moins : pour désigner tout d'abord la relation entre deux niveaux d'explication des sciences cognitives, c'est-à-dire pour marquer les étapes d'un programme réductionniste (heureusement fort tempéré) qui consisterait à se rapprocher “progressivement, par un processus complexe de compilation, de représentations qui seraient compatibles avec les structures neuronales des supports biologiques (du cerveau) identifiables par les neurosciences” (1990, p. 52) ; ou bien, ce qui est *a priori* différent, pour désigner la traduction de représentations centrales en représentations périphériques (ou vice-versa) par le cerveau lui-même, qui “procède très certainement en mettant en jeu des compilations entre représentations ... [et] engendre alors des représentations symboliques intermédiaires entre les représentations les plus externes et les expressions compilées” (*ibid*, p. 53). L'usage du terme de *compilation*, qui a certes des vertus métaphoriques, a aussi le défaut, si on le prend dans son sens classique, de suggérer des caractéristiques bien précises, comme une parfaite autonomie descriptive des niveaux reliés, l'unidirectionnalité de la traduction (toujours du “haut” vers le “bas”), le caractère à la fois nécessaire et suffisant (pour une exécution) du niveau le plus bas (c'est-à-dire le plus compilé), etc. A cela s'ajoute évidemment le présupposé logico-syntaxique de la programmation elle-même, qu'il paraît difficile de considérer comme une simple “métaphore” (si tentant que soit cet accommodement), alors qu'il s'agit bel et bien, dans le paradigme computo-représentationnel “classique”, du postulat fondamental constituant l'objectivité mentale en même temps que sa représentation littérale dans les champs scientifiques. Et l'on sait qu'à l'intérieur de cette tradition, les métaphores sont généralement analysées comme des effets de sens seconds ou dérivés : qualifier de cette façon le geste premier, fondateur, de ladite tradition serait donc une source de tension supplémentaire.

construction du système dépend étroitement d'une réduction de l'univers exploré à ce type d'objectivité.

Voilà donc, brièvement évoqués, les structures et processus, encore trop simples et uniformes, que l'on met en oeuvre pour modéliser des activités cognitives qui devraient être distinguées davantage, et autrement peut-être qu'une morphologie et une syntaxe formelles semblent le permettre. Quant à savoir sur quoi l'on pourrait appuyer les distinctions, peut-être non fondées, que l'intuition semble commander, ce n'est pas facile de le dire. Les typologies des connaissances et savoir-faire ne courent pas encore les rues. Les ergonomes, par exemple, semblent apprécier le modèle très général de Rasmussen, qui distingue entre trois types de comportements emboîtés les uns dans les autres : les comportements routiniers sans contrôle conscient, les comportements "opératifs" caractérisés par l'application non problématique de règles et procédures mémorisées, et les comportements de résolution de problèmes en situation nouvelle où l'on fait appel à des savoirs d'arrière-plan non techniques, aussi bien qu'à la réflexion ou aux rapprochements analogiques de toutes sortes (cf. Falzon, 1989). Quant aux "connaissances", on peut au moins envisager, comme l'a fait V. Prince, une classification par le mode d'acquisition : simple conditionnement, emprunt à un fonds théorique livresque, ou transmission et expérience réfléchie. Ces distinctions sont probablement liées à la rapidité d'application, à l'adaptabilité, à la possibilité d'une verbalisation <sup>11</sup>. Mais elles n'ont pour le moment aucune chance de se refléter dans le fonctionnement des systèmes symboliques, puisque les représentations qui les constituent n'ont pas d'*histoire*. Soulignons, au risque d'appuyer sur une évidence, que les enjeux, sur tous ces points, ne sont pas seulement "théoriques" ou "descriptifs" : leur importance pratique est considérable. Tant que l'expertise est traitée comme une donnée primitive et homogène, et qu'on ne sait pas distinguer entre ses différentes phases ou dimensions,

---

<sup>11</sup> Cf. Prince 1989. Cf. également l'opposition connaissances privées/connaissances publiques décrite par C. Vogel (1988).

on ne dispose en fait d'aucun critère pour identifier les moments et les domaines où une systématisation informatique devient possible et intéressante<sup>12</sup>. Certes, Newell et Simon avaient bien esquissé quelques remarques en ce sens, mais il n'a été tenu aucun compte de leurs observations dès lors qu'elles présentaient une tonalité "limitante" (dont ils se sont du reste assez vite désintéressés eux-mêmes). T. Winograd et F. Flores, H. et S. Dreyfus s'y sont également essayé – de façons plus radicales – mais leurs réflexions ont été jusqu'ici assez mal reçues dans les milieux de l'Intelligence Artificielle.

Cependant, en dépit de ces considérations plutôt négatives, il faut rappeler que c'est au développement de l'IA, dans son interface avec la psychologie, que l'on doit la reconnaissance d'une certaine diversité des modes de raisonnement ou de compréhension : qu'on pense seulement aux notions d'heuristique et de règles de production (héritées du modèle de la *rationalité limitée* mis en avant par Simon, en opposition au modèle de l'optimisation, pour rendre compte des mécanismes de la prise de décision) ; ou bien à la notion de *métacognition* (R. Davis, J. Pitrat) ; ou encore au concept de réseau sémantique, aux recherches sur la typicalité et les prototypes ; aux tentatives de formaliser le raisonnement par défaut ou dans l'incertain ; à l'introduction de primitives conceptuelles, de plans et de scénarii (Schank et Abelson) ; à la prise en considération de la complexité du sens commun et des savoirs d'arrière-plan, etc. C'est toute la contribution de l'IA à la représentation des connaissances et aux logiques non classiques qu'il faudrait mentionner ici. Sans oublier les recherches portant sur l'architecture des systèmes (*tableaux noirs*) ou les langages de

---

<sup>12</sup> C'est là tout le problème des fameux "x% de bonnes réponses" : comment les reconnaître ? à quelle dimension de l'expertise correspondent-elles (novice, débutant avancé...) ? est-il juste de concevoir alors les systèmes comme s'il s'agissait de remplir un objectif de 100%, alors que l'on construit en réalité un objet qui relève d'une toute autre dimension ?

programmation (*objets, acteurs*) qui ne sont pas dépourvues, souvent, de motivations cognitives ou épistémologiques<sup>13</sup>.

De plus, la modélisation du diagnostic est depuis SOPHIE ou GUIDON l'objet de travaux très importants, qui ont entraîné d'un côté le développement d'une physique qualitative (de Kleer, Forbus, Kuipers...), et de l'autre amené certains auteurs à souligner l'importance de l'opération de *catégorisation* dans la résolution de problèmes (W. Clancey)<sup>14</sup>. Il en est résulté parfois des prises de position fortes en faveur d'une théorie "prototypale" du concept, par opposition à une définition en termes de conditions nécessaires et suffisantes, ou à une caractérisation en termes "logiques". Malheureusement, les effets de prototypie modélisés se réduisent le plus souvent à une simple réification, c'est-à-dire à la représentation de quelques objets typiques des catégories concernées ; ou bien ils sont assimilés, mais à tort le plus souvent, à un phénomène d'appartenance graduelle ou d'incertitude. Plus grave encore, l'organisation interne des catégories est souvent trop peu contextualisée : les objets typiques sont toujours les mêmes, quelle que soit la perspective. Mais il ne faut pas croire pour autant que le relatif simplisme de ces procédés soit méconnu ; il est le plus souvent délibéré, et tient à l'extrême complexité des problèmes. Ainsi par exemple, Minsky évoquait déjà en 1961 la complexité des "exigences téléologiques de la classification" ; ce faisant il anticipait la position que Schank a soutenue, plus tard, à propos de la création et de la modification des catégories qui seraient, selon lui, profondément déterminées par les nécessités fonctionnelles de leur emploi

---

<sup>13</sup> Pour une anthologie de la représentation des connaissances en IA, voir Brachman et Levesque (1985). Sur les architectures de *tableau noir*, cf. Hayes-Roth 85 (et pour une complexification, Bachimont 90).

<sup>14</sup> Voir par exemple Clancey (1985), et le recueil *Readings in qualitative reasoning about physical systems*, édité par D. Weld et J. de Kleer (1990, Morgan Kaufmann). Les systèmes SOPHIE et GUIDON sont présentés dans le livre de Wenger (1987).

(contraintes de formation qu'il proposait d'appeler *pragmatiques*, par analogie avec le domaine linguistique)<sup>15</sup>.

On peut dire, en somme, que la complexité des problèmes abordés, la variété et la relative faiblesse des techniques disponibles sont maintenant des données connues de tous. Certaines déclarations outrancières, qu'il était possible d'entendre il n'y a pas si longtemps encore, sur l'efficacité des systèmes experts, l'enseignement assisté ou le traitement des langues naturelles, ne sont plus de mise aujourd'hui. Nous assistons à une unification, une mise en commun des résultats et des problématiques de divers secteurs de l'IA, et le domaine des systèmes experts en bénéficie comme les autres. La conception étroite de l'expertise, dite de première génération, est donc révolue, et avec elle l'espoir de gains de productivité faciles par le biais d'une automatisation généralisée. La voie pourrait donc sembler ouverte pour une approche moins simpliste des processus cognitifs, qui sache faire une place aux différents styles de comportements ou de connaissances que nous avons évoqués précédemment, et dont il y a tout lieu de penser qu'ils se retrouvent complètement enchevêtrés, imbriqués en situation réelle.

Mais il faut bien rappeler qu'à chaque fois, jusqu'ici — et c'est probablement une conséquence insurmontable du paradigme cognitiviste classique, c'est-à-dire de la tentative de faire directement supporter par une syntaxe formelle des notions qui n'en garderaient pas moins un sens psychologique intuitif — les processus de pensée représentés ont pris invariablement, *dès que l'on a recherché en même temps complexité et cohérence*, une forme hypothético-déductive très réductrice. Réductrice parce qu'elle finit toujours par imposer des règles strictes d'inférence

---

<sup>15</sup> Minsky (1961) est cité par Schank, Collins et Hunter (1986). Les positions évoquées ici semblent impliquer une conception intuitive des catégories beaucoup moins normative que celle que E. Rosch a développée. Les relativisations étudiées par Rosch sont davantage spécifiques aux cultures qu'aux individus et à leurs fins propres.

logique à des symboles aux dénotations sans équivoque (ou dont l'équivoque, le vague, ne parviennent jamais à rentrer en jeu convenablement)<sup>16</sup>. Réductrice encore parce qu'elle force à reconstituer sous forme d'inférences longues et disparates ce qui semble plutôt relever d'un schéma interprétatif immédiat et spécifique (C. Vogel, 1988, ch. 7). Réductrice également, parce qu'en dépit des trouvailles de Minsky, Schank ou Winograd par exemple, elle n'offre toujours pas de reproduction convaincante d'un raisonnement par analogie, ni de mécanisme de mémoire associative avec complétion ou redressement de la forme d'appel ; pas non plus de structure interne des catégories qui puisse guider les rapprochements, ni de formation de catégories adaptées à la situation courante.

Le problème semble bien être le suivant : une fois qu'on a admis une forme ou une autre de *schématisation* au centre des processus de pensée, en distinguant ses fonctionnalités de celles de l'inférence logique, on voit moins bien comment réduire les fonctions du concept (en tant qu'objet de la psychologie) à celles qu'occupe un simple terme à l'intérieur d'une langue formelle, serait-elle "profonde". On est alors amené à tenter de comprendre la conceptualisation, puis l'enchaînement et le regroupement des concepts, c'est-à-dire le raisonnement, en affinité avec un phénomène que le point de vue "orthodoxe" entend au contraire traiter d'une façon antagoniste : il s'agit évidemment ici de la *perception*. Les observations ou les introspections praticables dans ce domaine de la perception, où sont impliquées sensation et catégorisation, suggèrent en effet une affinité profonde avec ces conceptions schématisantes du raisonnement que l'on cherche à établir. Elles s'accordent également avec certaines modalités du sens linguistique mises en avant par des sémanticiens comme Lakoff,

---

<sup>16</sup> Cf. *infra*, notamment à la fin de la section 2.

Langacker, Rastier<sup>17</sup>. Elles soulignent par contraste l'importance des images mentales dans une tâche de compréhension. Lorsque l'expert examine une structure physique, écrit C. Vogel, on doit imaginer que pour lui "l'espace du problème à résoudre est au sens propre un espace physique du problème, représenté de façon plus ou moins métaphorique et traversé par des forces et des contraintes." (1988, p. 76). Les *espaces mentaux* (c'est-à-dire en fait leurs objectivations physico-mathématiques et informatiques) devraient donc présenter à la fois des aspects discrets, articulés, et des aspects continus, progressifs. La représentation des dimensions abstraites, symboliques de la pensée, et simultanément de ses modalités concrètes, morphologiques, constitue bien le défi incontournable que l'IA est censée relever<sup>18</sup>.

C'est ici, dans cet appel, métaphorique ou réaliste, à la perception, que se rejoignent certains partisans tout comme certains détracteurs de l'IA classique et du paradigme qui lui est associé. Pour Schank, Minsky et le Winograd des années 70, *comprendre* c'est d'abord percevoir et se remémorer, ou plus généralement *reconnaître*, c'est-à-dire apparier "d'un seul coup" des formes génériques avec des formes spécifiques, afin de les compléter, de les redresser, ou bien de les étendre par analogie. L'inférence est alors assimilée à une suite de réponses réflexes, qui ne devrait plus

---

<sup>17</sup> Pour ces auteurs, par exemple, la "faculté de langage" ne peut reposer sur des modules cloisonnés à la Fodor. Ces modules seraient en effet de simples corrélats, dans l'architecture cognitive, d'une nouvelle version de l'autonomie de la syntaxe, dont la reconnaissance aurait été réduite à une sorte de "réflexe perceptif". Le désaccord avec Fodor ne porte pas principalement sur l'appel, métaphorique ou réaliste, au domaine de la perception ; il porte sur la perception elle-même, que Fodor conçoit comme une sorte de réflexe, c'est-à-dire une "transduction" fermée à tout sémantisme. Cependant, les conclusions sur le découpage et l'interaction des disciplines (linguistique, psychologie...), la conception de la sémantique (symbolique ou iconique), de même que l'emploi des notions de *schème* ou de *concept* varient fortement suivant les auteurs.

<sup>18</sup> Dans la mesure où elle témoigne (avec perversité) de la multimodalité des représentations, l'interprétation des *rébus* devrait figurer parmi les questions-limites, les utopies fondatrices de l'IA.

grand chose aux règles d'inférences logiques, au souci d'exhaustivité, ou à la notion logique de vérité. Mais bien sûr tous ces processus restent, dans cette approche, strictement “symboliques et computationnels”.

Hubert Dreyfus de son côté, en grand pourfendeur de l'IA classique, est porté à ne concéder aux représentations mentales qu'un rôle minimal, “instrumental”. Pour lui, comme d'ailleurs pour le Winograd des années 80, le monde, s'il est perçu, ne peut l'être parce qu'il aurait *déjà* été représenté. Les représentations, si tant est qu'elles existent, jouent un rôle dans nos conduites uniquement dans la mesure où elles sont rapportées à un fonds, à des ressources, qui ne peuvent être représentationnelles, ni *a fortiori* calculatoires sur une telle base. De ce point de vue, par conséquent, l'IA a tort de croire que les conduites intelligentes se ramènent à un calcul sur des représentations. Mais Dreyfus est aussi d'accord pour privilégier le domaine sensori-moteur, la perception, dans l'approche de la cognition humaine. Et précisément, il s'appliquera à prouver que l'IA classique ne jette pas la moindre lumière sur les phénomènes perceptifs, qu'ils soient considérés en eux-mêmes ou comme métaphore de la cognition tout entière.

Et de fait, en dépit de travaux remarquables sur les schémas ou les prototypes, qui étaient les structures sur lesquelles l'IA fondait de grands espoirs en ce sens, et qui ont d'ailleurs renouvelé la conception des systèmes symboliques, on ne peut certainement pas dire que les limites auraient été radicalement repoussées sur ce plan. Il faut d'abord rappeler (mais nous y reviendrons dans la section suivante) que les “nouvelles” formes d'inférence que ces approches ont rendues possibles n'ont rien de si nouveau qu'on ne puisse facilement reformuler avec un peu de logique des prédicats, comme Hayes l'avait fait remarquer (1979). De plus, si la modélisation effective de la vision, par exemple, a quelque peu progressé dans l'intervalle, ce n'est pas à cela qu'elle le doit (cf. Marr, Ullman, Koenderinck). Si enfin l'on se préoccupe de compréhension de textes, il semble bien que dans toute situation un tant soit peu complexe, le nombre

de schémas nécessaires, qui ne sont pas dynamiquement assemblés par le système mais fournis par le modéliseur avant de commencer, doit être “infini”; cela compliquera d'autant le problème d'un choix pertinent de candidats pour une comparaison avec les occurrences à venir ; et le problème de la sensibilité du résultat au contexte n'en sera pas résolu pour autant.

Peut-être faut-il finalement admettre, écrit Dreyfus, que les aptitudes, les savoir-faire (*skills*) qui nous permettent d'interagir avec les objets et les personnes sont “des aptitudes que nous développons par la pratique, sans jamais avoir à nous représenter notre corps à la façon d'un objet, ni notre culture comme un ensemble de croyances, ni nos inclinations (*propensities*) comme des règles de la forme situation–action.” (1981, p. 198). Les modélisations de l'IA ne pourraient donc, selon lui, que déboucher sur une régression à l'infini. Parmi les aspects fondamentaux de ce que nous pourrions appeler une *situation cognitive*<sup>19</sup>, qui avaient jusque là totalement échappé aux systèmes symboliques stricts, et le resteraient peut-être toujours, il recensait en particulier les traits suivants (1979, trad. fr. 1984) :

- Le travail de la conscience marginale qui tisse en permanence une toile de fond sur laquelle se distinguent les formes méthodiquement scrutées et recensées par la conscience attentive.

- Le regroupement permanent des données en formes catégorisées remarquablement résistantes aux distorsions, au mouvement, au bruit de fond, aux dégradations de toutes sortes.

- L'élimination “directe” des ambiguïtés, sans recours apparent à des “descriptions” détaillées (c'est-à-dire des données stockées ou inférées dans un format de type propositionnel) : cela vaut particulièrement pour la

---

<sup>19</sup> Il est à noter que Dreyfus employait à l'époque le terme *cognitif* dans un sens restrictif et péjoratif, solidaire d'un mentalisme computationnel strict, qui n'est pas le sens neutre et ouvert que nous voudrions lui donner ici. Par ailleurs, l'expression *situation cognitive* me vient d'un exposé donné par J. Schlanger en 1987 au CREA.

compréhension des règles d'actions, toujours incomplètement spécifiées<sup>20</sup>. Cette élimination directe de l'ambiguïté n'empêche pas, le cas échéant, sa prise en considération dans un deuxième temps, après détection d'une "incohérence" ou "impossibilité".

- La distinction changeante et graduelle entre essentiel et accessoire, avant-plan et arrière-plan<sup>21</sup>, qui semble liée à une appréhension *globale* de la situation, fondée sur un sentiment de ressemblance avec toute une famille d'autres situations. Cette ressemblance semble impossible à décomposer et définir avec précision en termes de traits communs distinctifs ; leur analyse logique ou componentielle, qui certes joue toujours dans l'après-coup un rôle important, ne peut donc se confondre avec une analyse causale de ce qui a produit l'impression de ressemblance.

Sans crainte d'être contredit par son auteur, on pourrait ajouter à cette liste une mention du caractère fondamentalement *gestaltique* des configurations significatives : cela implique à la fois l'existence d'une structuration en parties qui ne soient pas simplement juxtaposées mais fonctionnellement reliées, une conscience de cette structuration qui soit variable et adaptable, la possibilité d'une saisie du tout plus facile et immédiate que celles de ses parties à l'état dispersé, la dépendance des parties relativement à leur réunion dans le tout qui les soude (c'est-à-dire l'antécédance éventuelle de la relation sur les *relata*), et enfin une compositionnalité faible qui ne peut s'analyser uniquement en termes de composants (car l'assemblage ajoute ou retranche quelque chose qu'on ne peut caractériser comme *partie* manquante ou excédentaire, et qu'il faut donc décrire d'une autre façon).

---

<sup>20</sup> Dans le point de vue classique, c'est aux règles dites *pragmatiques* de fournir les compléments nécessaires. Avec les problèmes habituels de régression à l'infini.

<sup>21</sup> Le réglage de cette distinction, dans certaines approches propositionnelles du sens, est dévolu aux mécanismes de la *pertinence*. Même remarque que précédemment sur la régression à l'infini.

On peut évidemment douter, comme l'ont fait Dreyfus, Winograd et Flores, ou Suchman, que les systèmes symboliques stricts de l'IA classique soient jamais capables de reproduire *par eux-mêmes* la variété et la complexité des phénomènes cognitifs que nous venons d'évoquer. Dimensions disparates des savoir-faire et des schèmes de représentations, catégorisation et typicalité, adaptation, schématisation et iconicité des "espaces mentaux", inhérence des aspects sensori-moteurs, conscience marginale et mémoire associative, construction gestaltique des structures significatives, (in)cohérences et analogies, arrière-plan culturel, contexte relationnel... : la liste est longue, des aspects de la cognition *tous décisifs*, que l'IA classique a bon gré mal gré contribué à souligner, en les déniaient bien souvent ou en prétendant comiquement les contrôler (l'année prochaine). *Mutatis mutandis*, des remarques analogues s'appliquent aux approches connexionnistes, parfois présentées à leur tour comme des panacées. Certes, l'importance de ces modèles dans le renouvellement des problématiques cognitives n'est pas contestable. Ils font entrevoir de nouveaux types de structures (non finitistes/combinatoires), permettent d'envisager véritablement la compatibilité de modalités cognitives distinctes (langage et perception), permettent de penser autrement la relation entre différents niveaux d'analyse (émergence et non plus compilation), etc. A travers eux, l'IA renouvelle son rapport au continu et à la mathématisation, et notamment à la théorie physico-mathématique des systèmes dynamiques. Leur rôle théorique est donc fondamental. Concernant leur importance pratique, il convient d'être plus nuancé. On estime parfois que sur certains des points cités plus haut, les modèles connexionnistes, directement appliqués au domaine cognitif, peuvent apporter une souplesse et une robustesse de fonctionnement supérieures, jointes à des capacités "d'apprentissage" par conditionnement qui pourraient bien leur faire jouer un rôle complémentaire de celui tenu par les systèmes classiques. Mais on sait aussi les problèmes qu'ils rencontrent à tenir convenablement ce rôle, dès que l'on s'écarte des quelques

domaines où ils sont véritablement opérationnels (comme la reconnaissance des formes ou la modélisation de schèmes sensori-moteurs). Leur intervention est encore limitée à des cas beaucoup trop simples, avec peu de structure inscriptible dans les représentations, et des processus souvent réduits à des mécanismes d'associations qui n'exploitent pas suffisamment cette structure lorsqu'elle existe. Les algorithmes "d'apprentissage" sont loin de garantir dans la pratique les résultats que leur description théorique fait espérer, et par ailleurs le problème de l'explosion combinatoire les concerne tout autant que les approches classiques. Enfin, le statut du *microfonctionnalisme* associé à ce type de modélisation cognitive reste bien problématique<sup>22</sup>. Dans une perspective raisonnablement optimiste, on peut envisager cependant que les réseaux connexionnistes permettent de modéliser des heuristiques de classifications ou de corrélations analogiques jouant sur un assez grand nombre d'observations ou de cas traités antérieurement par le système lui-même ; ils offriraient ainsi des suggestions qu'un traitement symbolique strict serait ensuite chargé de raffiner, valider, reconstruire en termes plus explicites. Ce serait en ce cas une partie non délibérative du raisonnement qui serait prise en charge par des réseaux. Reste évidemment à imaginer une traduction cohérente vers les systèmes classiques (ou à partir d'eux) permettant de construire des systèmes hybrides cohérents. Les représentations connexionnistes ont des propriétés qui les destinent plutôt à des traitements fondés sur des régularités statistiques ou sur des répartitions géométriques dans l'espace où elles prennent valeur ; leurs occurrences ont toujours un caractère approximatif, et appellent donc un traitement qui s'en accommode. Tous ces caractères semblent donc les vouer toujours davantage à l'univers connexionniste. A cela s'ajoute le fait

---

<sup>22</sup> Quelques références (parmi la foisonnante littérature disponible) sur les problèmes de la modélisation connexionniste dans les domaines cognitifs : Rumelhart et McClelland 86, Smolensky 88 et 89, Victorri 88 et 89, Andler 90, Petitot 90, des numéros spéciaux d'*Intellectica* et *Artificial Intelligence* (1990).

que ces représentations proviennent d'une "histoire", à savoir de la confrontation à un échantillonnage, voire à toutes les données déjà rencontrées : elles sont donc susceptibles de modifications que les systèmes classiques qui leur seraient liés devraient intégrer ; sans parler ici des configurations intempestives, non désirées par le modélisateur, que les algorithmes d'apprentissage ne permettent pas d'éliminer, et dont on ne saura que faire. On fera donc bien d'écarter une fois de plus les affirmations dithyrambiques, qui n'ont pas plus de valeur dans ce cas – mais qui sait ? – qu'elles n'en avaient auparavant à propos de l'apprentissage symbolique. Le rôle des modules à apprentissage *dans les systèmes à base de connaissances* serait plutôt, pour le moment, de faire tout simplement de l'analyse de données, ce qui implique naturellement révision et interprétation par des opérateurs humains.

On peut mettre également en question l'idée que les systèmes de l'IA classique puissent, non pas tant reproduire dans leur fonctionnement, mais seulement *évoquer* la complexité d'un comportement humain, tout en conservant des ressemblances significatives avec des processus psychologiques. Comme l'expliquent J.-P. Krivine et J.-M. David<sup>23</sup>, la recherche en milieu industriel, qui est pourtant authentiquement concernée par le problème du transfert d'expertise, n'y voit déjà plus une affaire de transcription, mais de "modélisation qualitative". On pourrait même dire de "modélisation" tout court, tant le changement de point de vue exprimé par Krivine et David se confond avec une réflexion épistémologique sur le rôle des modèles dans les sciences (nous reviendrons sur ce point en conclusion). Si l'on considère l'histoire récente de l'IA, on ne peut que noter, en tout cas, à l'intérieur des mêmes domaines, parmi les mêmes groupes de chercheurs, la juxtaposition d'approches hétérogènes de la modélisation, qui ne sont pas toujours distinguées comme elles pourraient l'être. Un cas exemplaire en est fourni par les travaux de ce qu'on pourrait

---

<sup>23</sup> Voir leur article dans ce numéro.

appeler l'école de J. de Kleer. Dans son approche du diagnostic de pannes, elle fait par exemple appel à de grandes catégorisations des modes de raisonnements, non dénués de pertinence psychologique : *case based*, *symptom based*, *model based*, *consistency based*, *explanation based*, etc. Mais elle organise aussi les opérations du diagnostic et les contrôles de cohérence à partir de considérations de la théorie *purement logique* des modèles (*alibis* de O. Raiman, par exemple). Elle fait appel à des modélisations de domaines physiques, tant numériques (par équations différentielles), que qualitatifs (par ordres de grandeurs) ou structurels/fonctionnels (au niveau de la machine instanciée par le dispositif physique). Ces descriptions concernent évidemment des niveaux d'organisation distincts, ce qui est parfaitement légitime, mais elles sont également de styles hétérogènes : certaines relèveraient plutôt d'une physique ou d'une morphologie spontanée, dite "naïve" ; d'autres sont similaires aux représentations explicites produites par une nouvelle branche de la technoscience qui aurait précisément pour tâche de concevoir le *General Diagnostic Engine* dont il est question<sup>24</sup>. D'où en fait le balancement constant de ces travaux entre la reconstitution de "modèles mentaux" dans un style psychologique, qui n'est que rarement leur idéal, et la modélisation réfléchie, techniquement très élaborée, des domaines de réalité sur lesquels le système doit raisonner : le risque est alors d'en arriver à des modèles appliquant une physique que l'on voudrait "naïve" et conforme au sens commun dans chacun de ses éléments de base, mais où l'on autoriserait en même temps des synthèses d'une cohérence et d'une exhaustivité qui n'appartiennent, du moins en principe, qu'aux théories scientifiques. On aurait ainsi confondu la physique *naïve* (au sens de Hayes) et la physique *qualitative* (qui relèverait des modes d'objectivation

---

<sup>24</sup> Cette branche de la technoscience n'aurait donc pas pour tâche de reproduire au plus près une activité préexistant à l'informatique, mais plutôt de redéfinir un travail dans un environnement encore inédit. Les considérations psychologiques ne joueraient donc *a priori* qu'à travers les préoccupations ergonomiques.

propres à la science). Confusion bien évidemment liée aux présupposés de l'IA, qui identifie les modalités cognitives de la compréhension à la possession d'une "théorie du domaine", et doit donc supposer, sauf à accepter une réorientation fondamentale, que le sens commun est lui-même structuré à la façon d'une théorie (cf. H. et S. Dreyfus, 1988, p. 318 et 326).

Enfin, dans la mesure où la modélisation dite cognitive du savoir de l'expert n'a pas pour but de respecter tous ses idiomatismes, on se trouve également confronté au problème de normaliser le modèle construit dans tous ses aspects : ontologie (taxinomies, actinomies, structures casuelles, temporelles...), modalités (intentions, croyances), etc. La question de l'universalité des structures cognitives fait retour ici, avec la multiplicité des experts et des utilisateurs. On peut ainsi se demander – mais ce n'est qu'un exemple – si la "logique mentale" des relations méréologiques maniées par l'expert ne risque pas d'exiger quelques rectifications. De même pour la décomposition de certains processus, trop enchevêtrée pour être acceptable. Ces problèmes de reconstitution et d'interconnexion des savoirs experts ouvrent nécessairement sur d'autres dimensions que le souci des réalités psychologiques individuelles — ce qui amène notre deuxième point.

## ***II. Les contraintes de la reconstruction logique.***

Que l'on croie ou non à la fidélité psychologique des modèles symboliques "stricts" de l'expertise, ou qu'on les considère plutôt comme des reconstructions, des artefacts, le problème se pose de toute façon d'élucider et contrôler leur fonctionnement en termes *formels* cette fois. En effet, plus les systèmes offrent de facilités à l'expression des connaissances (c'est-à-dire aux symbolisations construites par l'expert avec l'aide du cognicien), en permettant d'amalgamer dans la même structure des conditions portant sur le domaine étudié, sur l'organisation des tâches de raisonnement, ou sur l'état du processus d'inférence au moment du

déclenchement éventuel, plus leur fonctionnement devient logiquement imprévisible. Or une relative transparence logique des systèmes semble indispensable, dans la mesure (et seulement dans cette mesure) où les seuls critères opérationnels de cohérence, de validité, et de complétude dont nous disposons actuellement sont en fait des critères logiques. Nous prendrons ici *logique* dans un sens relativement restreint, comprenant la donnée d'une syntaxe de logique des prédicats avec opérateurs modaux éventuels, une sémantique dénotationnelle en termes de mondes possibles, et une équivalence stricte entre la validité sémantique et la déductibilité selon certains schémas d'inférence qui restent à préciser. Un sens bien classique, en somme, mais étendu pour y adjoindre les modèles que Kripke a proposés pour les logiques modales ou intuitionnistes, et qui forment toujours la base d'un grand nombre de travaux<sup>25</sup>. Notre thèse est que, sans le maintien d'un minimum de conformité à ces notions capitales, les systèmes symboliques de l'IA classique ne peuvent croître en complexité, et encore moins se connecter les uns aux autres. Nous voudrions maintenant préciser et argumenter cette thèse, sans exclure d'y apporter éventuellement quelques nuances.

À partir du moment, en effet, où les processus modélisés donnent lieu à des inférences longues, complexes, déjà elles-mêmes réduites à des enchaînements symboliques stricts, les formes de modélisations que nous appelons *reconstructions logiques* semblent indispensables<sup>26</sup>. Elles nous

---

<sup>25</sup> On pourra consulter les recueils parus chez Morgan Kaufmann sous le titre générique *Theoretical Aspects of Reasoning about Knowledge*, et les travaux de Y. Shoham parus dans la revue *Artificial Intelligence* ces dernières années. Voir aussi la thèse de L. Catach sur *Les logiques multimodales* (1989), .

<sup>26</sup> Soulignons qu'il n'y a pas de lien tautologique entre inférence et reconstruction logique, à moins bien sûr de présupposer : (a) une conception finitiste-combinatoire des structures symboliques, (b) une réduction exhaustive du processus à des manipulations de symboles, et (c) une identification de ces manipulations à des transformations par règles syntaxiques. Mais si l'on convient simplement d'appeler *inférence* tout processus

permettent, avec l'aide de prolongements externes à notre mémoire (papier, crayon...), d'analyser des processus dont nous ne pouvons contrôler la marche en nous appuyant uniquement sur nos synthèses "spontanées". Elles se laissent parfaitement réifier, à des niveaux de précision réglables (changements de notations), sous forme d'écritures qui instrumentalisent pour nous ce type particulier d'objectivité qu'est l'objectivité logico-formelle. À un autre niveau, nous avons également besoin d'une analyse logique des modèles cognitifs adoptés (ou plutôt des systèmes symboliques qui en tiennent lieu), qui nous permette d'anticiper sur les effets de leur mise en oeuvre. Soulignons qu'il ne s'agit pas seulement ici des problèmes posés par une implantation informatique, mais plus fondamentalement d'un problème de compréhension, d'évaluation de nos modèles de la connaissance. Et à cet égard les logiques modales, si caricaturales soient-elles, demeurent des points de repère importants. Elles sont d'ailleurs en progrès constant, et contribuent encore à quadriller, à baliser le potentiel véritable des systèmes réalisés.

Certains auteurs, et non des moindres<sup>27</sup>, ont suggéré que les structures symboliques propres à l'IA, par opposition aux formalismes construits dans le style "logique", permettraient, ou permettraient de contourner ces problèmes, et de développer indéfiniment des systèmes dans une sorte d'au-delà des contraintes logiques, c'est-à-dire (croit-on pouvoir conclure) par simple conformité aux processus psychologiques. On peut évidemment se demander comment cette conformité pourrait être garantie, et à propos de quoi. On ne voit de toute façon que l'introspection et le débat public pour remplir ce rôle, si l'on a en tête de constituer une base de connaissances encyclopédiques. Le projet CYC de R. Guha et D. Lenat, par exemple, met à contribution les productions de différents *éditeurs de connaissance*, et prétend aller vers la mise au point de procédures de

---

mettant en jeu des entités symboliques (par ex. une association lexicale), on peut facilement imaginer d'autres options...

<sup>27</sup> Cf. Schank et Rieger (1974) et Minsky (1981).

transfert analogique, permettant de les connecter entre elles au-delà de ce qui a été explicitement prévu par chaque rédacteur (Guha et Lenat, 1989). C'est donc typiquement un projet de représentation du sens commun, où l'on ne s'embarrasse pas, apparemment, de méthodologies précautionneuses analogues à celles du transfert d'expertise.

On comprendra que nous préférerions attendre de voir avant de croire. Il ne manquera sûrement pas alors de moyens de reconnaître la véritable utilité de ce qui aura été fait, et n'aura évidemment pas atteint les buts que la rhétorique coutumière des grands projets de la technoscience contraint parfois à déclarer. L'IA vit de ces retombées.

D'ici là nous nous contenterons de constater qu'à chaque fois qu'une structure ou un processus d'inférence "non standard" (mais néanmoins très classiquement symbolique) a été proposé pour faire droit à un aspect du raisonnement humain effectivement ignoré par les tenants d'une approche strictement logiciste, seule une reprise dans un cadre logique (convenablement étendu) a pu empêcher la trouvaille de retomber dans la trivialité ou le mystère. On peut citer ici tout ce qui tourne autour du raisonnement par défaut et des phénomènes de non-monotonie : la logique de la circonscription, l'inférence de la négation par échec fini ou la complétion des programmes en Prolog<sup>28</sup>. On peut mentionner également l'explicitation de la logique triviale des *frames* (Hayes 79), la conception logique de la programmation par contraintes (Jaffar et Lassez, 86 et 87), ou même les réseaux sémantiques que personne n'arrive vraiment à utiliser autrement que comme une variante notationnelle de la logique des prédicats. Dans chacun de ces cas, seule l'étude logique a permis de généraliser et systématiser certains procédés, voire de les suggérer. Ce qui a échappé à la reprise logique, comme on peut le penser jusqu'à un certain point à propos des langages objets, relève plutôt de la méthodologie de programmation ou de la mise au point des outils, c'est-à-dire d'un tout

---

<sup>28</sup> Pour une introduction, voir le livre de Genesereth et Nilsson (1987).

autre niveau d'analyse<sup>29</sup>. Ou bien alors c'est qu'il est question de bricolage privé, de l'adaptation "individuelle" des systèmes – qui par ailleurs reste absolument indispensable au fonctionnement des réalisations.

Pour radicaliser encore cette thèse quelque peu provocatrice, on peut ajouter que les meilleures analyses proposées par les chercheurs d'orientation formaliste en IA sont en fait celles qui restent les plus proches du cadre logique classique, mais en y incorporant évidemment toute la gamme des modalités : on doit citer ici les travaux de logique multimodale de Levesque, Halpern, Moses, Fagin, Vardi, Lehman, Shoham, Delgrande et bien d'autres, qui concernent des sujets aussi variés que la logique des défauts, les rapports entre connaissance et croyance, l'incohérence, le savoir implicite, la causalité, le temps... Autrement dit, quitte à formaliser, autant rester ou revenir dès que possible dans le dispositif légué par Tarski et enrichi par Kripke, qui à ce jour reste inévitable, et toujours le plus clair, le plus facilement transmissible<sup>30</sup>.

En somme, et compte tenu de ce que nous avons dit plus haut, le développement des systèmes symboliques "classiques" et celui de la logique seraient inéluctablement solidaires : ce que l'on peut interpréter, au choix, comme une célébration de la logique en tant qu'instrument d'élucidation, ou comme la constatation désolante que les systèmes symboliques de l'IA n'ont nullement réussi à déborder les possibilités du logicisme dont ils sont issus, et auquel ils prétendaient échapper. L'analyse logique a en tout cas un effet décapant : elle certifie ce qu'il y a de cohérent dans l'informatisation d'un modèle cognitif, et dissipe les illusions de

---

<sup>29</sup> Mais il faut nuancer cette affirmation (cf. *infra*).

<sup>30</sup> En attendant peut-être une mise en oeuvre du concept logico-mathématique de *topos* (Goldblatt 86 ou Lambek et Scott 86 sont ici des ouvrages de référence ; à noter aussi, dans le même ordre d'idées, le cours introductif de J. Bénabou sur les *Ensembles empiriques*, 1988). Pour ce qui concerne les autres noms mentionnés dans le texte, on trouvera quelques références dans la bibliographie. Shoham, par exemple, a écrit un article très éclairant sur la signification et l'utilité des logiques non monotones, en les rapportant justement à une sémantique de mondes possibles.

subtilité qui peuvent s'attacher temporairement à certains concepts de l'IA, lorsqu'on les présente comme les corrélats adéquats de certaines descriptions ou données psychologiques (typicalité et prototypes, par exemple).

Nuançons immédiatement ce qui précède. Certes, le cadre de référence logique reste une garantie très importante du caractère non insensé et suffisamment productif des systèmes symboliques. Mais cela n'implique nullement une priorité méthodologique : c'est bien souvent en dehors (et sans souci) des traditions logiciennes que les innovations intéressantes ont lieu en IA. Et de plus, quand une élucidation logique est proposée pour ces innovations, elle ne vient pas les éliminer : non seulement parce qu'en général l'élucidation est seulement partielle (elle n'est qu'un schéma régulateur, une idéalisation qui nous aide à contrôler le développement des systèmes sans le contraindre absolument), mais surtout parce que les structures symboliques qui ne sont pas exactement “logiques” de prime abord possèdent de remarquables propriétés “ergonomiques”, qui en font, dans le contexte de notre interaction avec les ordinateurs, d'indispensables *variantes notationnelles* des formalismes logiques censés après coup en fonder l'utilisation. Ainsi J. Sowa présente-t-il ses *graphes conceptuels* comme une variante notationnelle rigoureuse de logiques typées, remarquablement bien adaptée à un certain nombre d'intuitions sémantiques (taxinomies et structures casuelles) et susceptible de recevoir des annotations de divers niveaux (lexical, syntaxique) très utiles dans une perspective de traitement automatique de données linguistiques (1984). Cet exemple est évidemment très favorable à la thèse défendue ici, et il faut souligner d'un autre côté que la plupart des formalismes logiques “justificateurs” ne jouent aucun rôle opérationnel explicite dans la réalisation des systèmes. Ils en fournissent seulement une clarification nécessaire à leur développement, tout au moins au delà d'un certain seuil de complexité – ne serait-ce qu'en permettant de les classer, comme J. Fargues l'avait naguère suggéré à propos des moteurs d'inférence (1985).

D'une façon semblable, les structures des langages *objets* ou *acteurs*<sup>31</sup> ont certainement des qualités “iconiques” qui facilite dans bien des cas la modélisation. Cependant, l'information qu'elles contiennent est pour l'essentiel reformulable dans un cadre logique, sans modification substantielle du cadre épistémologique et en restant *grosso modo* au même niveau d'analyse<sup>32</sup>. En quelque sorte, la complexité de ces structures *objets*, pour autant qu'elle soit contrôlable, n'a pas atteint un niveau tel qu'elle exige une véritable mathématisation pour être comprise. Une logicisation directe y suffit. L'écart, bien réel pourtant, qui les sépare des formalismes logiques ne permet pas d'introduire un niveau explicatif propre pour les domaines modélisés. Le supplément d'analogie offert se comprend plutôt dans le cadre d'une *ergonomie des représentations*.

---

<sup>31</sup> Envisagés ici comme langages de représentation des connaissances, et pas seulement comme langages de programmation. Cf. J. Ferber 1987a, 1987b.

<sup>32</sup> L'affirmation peut sembler sommaire, et exigerait en effet d'être mieux argumentée. Deux remarques brèves (inspirées par les objections d'un des rapporteurs de cet article, que je remercie d'avoir soulevé ce point). Du point de vue de la *théorie du calcul*, tout d'abord, où rivalisent en effet les formalismes applicatifs et prédicatifs, les représentations inspirées du lambda-calcul (comme le sont dans une certaine mesure les langages objets), sont évidemment bien adaptées à la représentation des *opérations* des symboles les uns sur les autres. Mais elles se révèlent particulièrement opaques dès qu'il s'agit de leur attribuer un *contenu* étranger à l'opérativité symbolique : les grammaires catégorielles, par exemple, font appel au formalisme des combinateurs pour décrire la compositionnalité syntaxico-sémantique, mais les archétypes cognitifs qu'elles permettent de calculer sont, à défaut de mieux, exprimés par des structures prédictives. Newell ne disait pas autre chose : le niveau opératif des systèmes symboliques classiques est “lispien”, tandis que son interprétation en termes de connaissance est principalement logique (cf. également Birnbaum, 1991, p. 60). Si l'on considère maintenant l'opposition *théorie du raisonnement vs modèles*, notre thèse serait qu'une approche *finitiste* des objets modélisés, pratiquement la seule pratiquée en IA, ne permet pas de faire apparaître une “véritable” distinction entre les pôles hétérogènes du langage et du modèle : c'est précisément la raison pour laquelle les langages *objets*, en dépit de leur nom équivoque, sont si facilement convertibles en structures logiques “discursives”. Et c'est aussi la raison pour laquelle les tentatives d'introduire un peu de schématicité ou d'iconicité en IA à travers les structures de *frames* ont échoué pour l'essentiel (cf. le langage KRL et les désillusions de Winograd).

La position adoptée ici n'est donc nullement logiciste, en quelque sens que ce soit et encore moins en ce sens fort qu'on pourrait qualifier "d'éliminatif" – bien au contraire. Elle consiste simplement à dire que les systèmes symboliques de l'IA classique ne peuvent se développer, même en tant que prétendue simulation cognitive, si ce n'est en prenant *aussi* appui sur autre chose que les exigences de ladite modélisation cognitive ; et cette autre chose, *pour le moment*, s'identifie avec un dispositif logique plus ou moins classique (tout en incluant des modalités, comme nous l'avons déjà précisé). Rien d'étonnant à cela, si l'on admet que les systèmes logiques *définissent* véritablement, à travers les dénnotations que nous leur prêtons ou les instanciations physiques que nous en réalisons, ce que nous comprenons sous le concept d'un monde finiment construit à partir de *faits atomiques indépendants*, tels que les décrivent les langages de représentation de l'IA. Il s'agit donc d'une fatalité inhérente à ces approches : quoi qu'on fasse dans ce cadre, les relations des objets entre eux resteront finies et formelles, c'est-à-dire analogues et en fin de compte réductibles à la relation qu'un prédicat entretient avec ses arguments ou avec d'autres prédicats, tous donnés en nombre fini. Est-ce là le signe d'un grave déficit ? À notre sens, oui, et l'on peut sur ce point rappeler très brièvement deux arguments, le premier portant sur la différence entre logicisation et mathématisation, le second sur la confusion entre le rôle descriptif et le rôle causal des structures logiques.

Comme toutes les descriptions scientifiques, celles qui sont étroitement associées aux formes logiques reposent sur une délimitation assez précise de leurs conditions de validité, un choix rigoureux du niveau d'abstraction ou d'idéalisation, une division du domaine en régions séparées que la pratique reconnecte tant bien que mal, etc. C'est en ce sens qu'on peut les dire "sans équivoque" : la variété de leurs dénnotations formelles est suffisamment maîtrisée, pour autant qu'elle importe ici, et leur référence pratique, concrète correspond à une objectivité bien cernée, sur laquelle on

n'opère pas de n'importe quelle façon. Ces traits constitutifs de la modélisation scientifique sont cependant caricaturés par les approches logicistes : elles visent en effet des entités singulières qu'elles font entrer directement dans un dispositif d'analyse finitiste, sans les plonger au préalable dans un univers mathématique (possiblement fictif) d'entités infinies ou même continues, qui offrirait le “champ” nécessaire à d'autres approches de la relation, de l'identité, de la stabilité (rapports tout-partie, type-occurrence, etc.) et à partir de là, par exemple, une meilleure compréhension du vague, de l'équivoque, de l'imprécision caractéristique des formes significatives. Un développement complet de cet argument n'a pas sa place ici, mais nous pouvons au moins citer un texte très éclairant sur cette question. Dans un compte-rendu critique d'un livre récent sur les théories de la Gestalt, publié sous la direction de B. Smith en 1988, J.-M. Salanskis a en effet parfaitement exprimé ce privilège de la *mathématisation* dans la démarche scientifique moderne (qui implique généralement la théorie des ensembles), en la reliant à ce qu'il appelle le *décrochage de type* :

... les grandes démarches novatrices récentes de la théorisation mathématique du monde relèvent essentiellement de ce “point de vue ensembliste”, et en l'espèce d'une possibilité que ce point de vue est seul à apporter, possibilité que nous appellerions volontiers celle du “décrochage de type”. Le “décrochage de type” consiste à poser, dans une approche mathématique, l'étant singulier que l'on vise comme non singulier, comme divers, complexe, relationnel à l'intérieur de soi. Il ne faut pas confondre cette démarche avec celle du réductionnisme, qui quant à lui change la visée du discours, dénie la singularité de ce qui était d'abord interrogé pour mettre en avant une autre singularité, plus vraie dès qu'il existe un accès expérimental à son autonomie “au sein” du premier singulier. Ici, le singulier visé restera l'objet de la visée, mais sera modélisé par une entité mathématiquement non singulière, mathématiquement riche, sans référence à quelque autre singularité sur le versant ontique : la théorie des ensembles nous permet d'envisager le *cela* que nous théorisons comme “fictivement” composé ou “fictivement structuré”, le “fictivement” ayant ici deux valeurs possibles, puisqu'il peut signifier simplement que la trame ensembliste associée à l'entité est *factuellement* dépourvue de référence, ou bien qu'elle est *constitutivement* dépourvue de référence (dans ce cas elle “n'existe” qu'en raison des

présuppositions idéalisantes-infinitisantes de la théorie des ensembles) (1991, p. 532).

J.-M. Salanskis passe ensuite en revue quelques exemples de mathématisation avec décrochages de type, parmi lesquels figurent notamment les approches catastrophistes ou connexionnistes de la structure. Il souligne également que la représentation mathématique du continu repose elle-même sur un pareil décrochage de type, *qui se trouve donc impliqué chaque fois qu'il est fait appel au continu mathématique dans une modélisation*. Le lecteur intéressé pourra se reporter à l'article de Salanskis pour plus de détails sur ces points fondamentaux.

Passons maintenant à notre deuxième argument. Comme nous le savons (presque) tous, les contraintes induites par les dispositifs logiques entrent bien souvent en conflit avec certains traits particulièrement évidents de la phénoménologie de telle ou telle "tâche" cognitive. Faut-il d'ailleurs rappeler ici l'antipsychologisme notoire des pères fondateurs de la logique moderne, à commencer par celui de Frege, ou bien encore celui des auteurs du Cercle de Vienne ? La logique était pour eux l'instrument privilégié de la reformulation des contenus de la science, voire de la réforme ou de la refondation des significations foncièrement vagues véhiculées par les langues naturelles. C'était la forme réifiable et rigoureuse exigée par une herméneutique qui se voulait strictement rationnelle, à la fois littérale et complètement objectivable. Elle devait être décelée dans tous les discours chargés d'un sens attestable, non illusoire. Il n'était donc évidemment pas question de borner *a priori* la complexité et l'étendue de ses manifestations. Cependant, et pour autant que nous le sachions, la forme logique ne représentait nullement pour ces auteurs ce qu'elle est pour les partisans du *réalisme* des approches logicistes en matière de cognition : elle ne constituait pas un mode de description véridique des dynamiques mentales et sociales effectives, qui seules permettent l'émergence intersubjective du niveau logique et sont, d'un point de vue

anthropologique, réellement responsables de ses effets pratiques<sup>33</sup>. En d'autres termes, la valeur herméneutique que nous attribuons aux structures logiques (par exemple dans l'analyse des langues naturelles) peut fort bien reposer sur les qualités particulières, non logiques, des “machines à interpréter” que nous sommes. Par ailleurs, à supposer même que l'on accepte de réduire une part significative de notre monde phénoménologique à l'objectivité logico-formelle, rien n'autoriserait à confondre le produit et les instruments de cette réduction avec une structure causale sous-jacente (qu'il faudrait alors concevoir comme une factorisation de la causalité physique à travers une détermination de type syntaxique). A titre de cas-limite, on peut citer ici le cas du philosophe J. Searle, qui fait un certain usage descriptif (voire normatif) des formalismes logiques, sans reconnaître pour autant l'existence d'un niveau représentationnel autonome, de type syntaxique, qui constituerait l'essentiel des processus cognitifs.

Sans prétendre clore le débat, nous pouvons maintenant résumer dans ses grandes lignes l'argumentation présentée dans cette section. On a d'abord cherché à montrer que les systèmes symboliques de l'IA classique relèvent principalement d'une approche logique du cognitif, qu'ils ne peuvent s'en écarter sous peine d'y perdre toute cohérence, et que leur diversification incessante ne s'interprète pas comme l'apparition d'un univers syntaxique “au-delà de la logique”, mais plutôt comme un progrès dans l'ergonomie de la “mise en logique”. On a ensuite distingué entre logicisation et mathématisation, cette dernière étant seule à offrir un répertoire de formes assez riche, en même temps qu'un accès au monde physique et une connexion possible aux théories contemporaines de la perception. On a également reconnu la valeur descriptive des formalismes logiques, en rappelant qu'elle procède d'une tradition antipsychologiste, et

---

<sup>33</sup> Sur l'étrangeté de l'empirisme revendiqué sous l'appellation d'*empirisme logique*, cf. notamment la conclusion du livre de J. Proust, *Questions de forme* (1986, pp. 406 sq.).

en la distinguant pour finir du rôle réducteur que lui fait jouer l'approche syntaxique des *processus* cognitifs<sup>34</sup> Savoir quelles sortes d'intentions ou d'objets peuvent en dépit de cela être utilement spécifiés, voire contrôlés, à travers ce type d'inscription à l'intérieur de la "région logico-symbolique", et de quelle relative autonomie causale ils pourront alors bénéficier, définit bien le problème du sens et de l'utilité des systèmes de l'IA classique, qui sont donc pour ainsi dire en charge du développement de cette région (ou archipel) du logico-symbolique, sous le double aspect du mécanisme et de la facilitation.

Revenant alors aux systèmes à base de connaissances, on peut se demander, à titre d'illustration triviale de ce qui précède, si l'on fait une si bonne affaire que cela en amalgamant description du domaine et métalangage de contrôle, comme nous y invitent parfois les argumentaires de vente des générateurs de système : l'opacité logique augmente, et on perd peut-être, *sous l'influence de l'expert humain*, une occasion d'énoncer les connaissances sous les formes les plus modulaires et déclaratives possibles, qui conviennent mieux à la complexification des systèmes symboliques artificiels. Des remarques analogues pourraient être faites à propos des diverses et inextricables conceptions des domaines étudiés (les modélisations qualitatives d'un dispositif physique, par exemple, n'émergent pas si naturellement que cela du modèle nomologique par équations différentielles, et ne se correspondent pas toujours clairement entre elles). On peut en effet se demander jusqu'à quel point il faut les représenter par des modèles disjoints, possédant chacun ses propres structures et processus : c'est évidemment une contrainte de leur analyse logique, mais une fois séparés, il est bien difficile de les reconnecter de façon appropriée. Il est clair qu'il y a là un dilemme : rester proche des verbalisations de l'expert, et perdre progressivement tout contrôle sur les

---

<sup>34</sup> L'émergence, la construction ou la traduction des "structures" sont évidemment de tels "processus"...

structures construites, ou bien clarifier l'architecture logique d'un système dans lequel on aura de plus en plus de difficulté à intégrer les connaissances initialement recherchées. Et ce n'est pas se hasarder beaucoup que de l'affirmer : dans un conflit de ce type, les contraintes constitutives de l'instrument ont de bonnes chances de l'emporter.

Peut-être faut-il renoncer dans la plupart des cas, non seulement à une automatisation complète des raisonnements, mais aussi à ce que cette automatisation, si partielle soit-elle, se fasse sur la base exclusive de connaissances ou de démarches *préexistantes à leur nouveau contexte d'emploi* : le savoir expert ne peut être investi qu'après avoir subi une profonde recomposition, et reçu le renfort de connaissances inédites, spécialement conçues pour assister une modélisation à support informatique. Nous devons envisager alors les systèmes comme des "supports pour la résolution de problèmes", c'est-à-dire des outils limités et non pas des prothèses parfaitement substituables à une compétence absente. Dans ce cas, le rôle de l'utilisateur-opérateur ne pourrait plus se limiter à la fourniture d'une requête et de quelques données en cours de diagnostic ; un apprentissage spécifique, une bonne pré-compréhension du domaine seraient alors *reconnues comme nécessaires*. Il s'agirait en effet de l'interaction d'un système et d'un sujet *déjà relativement compétent*, jouant avec des modèles du domaine et des reconstructions logiques de ses propres activités ou raisonnements, souvent différentes de leur cours naturel. La conception des interfaces en serait notablement transformée, la résolution de problèmes *dépendant* alors de la réussite d'un processus de communication élaboré. Mais cela nous amène à notre troisième point.

### ***III. Les contraintes de la communicabilité et de l'interaction en situation réelle.***

Quelles que soient les fonctions dévolues au système, elles comportent au moins celle d'accroître, d'une façon ou d'une autre, les connaissances de l'utilisateur. Si l'on admet que les seules connaissances véritablement communicables et utilisables sont celles qui sont munies de *justifications*, on se retrouve devant toute une série de problèmes qu'on ne fera que mentionner ici. La plupart de ces questions sont d'ailleurs des thèmes de réflexion tout à fait reconnus en IA, couramment évoqués à propos des systèmes experts, de l'informatique linguistique ou de l'enseignement assisté. Les problèmes de contrôle, d'explication, de transfert de connaissances, d'ergonomie, s'y recroisent en un écheveau inextricable qui semble défier toute division rationnelle du travail.

3.1. Les exigences de la communicabilité sont profondément différentes selon les utilisateurs et les concepteurs, l'avancement et le type de leur compétence ; elles définissent par conséquent de nouveaux besoins en connaissances. C'est un lieu commun de l'EIAO qui a pour conséquence, comme l'ont démontré par excellence les travaux de Clancey autour de MYCIN, qu'il faut totalement restructurer (et donc en fait dupliquer) le système pour qu'il puisse devenir la base d'explications compréhensibles ; ces explications ne sont pas seulement utiles dans la perspective d'une aide à la formation, mais sont absolument indispensables, encore une fois, si l'on veut que les résultats produits par le système méritent et suscitent la confiance lorsqu'il est employé en situation réelle (pour un écho, voir Falzon 1989). Notons en passant que cette restructuration des systèmes est bien loin d'être automatisée, si bien qu'on ne peut prétendre pour le moment que les règles-pour-la résolution soient une forme "compilée" des règles-pour-l'explication. Par ailleurs, l'utilisateur n'attend pas nécessairement qu'on lui retrace un parcours déductif, mais plutôt qu'on lui

dresse un tableau cohérent de la situation, compatible avec les données, et sur lequel il puisse appuyer les actions à entreprendre. Dans tous les cas, les connaissances du système, telles que nous les avons présentées jusqu'ici, tentent de respecter des impératifs conflictuels de modélisation cognitive, de clarté à la réalisation, d'efficacité dans l'exécution. Elle ne contiennent généralement pas les représentations “redondantes” qui seules permettraient la justification de certaines modalités implicites du fonctionnement (pour la modélisation des domaines aussi bien que pour le contrôle des raisonnements). Il y a donc ici conflit de priorités entre les différents aspects que nous venons d'évoquer et les exigences anticipables de l'explication aux utilisateurs. À quel moment de la réalisation, de la conception, faut-il les prendre en compte ? Il est bien difficile de prendre parti (cf. Bouri, Dieng, Kassel, Safar, 1989).

3.2. Soulignons à présent quelques problèmes touchant au recueil des règles expertes, ou plus généralement à l'analyse des schémas interprétatifs de l'expert (*elicitation*).

(i) Comme tout énoncé, les verbalisations de l'expert explicitent bien moins qu'elles n'implicent : l'expert ne donnera pas la liste exhaustive des conditions où les règles et autres schémas ne doivent pas s'appliquer, et dans ce cas le système déduira “trop” ; l'expert oubliera aussi de mentionner toutes sortes de connaissances, parce qu'il n'en éprouve plus la spécificité et les considère comme relevant plutôt du sens commun, et dans ce cas le système ne pourra pas déduire “assez”. Cette incomplétude inévitable entraîne également l'incohérence de la connaissance transmise, ou du moins de son libellé, s'il n'est pas fait mention des conditions qui préviennent l'interférence de points de vue potentiellement contradictoires (qu'il faut malgré tout envisager simultanément !).

(ii) L'introspection, l'enquête réflexive menée par l'expert à la demande du cogniticien linéarise nécessairement ses schémas interprétatifs et projette sa dynamique mentale sur la forme a priori de “règle”. Elle

amplifie presque sûrement le rôle de la délibération dans le processus cognitif ; alors que la conduite de l'expert, y compris en situation problématique, est peut-être ordonnée, orientée, sans le recours exclusif à des règles (même subliminales), ou à des inférences (au sens logique du terme, qui implique un chaînage de règles d'une certaine importance).

(iii) L'échange est contraint par la forme linguistique, et se déroule donc dans un médium qui obéit à des lois bien différentes de celles qu'on suppose régir un hypothétique langage de la pensée et qui régissent effectivement les systèmes de représentations de l'IA. L'évocation simultanée de niveaux de réalité distincts sous le même terme est déjà en soi un problème<sup>35</sup> ; et la dissociation de ces niveaux, nécessaire à la cohérence de leurs représentations, introduit d'un autre côté des clivages qui empêchent la saisie des *contextes* et des *afférences*. Que pourraient bien être les analogues des tropes pour ces langages de concepts ? Certaines formes de raisonnement par analogie comparables au jeu de la métaphore, ou certains déclenchements "métonymiques" de processus, comme l'écrit C. Vogel (1988) ? Pour intégrer ces dimensions du raisonnement, il ne faudrait pas moins qu'une révolution du concept de *langage de la pensée*, qui le rendrait probablement incompatible avec sa représentation par un système symbolique classique ; il y faudrait aussi le développement d'une théorie sémantique capable de repérer dans le discours de l'expert la trace de ces opérations conceptuelles pour le moment irréprésentables.

(iv) La définition, la cohérence et la validité des croyances, lorsqu'elles sont exprimées dans une langue naturelle et engagent la conduite d'un sujet, ne survivent que très difficilement à une confrontation avec des normes strictement logiques du sens. Inversement, les structures qui satisfont parfaitement à ces normes logiques ne peuvent peut-être pas,

---

<sup>35</sup> Cf. par exemple l'analyse des emplois de *livre* ou *rue* par D. Kayser (1987) ; voir également les commentaires de Winograd et Flores sur les sens du mot *water* (1986, p. 55 et suivantes).

sauf cas très simples, être exprimées en langue naturelle sans qu'il subsiste quelque trace (notamment argumentative) du régime particulier de sens dans lequel elles auront été construites. Et cela, naturellement, ne serait pas sans effets sur le type d'explications qu'un système à base de connaissances serait capable de donner.

3.3. Une question qui n'a pas moins d'importance est celle de la présentation des données. Sont simultanément impliqués des problèmes de prétraitement, de changements de représentations exigés par la modélisation, et d'ergonomie des représentations. On peut imaginer différentes formes d'aide (cf. Falzon, 1989) : rendre plus accessible l'information pertinente, c'est-à-dire éliminer les données parasites, regrouper et résumer les données intéressantes, notamment les inférences, les compléter ou annoter de façon coopérative ; effectuer des prétraitements sur les données brutes de façon à les reconstituer dans un système de représentation adéquat ; changer de système de représentation à la demande ; offrir des capacités de prédiction par simulation de modèle, par raisonnement qualitatif ou caricatural. Sur tous ces aspects il est raisonnable d'espérer des progrès à relativement court terme : en effet, il n'est pas question ici de résoudre automatiquement des problèmes difficiles, mais seulement d'offrir une meilleure vue sur ce que le système fait, sur ce qu'on lui donne ou pourrait lui donner à traiter. Les interfaces graphiques, les environnements de simulation jouent ici le rôle principal. Des interfaces en langage quasi-naturel sont également envisageables à ce niveau : il s'agirait en effet d'explorer, de contrôler d'une façon relativement vague des structures et processus *qui ont été déterminés ailleurs*. Il ne s'agirait donc pas de compréhension en profondeur, mais seulement d'extraire rapidement et faire apparaître à l'écran les structures requises, en variant au besoin les formes de leur présentation. Les tâtonnements et les erreurs ne seraient donc pas susceptibles de

conséquences graves ; l'utilisateur pourrait probablement les détecter avec un peu d'expérience.

3.4. À moins d'épuiser et déconcerter perpétuellement, le système devra s'efforcer de fournir des résultats compatibles avec les préconceptions de son utilisateur, quand bien même elles ne seraient pas optimales. Ce serait d'ailleurs conforme à ce que révèle une analyse fine du comportement des experts vis à vis de leurs contractants (Falzon, 1989). Les difficultés extrêmes de la modélisation "implicite" de l'utilisateur étant maintenant bien connues, il vaudra mieux s'orienter vers la définition de langages de contrôle et de description des inférences recherchées ou des situations explorées, qui permettent de les orienter explicitement, de les spécifier à des précisions variables. Ces langages pourront être écrits, artificiels ou même quasi-naturels sous certaines conditions (voir ci-dessus, fin du paragraphe 3.3). Mais ils pourront également être iconiques, puisque les icônes peuvent être systématiquement assemblées et désassemblées, et permettent ainsi de réifier les structures et processus les plus variés. Leurs capacités représentatives couvrent d'ailleurs un spectre très large, depuis la simulation d'environnements physiques jusqu'à l'écriture d'idéogrammes dynamiques (P. Lévy, 1990-91)<sup>36</sup>.

---

<sup>36</sup> Les termes d'*idéogramme* et d'*idéographie* présentent quelques risques de confusion. Les idéogrammes dont il est question ici ne sont l'écriture d'aucune langue phonétique, comme le sont les idéogrammes chinois. Par ailleurs, ils ne sont en aucune façon la projection "directe" d'une idée, au sens psychologique du terme. Ils ne sont donc pas exactement des "modèles mentaux", mais plutôt des *figurations* associées à une objectivité informatique sous-jacente, qui exploitent les mécanismes (sémiotiques) de la ressemblance iconique pour conférer à ces structures informatiques *internes* une sémantique *externe* (conceptuelle ou extensionnelle, peu importe ici). Cette sémantique leur convient jusqu'à un certain point seulement, mais suffisamment pour faciliter leur manipulation et leur interprétation (comme dans le cas des langages de programmation par icônes) ; le cas échéant, elle permet en même temps une exploration non triviale de l'objectivité externe évoquée par la figuration (cas des environnements de simulation et des idéographies dynamiques théorisées par P. Lévy).

Il va sans dire qu'une bonne partie des problèmes évoqués dans cette section ne pourra commencer de recevoir des solutions qu'à partir du moment où l'évaluation des systèmes se fera suivant le point de vue des utilisateurs, et non des réalisateurs qui considèrent trop souvent comme un succès toute situation où le système produit une réponse que l'expert accepterait de valider. Encore une fois, ce genre de résultats bruts n'est pas d'une grande utilité pour l'utilisateur lui-même, qui a besoin de bien autre chose pour les obtenir et se former une opinion. Il faut donc espérer, pour le bien de tous, que ce type d'évaluations se généralise, si mauvais soient les résultats qu'il faudra dans un premier temps consentir à publier.

#### ***IV. Vers un nouveau schéma régulateur.***

La première section de cet article, consacrée à la modélisation cognitive, a rappelé diverses raisons, phénoménologiques ou épistémologiques, de récuser le paradigme dominant en Intelligence Artificielle, qui est celui des systèmes symboliques stricts, directement interprétables à partir de contenus psychologiques intuitifs. Nous n'avons cependant posé, ce faisant, aucune limite précise *a priori* sur le potentiel des ordinateurs. Les processus de pensée, qu'ils soient analysés au niveau mental ou neuronal, sont à présent les objets de diverses (techno)sciences dont les progrès sont manifestement liés au développement concomitant de l'informatique. Mais cela n'entraîne pas, loin de là, de recours croissant à un style fortement "anthropomorphe" de représentations, auxquelles on appliquerait, de surcroît, des contraintes syntaxiques strictes. D'un point de vue "artificialiste", également, il se pourrait qu'en s'éloignant progressivement de ce sens commun qui reste actuellement leur guide de construction en IA, et donc en intégrant davantage de mathématique ou de physique tout en continuant à faire de la logique, les systèmes réalisent des performances très complexes, par des moyens internes et avec des résultats externes apparemment différents des nôtres. Les symboles intervenant dans la

construction de ces systèmes seront simplement ceux de langages de programmation, possiblement très sophistiqués, et leur contenu psychologique pourra bien se révéler parfaitement inscrutable<sup>37</sup>. Quel que soit d'ailleurs le domaine, naturel ou artificiel, il ne serait pas étonnant de découvrir que l'étude ou la simulation de nombreuses performances, qui ont pu sembler dans un premier temps relever de dimensions psychologiques relativement simples et indépendantes, exigent en fait des approches très différentes de celles que l'IA a tentées jusqu'ici. Le problème de l'Intelligence Artificielle, convenablement réorientée, serait alors de faciliter le *couplage* entre les humains et des systèmes de cette (toute autre) complexité.

Mais bien sûr nous n'en sommes pas encore là. La brève histoire de l'IA est marquée d'un certain nombre de succès et de déboires, tous également significatifs, et son développement actuel est soumis à des contraintes, toujours les mêmes, qui semblent décidément incontournables et parfois inconciliables. Nous en avons analysé, ou tout au moins mentionné, un certain nombre, en les regroupant sous nos trois rubriques. Quelles conclusions, provisoires bien sûr, peut-on maintenant proposer ?

Les conflits de priorité entre les trois sortes de contraintes évoquées, l'extrême difficulté de l'apprentissage des structures complexes, la faiblesse opérationnelle, pour ne pas dire l'inexistence, des modélisations classiques de phénomènes tels que la perception visuelle et la catégorisation, la mémoire associative et le transfert analogique, conduisent à abandonner le

---

<sup>37</sup> Dire à ce propos que l'IA (classique) "fait l'hypothèse que les processus de pensée peuvent être simulés sur ordinateur" (G. Sabah, conférence à l'ISCC, octobre 1990) ne donne donc aucune indication sur l'IA, si ce n'est qu'on souhaite l'insérer dans le réseau des technosciences. Si l'on réalise, en effet, le rôle des ordinateurs dans les modélisations produites par les disciplines scientifiques, la phrase précédente signifie simplement que les processus de pensée peuvent être pris comme objets de (techno)science. Elle ne dit rien sur les tâches spécifiques de l'IA, ni sur le statut particulier des symboles informatiques dans ce contexte.

concept de *système symbolique strict, intelligent et autonome* comme schéma régulateur de la recherche en matière de systèmes à base de connaissances.

Pourrait peut-être lui être substituée la notion de *système interactif d'aide à la modélisation et à la validation des raisonnements*. Le mot *aide* indique la modestie, démobilisante comme il se doit, de ce “nouveau” concept ; et le terme *modélisation* est là pour écarter d’emblée toute tentation de réalisme psychologique par trop naïf, étant bien entendu qu'on ne peut, à l’heure actuelle, espérer obtenir un peu de complexité conceptuelle et inférentielle, sinon en reconstruisant, le cas échéant à partir de modèles dits *cognitifs*, des systèmes disposant aussi d'un schéma régulateur logique rigoureux (même s'il n'est pas rigoureusement appliqué, comme on l'a vu à la section 2). Dans ces conditions, il importe de se rappeler que les descriptions psychologiques, tout comme celles de la microsociologie et de la sémiologie des interactions, seront profondément altérées par cet inévitable embrigadement logique. Et il convient également de rester neutre vis-à-vis de cette métamorphose, dès le moment où elle n'est pas déniée. Toute technologie délimite et fige le potentiel cognitif humain ; mais en même temps elle le développe, en lui permettant de s'exprimer différemment dans son champ d'intervention spécifique. Ici également, les sujets pourront s'adapter, et apprendre à apprécier les résultats obtenus en suivant ces principes particuliers d'*objectivation*<sup>38</sup>. En

---

<sup>38</sup> Le sème de la réitération, porté par le préfixe *re-* des *reconstructions logiques*, est déjà de trop, si l'on entend conjurer les tentations réalistes. Et il n'y a aucune raison de conformer prématurément les descriptions psychologiques aux exigences de leur future objectivation logique ou informatique (à moins que cela soit dès le début conforme à l'intuition) : en faisant cela, on procède à des assimilations hâtives qui empêchent d'identifier précisément les points sur lesquels une informatisation pourrait se révéler possible et utile. Mieux vaut partir d'une approche très “descriptive”, c'est-à-dire riche en analogies, en métaphores sans contrepartie convaincante dans l'univers logico-informatique ou mathématique. C'est seulement à ce prix que l'on peut espérer décider en toute conscience des dégradations que la description *doit* subir pour être projetée

somme, le développement remarquable des systèmes symboliques opéré sous l'égide du paradigme "classique" des sciences cognitives n'est pas en lui-même un facteur suffisant d'*autonomie* cognitive : les succès, la valeur de ces systèmes dépendent toujours en dernière analyse de l'implication constante et judicieuse de leurs utilisateurs.

Le terme de *système expert* doit être abandonné – et d'ailleurs il commence à l'être, comme on l'a noté en introduction : il faisait espérer des qualités qui sont hors de portée de ces systèmes, comme celle de se substituer véritablement à l'expert, au décideur responsable qu'ils caricaturent. Leur utilité réelle serait plutôt à chercher dans la diminution de la charge cognitive sur tel ou tel des aspects de la tâche de raisonnement<sup>39</sup> ; mais il faudra pour cela des utilisateurs relativement avertis du domaine modélisé et des limites de la technologie informatique. Il s'agit en somme d'offrir davantage de productivité à des compétences plutôt "réceptives", de leur permettre d'approcher ainsi les performances d'une compétence nettement "productive", avec une offre combinée de savoir encyclopédique, de modélisations plus ou moins schématiques avec réalisations graphiques (environnements de simulation, voire idéographies dynamiques telles que les décrit P. Lévy), et enfin de *suggestions de solutions que l'utilisateur aurait peut-être pu trouver avec l'effort nécessaire, et qu'il pourra ici se contenter de compléter, corriger et valider*. L'ensemble du système ne doit pas être conçu comme la simulation d'une intelligence experte, même si l'on admet qu'un moment important de sa conception procède de cette approche, mais plutôt comme un *artefact*

---

dans ces divers types d'objectivités, et le cas échéant, prévoir les contrôles "compensateurs" dont le système devra offrir la possibilité à l'utilisateur.

<sup>39</sup> Sans parler ici d'éventuels effets épistémologiques ou sociologiques au niveau des organisations : description et archivage des savoir-faire spécifiques, réflexion sur les méthodes, refonte des terminologies, clarification et transformation des rôles professionnels, etc. (cf. P. Lévy, ce numéro). Ces effets ne dépendent pas de la valeur d'usage "substitutive" des systèmes réalisés ; ils apparaissent plutôt comme des effets secondaires du processus de réalisation lui-même.

*interactif* (selon l'expression de L. Suchman) qui tire sa valeur d'un compromis passé entre plusieurs fonctions à la fois indispensables et imparfaitement conciliables. L'interface elle-même ne doit pas s'appuyer nécessairement sur le projet de simuler la communication humaine ; dans la plupart des cas, elle ne doit pas chercher, ce qui serait vain, à masquer systématiquement la profonde asymétrie des partenaires.

“Les erreurs commises dans le domaine de l'IA, écrivait déjà D. Marr en 1977, ne résident pas dans le fait d'avoir mené des recherches de ce genre [...] mais essentiellement dans une faute de jugement quant à leur valeur”<sup>40</sup>. On ne saurait mieux dire qu'il n'y a rien à abandonner de ce qui a été fait depuis les débuts. Les représentations actuelles de l'IA, ainsi que celles à venir, peuvent toujours être mises à contribution, à condition de se rappeler cependant qu'il s'agit de *modélisation*, et non de *simulation*<sup>41</sup>. Comme on l'a dit, ces modélisations répondent à divers objectifs et contraintes ; et même si elles intègrent des descriptions linguistiques ou psychologiques, ce n'est pas exactement pour simuler des comportements humains.

*Modèle* doit plutôt s'entendre ici au sens de figuration, de maquette susceptible de servir les buts de la connaissance, de schéma directeur pour

---

<sup>40</sup> Cité par D. Andler dans son avant-propos à la traduction française de *What computers can't do* de H. Dreyfus (1984).

<sup>41</sup> Le terme de *simulation* est employé ici dans le sens fort, mythique depuis Turing, qui fait généralement référence en IA : il suggère la production d'une réplique indiscernable, ou à tout le moins fonctionnellement substituable à un original sans provoquer de déficit *significatif* en situation d'interaction réelle. Cet usage du terme s'oppose à celui d'autres champs scientifiques (par exemple en physique), où les “simulations”, de par leur caractère “partiel” et... d'autant plus effectif, s'identifient à ce que nous appelons ici *modélisations*. Ce dernier terme a de son côté des connotations scientifiques qu'on jugera peut-être trop prononcées (par exemple si on a en vue un contexte d'emploi “grand public”). Au moins a-t-il le mérite de suggérer l'importance des aspects “schématisant” ou “visuels”, et de ne pas masquer les difficultés (en évoquant à tort des interactions aussi spontanées qu'une conversation).

un certain type d'objectivation<sup>42</sup>. Tout comme le modèle proprement scientifique, il possède deux fonctions majeures : “il offre une contrepartie, dans l'ordre des structures figuralement claires, mathématiquement exactes, aux états de choses diffus que décrivent les sciences empiriques. Mais il intervient aussi en regard des élaborations théoriques [...] et il fournit à celles-ci une référence objective qui peut consister dans une figuration géométrique ou dans un symbolisme algébrique”. Premier cas : transformation d'une objectivité diffuse en objectivité exacte. En contrepartie de sa relative simplicité, le modèle offre alors des possibilités de validations ou réfutations partielles, à raison du caractère bien cerné du type d'objectivité dans lequel il est construit. Deuxième cas : figuration concrète du théorique. Il facilite alors l'approche par sa fonction d'icône. Le modèle sert à la fois “les intérêts de la compréhension, laquelle bénéficie de la clarté des figures, ceux de la recherche, qui utilise la puissance d'analogie des schémas, ceux mêmes de la preuve, qui trouve dans l'objectivité des modèles une part de ses garanties”. En un mot, c'est une “fiction surveillée”, qu'elle soit mathématique, physique, informatique, etc., dont les règles d'interprétation appellent à leur tour une autre modélisation, et ainsi de suite à l'infini... Le sens d'un modèle ne peut se déduire entièrement des lois de fonctionnement du domaine d'objets dans lequel il est construit. Il y faut également le temps de l'interprétation, du transfert au domaine initial, qui exige de l'intuition et un certain sens de *l'analogie* qui est précisément à la base de la modélisation. On ne peut utiliser un modèle dans l'ignorance totale de ce qui le motive.

Si l'on revient alors aux systèmes de l'IA, la discussion qui précède suggère à l'évidence que ce n'est pas seulement le processus d'acquisition des connaissances qui mérite d'être qualifié de modélisation qualitative. *C'est l'utilisation elle-même qui est aussi une modélisation* : elle réclame

---

<sup>42</sup> Les citations qui suivent dans ce paragraphe sont extraites de l'article “Modèles” de *l'Universalis*, et plus précisément de l'introduction et de la section 9 dues à N. Mouloud (première édition, 1968 ; conservé dans la troisième édition, 1989).

par conséquent des interventions fréquentes de l'utilisateur sur des représentations qui ont un statut purement instrumental, et ne peuvent acquérir une sémantique que par une opération de transfert analogique, très imparfaitement simulable par le système, on s'en doute.

L'Intelligence Artificielle, suivant en cela les deux fonctions des modèles citées plus haut, procéderait alors à partir de deux mouvements fondamentaux : celui de la réification des signes et des théories, et celui de la sémiotisation<sup>43</sup> des structures et processus de divers domaines. Dans un tel medium représentationnel, simultanément dynamique, versatile, et interactif<sup>44</sup>, s'engendrent alors des univers qui sont, étrangement, *à la fois exacts et concrets*. Les symboles, icônes, idéogrammes manifestés à l'écran, renvoient évidemment à des objets informatiques sous-jacents (appelons-les des *représentations*) qui ont leurs propres lois objectives de fonctionnement, dictées par la modélisation. Mais en même temps ce sont des entités sémiotiques, des *signes* pour nous, dont la valeur dépend, de proche en proche, d'un gigantesque complexe interprétatif qui s'égalé peut-être à la *forme de vie humaine*<sup>45</sup> dans son entier, en tels temps et lieu, pour telle personne, dans tel environnement physique et social, etc. Selon une argumentation bien connue, un modèle physico-mathématique ou informatique s'inscrit dans une objectivité bien particulière, fondée sur un "oubli" très ingénieux et obstiné de ce qui contrevient à sa détermination ou ne lui est pas nécessaire. Le monde sémiotique, quant à lui, n'est pas un objet au même titre, puisqu'il inclut en lui la possibilité de tous les objets.

---

<sup>43</sup> On nous accordera cet emploi, avec le sens de "trouver ou inventer un signe pour".

<sup>44</sup> Cf. la remarquable argumentation de W. Hill (1989) en faveur d'une conception de l'IA comme construction informatique de *media représentationnels* (se distinguant non seulement de la simulation cognitive, mais également de la réalisation d'intelligences artificielles "de substitution"). À rapprocher également des réflexions de P. Lévy sur l'informatique comme moment dans l'histoire des systèmes d'enregistrement et d'écriture (1990c).

<sup>45</sup> Expression wittgensteinienne, mais que Wittgenstein n'a pas employée, semble-t-il (l'adjectif *humaine* est de trop). Je l'emprunte à un commentaire de J.-M. Salanskis sur un passage des *Investigations Philosophiques*.

Les fonctions sémiotiques (inépuisablement interprétables) d'un symbole et ses fonctions représentationnelles (liées à une objectivité spécifique) ne peuvent donc jamais correspondre exactement<sup>46</sup>. Si elles le peuvent parfois, ce ne peut être par l'intermédiaire exclusif d'un domaine d'objectivité particulier, quel qu'il soit. Pour autant que nous le sachions, seul le sujet humain, au vu de ce qu'il sait, de ce qu'il perçoit, de ce qu'il sent, peut "décider" l'invariance des corrélations. S'il lui apparaît, selon ses ressources propres (notamment informatiques), qu'il est meilleur de modifier ces corrélations, il doit réformer les contenus représentationnels et accepter peut-être une modification de valeur des signes, ou bien même introduire de nouveaux signes. Cette tâche d'adaptation finit toujours par le requérir, à un moment ou un autre, même si, effectivement, il peut décider de s'en remettre, provisoirement et localement, à des mécanismes ou des mesures. L'existence de domaines scientifiques, où la sémiotique semble totalement verrouillée aux contenus représentationnels, ne contredit pas ce point de vue. Ce sont en effet les sujets humains qui assurent eux-mêmes que les conditions d'appartenance au domaine d'expérience couvert par la science sont bien réunies (modes de raisonnements, observations, etc.) : ces conditions ne sont que très partiellement objectivées par le processus scientifique. De la même façon, les micro-mondes explorés par l'IA des années 70 ne pourraient modéliser les environnements réels de la robotique sans l'intervention *décisive* des opérateurs humains qui les contrôlent effectivement et sont seuls, en définitive, à pouvoir les *maintenir* et les *valider*.

Si donc l'*utilisateur* est lui aussi un *modéliseur*, il ne peut être passif et ignorant de ce qui se trame dans le système – comme s'il n'avait qu'à appliquer les résultats sans la responsabilité de les maintenir et valider, encore une fois. Les résultats seront aussi *ses* résultats. On peut certes

---

<sup>46</sup> Ce passage est librement inspiré d'un texte de F. Rastier, où se trouve notamment présentée une distinction entre monde sémiotique et monde représentationnel (1991, épilogue).

imaginer des cas où l'utilisation ne semble pas entraîner cette responsabilité particulière, et où elle se fait presque machinale : cela peut arriver, en effet, si l'on estime qu'il s'agit de choses qui ont été *décidées une fois pour toutes*, que ce soit par soi-même ou par d'autres. Il peut s'agir tout d'abord de domaines scientifiques ou techniques, où les conditions de vérité, la valeur de connaissance, la pérennité et l'efficacité des résultats ont été assez strictement établies. Mais cela peut également concerner des domaines où les mécanismes de la réfutation ne peuvent pas jouer, ou bien qui sont, à des titres divers, par trop imprévisibles : les raisonnements y conservent parfois une forte plausibilité, même si leurs conclusions s'avèrent souvent fausses ; et l'application généralisée des meilleures anticipations peut précisément avoir pour effet de les invalider (analyse des valeurs financières, par exemple). Cependant, c'est ici la valeur performative, plus que la valeur de connaissance, qui importe principalement : ce type de systèmes produit généralement des résultats à portée éphémère, jetables après usage, et s'ils sont utilisés, c'est parce que les organisations se trouvent bien, tout simplement, de les avoir intégrés à leur fonctionnement. Ils ont donc le statut ambigu d'*oracles* : qu'on prenne ou non le parti de les suivre, ce n'est pas d'eux que l'on attend véritablement un développement de la *rationalité* des organisations – même si leur complexité, leur systématité en sont parfois accrues.

Encore une fois, et y compris dans tous les cas mentionnés ci-dessus, l'utilisateur ne peut être dispensé d'apprécier intuitivement les résultats du système, en explorant, le cas échéant, les structures intermédiaires qui les sous-tendent. C'est bien ce que signifie, en somme, le mot *résultat* : une *sortie* que l'utilisateur doit *interpréter* avant d'en faire éventuellement une *entrée* pour la procédure suivante. Dans ces conditions, le rôle du système est essentiellement de fournir des schémas de représentations (de structures et de processus), et de faciliter leur manipulation et leur compréhension en les couplant à des entités qu'on peut raisonnablement doter d'une valeur sémiotique, *quitte à procéder aux ajustements qui rendent le couplage*

*acceptable*. L'IA des systèmes de raisonnements, vue sous cet angle “intime” du couplage entre des postes informatiques et des travaux, des oeuvres ou des actions individuels, serait alors comparable à une sorte d'*ergonomie de la sémiotisation et de la modélisation des connaissances*. Cette ergonomie ne peut se réduire totalement à des normes édictées *a priori* ou à des anticipations, si géniales soient-elles. Même si l'offre d'outils génériques précède les demandes spécifiques qu'elle contribue d'ailleurs à déterminer, elle devra aussi s'appuyer récursivement sur l'observation et la consultation des utilisateurs-modélisateurs des versions successives. On voit mal, en effet, comment il serait possible de bien utiliser un système sans comprendre, ou tout au moins *se figurer*, quelques-uns de ses principes de fonctionnement. Et à partir de là on devra probablement admettre, avec la montée en complexité, une adaptation importante et relativement imprévisible aux exigences individuelles de cette compréhension, de cet emploi. Comme on l'a déjà par trop souligné, l'utilisateur ne peut être dénué de compétences dans les champs de connaissances mises en oeuvre ; avec la diffusion de la culture informatique et le raffinement progressif des capacités sémiotiques des systèmes, d'autre part, une certaine compréhension intuitive des processus sous-jacents va se développer. Il faut donc s'attendre à une intervention croissante des utilisateurs lors des phases de spécification, et peut-être même à une remise en cause de leur trop nette séparation d'avec les phases de conception.

Enfin, et ce n'est pas le moins important, il n'est pas impossible que nous assistions, chemin faisant, à une relativisation, voire une dissolution, de la problématique de la “simulation des comportements” qui a jusqu'ici tenu lieu de schème régulateur principal au développement de l'Intelligence Artificielle<sup>47</sup>. Construire un système n'est pas exactement offrir une

---

<sup>47</sup> La simulation des comportements évoquée ici ne se réduit pas à celle que l'on tenterait de fonder sur des mécanismes internes psychologiquement plausibles (c'est-à-

réplique de substitution à un segment de savoir-faire défini *a priori* au moyen de catégories pré-informatiques ; c'est bien plutôt modifier l'exercice de ce savoir-faire en lui fournissant un nouvel appui objectif (qui présente, le cas échéant, une réification mimétique de certains aspects de la situation antérieure). Cette simple *constatation* devrait empêcher que l'on confonde encore tentatives de simulation et ingénierie des connaissances, comme l'ont si souvent fait, pour des raisons par trop évidentes, la plupart des acteurs concernés, langues de bois universitaires et mythologies industrielles se répondant ici parfaitement. Disant cela, nous ne voulons évidemment pas prétendre que ce soit "mal" de simuler lorsque cela semble possible. Mais en intégrant sans trop s'en inquiéter et dès le départ le principe de simulation dans leur philosophie spontanée, les chercheurs du domaine n'ont fait que retarder le moment de la confrontation avec certaines questions pourtant capitales. Il nous semble à présent que la force d'entraînement indéniable qui a résulté de cette méconnaissance est pour l'essentiel épuisée, et qu'il serait temps de donner leur chance à d'autres méconnaissances, possiblement tout aussi intéressantes. Dans son célèbre article de 1950 *Computing Machinery and Intelligence*, Alan Turing avait pourtant bien défini son *Jeu de l'Imitation* comme une mise en scène cryptée des paradoxes attachés au concept de simulation. En lui assignant simultanément pour thèmes l'intelligence et la différence des sexes, il transmettait au passage, tout en les déniait, un certain nombre d'enseignements valables de ces deux points de vue : simuler est aussi tromper, les êtres parlants sont en même temps sexués, leur inscription dans cette différence est décisive et déjà elle-même affaire de simulacres, les critères de sa simulation n'en sont que plus incertains, etc. Mais cela ne suffisait pas. Il fallait encore envisager que ce Jeu de l'Imitation puisse

---

dire à ce qu'on appelait la simulation cognitive). Elle comprend également les interactions à caractère mimétique obtenues par tous les moyens possibles (par exemple, une interface avec une base de données faisant appel à une sémantique logique inspirée par la structure de la base, bien plus que par des considérations linguistiques ou psychologiques).

avoir une histoire, que sa généralisation-même, sa socialisation massive puisse en modifier profondément la valeur (en particulier celle de critère possible d'intelligence pour les machines capables de s'y intégrer). N'est-ce pas précisément ce qui est arrivé au marivaudage télématique en France ? L'impatience des utilisateurs des messageries, l'opacité et la lenteur des tours de parole, la faible qualité esthétique et ergonomique du support, et finalement l'intervention d'opérateurs stipendié(e)s auront progressivement créé un genre discursif très particulier, qui semble parfois à la portée de l'informatique linguistique contemporaine<sup>48</sup> : et cela dans une situation proche de celle imaginée par Turing, et sans que l'on puisse pour autant en tirer la moindre conclusion sur "l'intelligence" des programmes susceptibles de donner le change. Ainsi donc, faute d'avoir véritablement pris en compte le caractère *situé* des interactions qu'ils s'efforcent de concevoir, les chercheurs en IA auront trop souvent manqué d'intégrer à leur dispositif *théorique* certains des éléments qui se manifestent de façon décisive dans les contextes dits *applicatifs*<sup>49</sup>. Par ailleurs, comme nous le savons tous, la simulation est généralement si difficile qu'il apparaît maintenant bien imprudent d'avoir trop étroitement subordonné la valeur de l'IA à l'obtention de succès marquants de ce point de vue : c'était préparer la voie à la diffusion massive de simplismes justificateurs (affectés ou sincères, peu importe ici) dont la dissipation brutale a pu parfois entraîner quelque discrédit. Un certain fonctionnalisme psychologique, particulièrement caricatural à cet égard, s'est même laissé aller à suggérer que des potentiels cognitifs substantiels pourraient être engendrés à partir de simples "transcriptions" d'observations à peine élaborées : ce qui s'actualisait en situation se réduisant dans cette

---

<sup>48</sup> Affirmation à nuancer selon le style de la messagerie et les critères de succès retenus (qui ne devraient pas aller au-delà de la participation sous pseudonyme à un certain type d'effervescence convenue, les effets de rencontre personnelle restant probablement hors de portée).

<sup>49</sup> Thèse argumentée par Marc Nossin dans "La dimension applicative du traitement automatique des langues naturelles" (1991).

perspective à une plate “instanciation” d'un potentiel lui ressemblant à s'y méprendre (à quelques indices occurrenceiels près). C'était rétrospectivement pousser l'anthropomorphisme un peu trop loin, alors même qu'une riche tradition préexistante de descriptions phénoménologiques, pourtant par essence anthropocentriques, s'attachait de longue date à déconstruire les fausses évidences de notre univers de formes significatives. C'était également chercher un peu vite à absorber dans les dispositifs informatisés de la raison technologique des performances “expertes” qui se tiennent à la fois aux fondements et aux marges de cette rationalité (même astucieusement limitée et focalisée pour l'occasion). Il était peut-être plus prudent et profitable de supposer que ces performances relèvent en effet de l'objectivité informatique non parce qu'elles s'y réduisent, mais parce qu'elles peuvent y trouver un appui efficace.

Enfin, comme y ont suffisamment insisté plusieurs auteurs déjà cités dans cet article, le principe de simulation, outre qu'il présuppose que la simulation de l'humain est dans tous les cas le meilleur type d'interaction, le fondement le plus judicieux du couplage avec les machines, passe à côté de la possibilité la plus réelle offerte par l'IA, qui est d'étendre les possibilités d'*objectivation* par le truchement de nouveaux media représentationnels, qui manifestent et reconfigurent le potentiel cognitif humain sur des modes de plus en plus variés. Les éventuelles qualités mimétiques de ces représentations dépendent évidemment, entre autres choses, de nos capacités d'adaptation. Les pensées d'un(e) mathématicien(ne) ressemblent-elles aux notations manipulées ? Oui, en un certain sens, dès le moment où une pratique suffisante vient renforcer les couplages... Avant de s'interroger sur l'authenticité des simulacres, il conviendrait peut-être d'aborder plutôt, à chaque fois, le problème de l'objectivation, de son complexe instrumental et sémiotique, de ses situations de référence. Il serait intéressant, en particulier, de chercher à préciser, dans cette perspective et à un horizon raisonnable, les rôles

informatiques que pourraient jouer les langues naturelles dans ces procès d'objectivation. Mais c'est là le sujet d'un autre travail.

Yves-Marie VISETTI  
LIMSI-CNRS  
B.P. 133 – 91403 Orsay

**N.B. :** Les positions exprimées dans cet article n'engagent évidemment que moi. Je crois cependant devoir beaucoup à un certain nombre de discussions et d'indications bibliographiques, pour lesquelles je voudrais tout particulièrement remercier Philippe Dague, Yves Gueniffey, Violaine Prince, François Rastier, et Jean-Michel Salanskis. La version présentée ici s'appuie dans une certaine mesure sur un rapport de lecture fort clair (et très critique) dont je voudrais également remercier l'auteur.

### *Bibliographie*

- ANDLER, D. (1990) Connexionnisme et cognition : à la recherche des bonnes questions, in *Revue de synthèse*, numéro spécial sur les sciences cognitives, n° 1-2, janvier-juin 90, p. 95-127.
- ARTIFICIAL INTELLIGENCE* (1990), numéro spécial consacré aux modèles connexionnistes (sous la direction de G. Hinton), 46, 1-2.
- ARTIFICIAL INTELLIGENCE* (1991), numéro spécial consacré aux fondations de l'IA (sous la direction de D. Kirsh), 47, 1-3.
- BACHIMONT, B. (1990) *Cohérence et convergence dans un tableau noir : organisation, formalisation et sémantique de l'architecture de contrôle ABACAB*, Thèse de Doctorat de l'Université Paris 6.
- BENABOU, J. (1988, 1990) Théorie des ensembles empiriques 1 et 2, *Cahiers de poésie comparée*, Mezura n° 17 et 21 [notes de cours rédigées par P. Lusson].
- BIRNBAUM, L. (1991) Rigor mortis : a response to Nilsson's "Logic and artificial intelligence", *Artificial Intelligence*, 47, 1-3, p. 57-77.
- BOURI, M., DIENG, R., KASSEL, G., SAFAR, B. (1989a) Systèmes à base de connaissances et explications, *Actes des 3èmes journées nationales du PRC-GDR IA, Paris (5-7 Mars )*, p. 328-339.
- BOURI, M., DIENG, R., KASSEL, G., SAFAR, B. (1989b) Vers des systèmes experts plus explicatifs, *Actes des 3èmes journées nationales de PRC-GDR IA, Paris (5-7 Mars )*, p. 340-355.
- BRACHMAN, R., LEVESQUE, H. (1985) *Readings in Knowledge Representation*, Morgan Kaufmann.

- CATACH, L. (1989) *Les logiques multimodales*, Thèse de Doctorat de l'Université Paris 6.
- CHANDRASEKARAN, B. (1987) Towards a functional architecture for intelligence based on generic information processing tasks, *IJCAI-87* (Milan), p. 1183-1192.
- CLANCEY, W. (1985) Heuristic classification, *Artificial Intelligence*, 27, p. 289-350.
- DE KLEER, J., WELD, D. (1990) *Readings in Qualitative Reasoning about Physical Systems*, Morgan Kaufmann.
- DELGRANDE, J. (1987) An approach to default reasoning based on a first order conditional logic, *Proc. of the AAAI Conf. (Seattle)*, p. 340-345.
- DESCLES, J.-P. (1990) Les représentations intermédiaires, in *Revue de synthèse*, numéro spécial sur les sciences cognitives, n° 1-2, janvier-juin, p. 33-56.
- DOLAN, Ch., SMOLENSKY, P. (1988) Implementing a connectionist production system using tensor product, *Proceedings of the 1988 Connectionist Models Summer School*, Carnegie Mellon University.
- DREYFUS, H. (1972, 1979) *What Computers Can't do : The Limits of AI*, Harper & Row ; trad. fr. (1984) : *Intelligence Artificielle : mythes et limites*, Flammarion .
- DREYFUS, H. (1981) From micro-worlds to knowledge representation : AI at an impasse, in *Mind Design* (éd. J. Haugeland), p. 161-204, MIT Press.
- DREYFUS, H., DREYFUS, S. (1986) *Mind over Machine*, New York, Macmillan.
- DREYFUS, H., DREYFUS, S. (1988) Making a mind versus modelling the brain : AI back at a branch point, *Artificial Intelligence* 117, n° 1 [repris dans *Philosophy of AI* (éd. M. Boden), p. 309-333, Oxford U.P., 1990].
- FAGIN, R., HALPERN, J. (1985) Belief, awereness, and limited reasoning, *Proc. of the Ninth Int. Joint Conf. of Art. Int.*, p. 491-501.
- FALZON, P. (1989) Assister l'activité. *Actes de la Conférence "Les Systèmes Experts et leurs applications"*, Avignon, p. 383-391.
- FARGUES, J. (1985) Non classical logic as a basis for a typology of inference engines, in *Computers and Computing, Proceedings of the Int. Conf. Future Trends in Computing*, p. 148, Masson et Wiley & sons.
- FERBER, J. (1987a) Approches réflexives en informatique, *Actes de Cognitiva 87*, p. 401-407.
- FERBER, J. (1987b) Des objets aux agents : une architecture stratifiée, *Actes de AFCET-RFIA 87*.
- FODOR, J. (1983) *The Modularity of Mind, an Essay on Faculty Psychology*, MIT Press ;[trad. fr. (1986) : *La modularité de l'esprit*, Ed. de Minuit].
- GENESERETH, M., NILSSON, N. (1987) *Logical Foundations of AI*, Morgan Kaufmann.
- GOLDBLATT, R.I. (1986) *Topoi : the Categorical Analysis of Logic*, North Holland.
- GOSCHKE, T., KOPPELBERG, D. (1990) The concept of representation and the representation of concepts in connectionnist models, Rapport no 36 du groupe de recherche *Mind and Brain* de l'Université de Bielefeld.

- GUHA, R., LENAT, D. (1989) *Building Large Knowledge Based Systems*, Addison Wesley.
- HALPERN, J. (1988?) Reasoning about knowledge : an overview, in *Theoretical Aspects of Reasoning about Knowledge (II)*, p. 1-17, Morgan Kaufmann.
- HALPERN, J., MOSES, Y. (1985) A guide to the modal logics of knowledge and belief: preliminary report, *Proc. of the Ninth Int. Joint Conf. of Art. Int.*, p. 480-490.
- HAYES, P. (1979) The logic of frames, in *Frame Conceptions and Text Understanding* (éd. Metzging), p. 46-61, Berlin, De Gruyter [repris dans *Readings in Knowledge Representation* (éds. : Brachman & Levesque), 1985, p. 287-295, Morgan Kaufmann].
- HAYES-ROTH, B. (1985) A blackboard architecture for control, *Artificial Intelligence*, vol. 26, p. 252-321.
- HILL, W. (1989) The mind at AI : Horseless carriage to clock, *AI Magazine*, vol. 10, n° 2, p. 29-41.
- INTELLECTICA (1990) numéro spécial sur les *Modèles connexionnistes* (éds. : D. Memmi et Y.-M. Visetti), n° 9-10.
- JAFFAR, J., LASSEZ, J.-L. (1986) Constraint logic programming, publication interne IBM 86.
- JAFFAR, J., LASSEZ, J.-L. (1987) Constraint logic programming : "A reader", et "Summary of CLP theory", *Fourth IEEE Symp. on Logic Programming*, San Francisco.
- KAYSER, D. (1987) Une sémantique qui n'a pas de sens, in *Sémantique et intelligence artificielle* (éd. F. Rastier), *Langages*, septembre 87, p. 33-45.
- LAKOFF, G. (1987) *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*, University of Chicago Press.
- LAMBEK, J., SCOTT, P.J. (1986) *Introduction to Higher Order Categorical Logic*, Cambridge University Press.
- LANGACKER, R. (1986) An introduction to cognitive grammar, *Cognitive Science*, 10, p. 1-40.
- LANGACKER, R. (1987) *Foundations of Cognitive Grammar*, vol. 1, Stanford University Press.
- LEHMAN, D. (1984) Knowledge, common knowledge and related puzzles, *Proc. of the third annual ACM Conf. on Principles of Distributed Computing*, p. 62-67.
- LEVESQUE, H. (1984), A logic of implicit and explicit belief, *Proc. of the Nat. Conf. of AI*, p. 198-202.
- LÉVY, P. (1990a) Vers une idéographie dynamique, *Actes du colloque international sur les instruments de communication évolués, les hypertextes et les hypermédias* (15-17 Mai). Université Paris 8, p. 37-82.
- LÉVY, P. (1990b) Interfaces informatiques et invention culturelle, in *Actes du second colloque sur l'ingénierie des interfaces hommes-machines*, AFCET, Biarritz (17-18 Sept.), p. 53-62.

- LÉVY, P. (1990c) *Les technologies de l'intelligence. L'avenir de la pensée à l'ère informatique*. La Découverte.
- LÉVY, P. (1991) *L'idéographie dynamique*, Le Concept Moderne/Editions, Genève (à paraître aux éditions La Découverte).
- MARR, D. (1977) Artificial Intelligence : A personal view, *Artificial Intelligence*, 9, p. 37-48 [repris dans *Philosophy of AI* (éd. M. Boden), p. 133-146, Oxford University Press, 1990].
- MARR, D. (1982) *Vision*, Freeman.
- McDERMOTT, D. (1987) A critique of pure reason, *Computational Intelligence* 3, p. 151-160 [repris dans *Philosophy of AI* (éd. M. Boden), p. 206-230, Oxford University Press, 1990].
- McDERMOTT, D. (1988) Preliminary steps towards a taxonomy of problem solving methods, in *Automating Knowledge Acquisition for Expert Systems* (éd. par S. Marcus), Kluwer Academic Publishers.
- MINSKY, M. (1961) Steps toward artificial intelligence, *Proceedings of the Institute of Radio Engineers*, 49, p. 8-30 [repris dans *Computers and thought* (éds : Feigenbaum & Feldman), 1963, McGraw Hill].
- MINSKY, M. (1981) A framework for representing knowledge, in *Mind Design* (éd. : J. Haugeland), p. 95-128, MIT Press.
- MOULOUD, N. (1968) Article "Modèles" de l'*Universalis* (repris dans l'édition de 1989).
- NEWELL, A. (1980) Physical Symbol Systems, *Cognitive Science*, 4, p. 135-183.
- NEWELL, A. (1982) The knowledge level, *Artificial Intelligence*, 18, p. 82-127.
- NEWELL, A., SIMON, H. (1976) Computer Science as empirical enquiry : symbols and search, *Communications of the ACM* 19 [repris dans *Philosophy of AI* (éd. M. Boden), p. 105-132, Oxford University Press, 1990].
- NOSSIN, M. (1991) La dimension applicative du traitement automatique des langues naturelles, actes du colloque *Informatique et Langue Naturelle*, Nantes.
- PETITOT, J. (1990) Why connectionism is such a good thing. A Criticism of Fodor's and Pylyshyn's Criticism of Smolensky, *Rapports du CREA n° 9020B*, Paris, Ecole Polytechnique.
- PRINCE, V. (1989) L'automatisation de l'expertise peut-elle rendre compte des automatismes des experts ?, *Deuxièmes Journées de Saint-Riquier sur le thème "Connaissances implicites et connaissances explicites"* ; article à paraître dans la *Revue Internationale de Systémique*.
- PROUST, J. (1986) *Questions de forme*, Fayard.
- PYLYSHYN, Z. (1984) *Computation and Cognition*, MIT Press.
- RASMUSSEN, J. (1986) *Information Processing and Human-Machine Interaction : an Approach to Cognitive Engineering*, Amsterdam, North Holland.
- RASTIER, F. (1988) Paradigmes cognitifs et linguistiques universelles, *Intellectica* 6, p. 43-74.
- RASTIER, F. (1989) *Sens et textualité*, Hachette.

- RASTIER, F. (1991) *Sémantique et recherches cognitives*, P.U.F.
- ROSCH, E. (1978) éd. *Cognition and Categorization*, Erlbaum.
- ROSENBLOOM, P., LAIRD, J., NEWELL, A., McCARL, R. (1991) A preliminary analysis of the SOAR architecture as a basis for general intelligence, *Artificial Intelligence*, 47, 1-3, p. 289-325.
- RUMELHART, D. & McCLELLAND, J., édés (1986) *Parallel Distributed Processing: The Microstructure of Cognition*, MIT Press.
- SALANSKIS, J.M. (1990) Le concept de Gestalt et la situation contemporaine de la philosophie des sciences, *Les Etudes philosophiques*, 4, p. 519-536.
- SCHANK, R., COLLINS, G., HUNTER, L. (1986) Transcending inductive category formation in learning, *The Behavioral and Brain Sciences* 9, p. 639-686 [suivi de la réponse des auteurs aux commentaires].
- SCHANK, R., RIEGER, C. (1974) Inference and the computer understanding of natural language, *Artificial Intelligence* 5, p. 373-412.
- SHOHAM, Y. (1988) Non monotonic logics : meaning and utility, article repris dans son livre *Reasoning about Change : Time and Causation from the Standpoint of AI*, MIT Press.
- SMITH, B., dir. (1988) *Foundations of Gestalttheorie*, München-Wien, Philosophia Verlag.
- SMOLENSKY, P. (1988) On the proper treatment of connectionism, suivi de la réponse de l'auteur aux commentaires, *The Behavioral and Brain Sciences* 11, p. 1-74.
- SMOLENSKY, P. (1989) Connectionism and constituent structure, in *Connectionism in Perspective* (édés.: R. Peifer, Z. Schreter, F. Fogelman-Soulié, L. Steels), p. 3-24, North-Holland.
- SOWA, J. (1984) *Conceptual Structures : Information Processing in Mind and Machine*, Addison Wesley.
- SUCHMAN, L. (1987) *Plans and Situated Actions : The Problem of Human Machine Communication*, Cambridge U.P.
- TOURETZKY, D., HINTON, G. (1988) A distributed connectionist production system, *Cognitive Science* 12, p. 423-466.
- TURING, A. (1950) Computing Machinery and Intelligence, *Mind*, vol. LIX, n° 236.
- VICTORRI, B. (1988) Modéliser la polysémie, *T.A. Informations*, 1-2, p. 21-42.
- VICTORRI, B., RAYSZ, J.P., KONFE, A. (1989) Un modèle connexionniste de la polysémie, *Actes de Neuro-Nîmes*, p. 97-108, EC2.
- VOGEL, C. (1988) *Génie cognitif*, Masson.
- WENGER, E. (1987) *Artificial Intelligence and Tutoring Systems : Computational and Cognitive Approaches to the Communication of Knowledge*, Morgan Kaufmann.
- WINOGRAD, T., FLORES, F. (1986) *Understanding Computers and Cognition*, Ablex Publishing Corporation.