



L'ingénierie des connaissances, entre science de l'information et science de gestion

Jean Charlet

► **To cite this version:**

Jean Charlet. L'ingénierie des connaissances, entre science de l'information et science de gestion. Entre la connaissance et l'organisation, l'activité collective, 2004. <sic_00000805>

HAL Id: sic_00000805

https://archivesic.ccsd.cnrs.fr/sic_00000805

Submitted on 12 Nov 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'ingénierie des connaissances, entre science de l'information et science de gestion*

Jean Charlet[†]

Résumé

Dans cet article, nous voulons expliciter et analyser les méthodes de l'ingénierie des connaissances. Nous voulons aussi étudier la position des outils de l'ingénierie des connaissances par rapport à l'organisation et aux sciences de gestion. Pour cela, nous pensons qu'il est nécessaire de revenir sur les fondements des systèmes à base de connaissances issus de l'ingénierie des connaissances et de l'intelligence artificielle pour positionner l'ingénierie des connaissances au sein des sciences de l'artificiel telles que décrites par H. Simon et tirer tous les bénéfices d'un travail de réflexion épistémologique, historique et pragmatique.

Nous commencerons par aborder les problèmes épistémologiques de l'ingénierie des connaissances et nous utiliserons les résultats obtenus pour proposer un programme et des méthodes de travail pour l'ingénierie des connaissances. Nous mettrons ensuite ce programme et ces méthodes à l'épreuve en essayant d'analyser, à l'aide d'exemples en ingénierie ontologique et sur le dossier médical, la portée des propositions faites. Enfin, nous essaierons d'abord de situer l'ingénierie des connaissances dans une perspective historique pour comprendre son statut scientifique et la validité des connaissances qu'elle nous permet d'acquérir puis nous essaierons ensuite de positionner les outils de l'ingénierie des connaissances par rapport aux sciences de gestion pour étudier les apports réciproques de ces 2 disciplines dans la constitution des outils qu'elles ont vocation à mettre en œuvre.

Mots-clés : sciences de gestion, gestion des connaissances, ontologies, épistémologie.

*Cet article est la version allongée d'un chapitre de livre coordonnée par R. Teulier et Ph. Lorino [2005] et faisant suite au colloque de Cerisy « Activité, connaissance, organisation ».

[†]STIM/DSI/AP-HP, & INSERM ERM 202, Jean.Charlet@spim.jussieu.fr

Table des matières

1	Introduction	3
2	Épistémologie et définitions	5
2.1	De l'information à la connaissance	5
2.1.1	Point de vue épistémologique	5
2.1.2	Point de vue systémique	6
2.2	Des systèmes à base de connaissances	7
3	Un programme de travail pour l'ingénierie des connaissances...	8
3.1	En guise de programme	9
3.2	En guise de méthode	9
3.3	En guise de champ d'action	10
4	... applicable et appliqué ?	10
4.1	HOSPITEXTE	10
4.1.1	Arguments	11
4.1.2	Du support papier au support électronique	11
4.1.3	Un prototype et des perspectives	12
4.1.4	Le point de vue de l'ingénierie des connaissances	13
4.2	Des thésaurus à la construction d'ontologies	13
4.2.1	Des ontologies en Ingénierie des connaissances	13
4.2.2	Des taxinomies	15
4.2.3	Des thésaurus en médecine	16
4.2.4	Une méthodologie de construction d'ontologies	16
4.2.5	Le point de vue de l'ingénierie des connaissances	18
5	Réflexions et discussion	19
5.1	Introduction	19
5.2	Un domaine scientifique, un enseignement ?	19
5.2.1	Des épistémologies positivistes aux épistémologies constructivistes	20
5.2.2	L'ingénierie des connaissances	21
5.3	Une ingénierie des inscriptions numériques des connaissances	23
5.4	La généralité des outils	24
6	Conclusion	25

1 Introduction

L'ingénierie des connaissances constitue depuis de nombreuses années un domaine actif des recherches menées en intelligence artificielle autour de la conception et de la réalisation des systèmes à base de connaissances (SBC). À l'instar de bien d'autres disciplines modélisatrices, elle consiste à concevoir des systèmes dont le fonctionnement permet d'opérationnaliser des connaissances portant sur le traitement ou la résolution d'un problème donné. La résolution (semi-)automatique de problèmes implique deux étapes essentielles : la modélisation du problème et d'une méthode de résolution dans un cadre théorique donné, l'opérationnalisation informatique du modèle obtenu. Longtemps, le cadre théorique de l'ingénierie des connaissances fut celui de l'acquisition des connaissances : modélisation psychologique ou empirique des connaissances d'un expert dans le but de les coder dans un système expert (SE).

La période actuelle se concentre davantage sur la modélisation conceptuelle du monde : on tente, le plus souvent à partir d'une formulation linguistique du problème – transcriptions d'interviews d'experts, descriptions techniques, notices de maintenance, etc. –, d'élaborer une représentation qualitative et formelle du problème. Il arrive également qu'on obtienne une première représentation du problème à partir d'observations de l'activité faites, par exemple, avec des méthodes issues de l'ergonomie ou de l'ethnométhodologie¹.

Définie en ces termes, il paraît naturel qu'une recherche menée sur les SBC comprenne une activité correspondant à l'ingénierie des connaissances. En effet, il s'agit de définir quelles connaissances sont pertinentes pour les exploiter en formulant un modèle qualitatif et formel du problème. Classiquement, l'ingénierie des connaissances propose des méthodes permettant de modéliser des raisonnements et le domaine d'activité dans lequel il s'insère. Un modèle conceptuel du SBC est ainsi construit : défini comme la description abstraite du comportement d'un système, en termes de domaine, méthodes de raisonnement, tâche dévolue au système, ce modèle conceptuel recouvre donc une réalité bien précise, différente de celle communément admise en Systèmes d'information²(SI). Malgré son abstraction, ce modèle conceptuel du problème doit être compréhensible et réappropriable par un spécialiste du domaine³. Cette description abstraite est réalisée au « niveau des connaissances » [Newell, 1982]. Enfin, un modèle opérationnel, défini « au niveau des programmes », traduit en termes algorithmiques les connaissances formalisées du modèle conceptuel. Ce dernier est à la fois la spécification et l'instrument de lecture du système opérationnel.

¹Le recours au texte, ou corpus, comme matière première n'est pas la pensée unique de l'ingénierie des connaissances : l'ergonomie, la sociologie sont des vecteurs d'entrée parmi d'autres. Mais deux constats doivent être faits : a) la formulation linguistique fournit une description qualitative du monde qui contribue à déterminer, en contexte, le contenu de l'information véhiculée et fait des textes le mode privilégié d'accès aux connaissances ; b) un certain nombre d'avancées méthodologiques dans ce domaine font de l'acquisition des connaissances à partir de corpus une des directions de travail prometteuse de l'ingénierie des connaissances [Bourigault, 2000; Bourigault *et al.*, 2004].

²De nombreuses personnes se sont préoccupé de ce modèle conceptuel et de son statut. Parmi une bibliographie fournie, nous citerons simplement des travaux qui proposent des définitions de base sur les modèles conceptuels [Aussenac-Gilles *et al.*, 1992], des réflexions sur les rôles au sein de ces modèles [Aussenac-Gilles *et al.*, 1997] et les travaux autour de la méthodologie KADS [van Heijst *et al.*, 1997] qui développent une large vue sur la question et plus précisément sur les rapports qu'entretiennent les connaissances du domaine et les connaissances de raisonnement. Cette courte bibliographie n'épuise évidemment les débats qui ont cours, en particulier sur la réutilisabilité des différentes parties du modèle conceptuel.

³Il y a une certaine contradiction entre un modèle compréhensible par un expert du domaine et un raisonnement modélisé à un niveau générique où il est fait abstraction du domaine. Ce problème est à l'origine de bien des incompréhensions des experts devant les modèles « génériques » de l'ingénierie des connaissances [Lépine & Aussenac-Gilles, 1996]. Des réponses sont recherchées dans diverses adaptations des modèles génériques et dans des modèles intermédiaires tels les bases de connaissances terminologiques [Condamines & Aussenac-Gilles, 2001].

Les recherches en Ingénierie des connaissances ayant évolué ces dernières années, elles concernent aujourd'hui tout système informatique utilisant des connaissances pour peu que ces connaissances soient explicitement modélisées en tant que telles. Par ailleurs, les travaux développés en ce domaine tirent parti d'apports d'autres domaines comme le génie logiciel et, inversement, les réflexions et techniques d'ingénieries des connaissances mises au point dans le cadre de la conception de SBC, sont réutilisables dans des contextes plus larges, en particulier au sein des SI. Ainsi, tout système, *a fortiori* SI, qui manipule et transporte des informations destinées *in fine* à être interprétées par des humains est candidat à être conceptualisé et modélisé avec les méthodes de l'ingénierie des connaissances⁴. Dans ce cadre, la question des méthodes concourant à l'élaboration des SBC et la réutilisabilité de ces mêmes méthodes est centrale pour les SI comme pour les SBC. En effet, une discipline qui se veut une ingénierie – des connaissances ou des SI – doit apporter, aux utilisateurs, des réponses en termes de méthodes et techniques reproductibles.

Enfin, les systèmes experts anciennement, les SBC maintenant ont, souvent implicitement, vocation à être mis en place dans des organisations, au sens où Ph. Lorino [2005] parle d'organisation, c'est-à-dire, *L'organisation à laquelle les chercheurs en gestion s'intéressent principalement, l'entreprise produisant des richesses économiques, est une institution spécifique, dont l'apparition et le développement constituent un processus historique qui peut être daté et décrit dans le mouvement général de la société [...], avec ses dimensions économiques, technologiques, juridiques, sociologiques*. J.-C. Moisdon s'intéressant à la doctrine d'usage des outils de gestion, étend l'objet d'étude aux administrations, en particulier hospitalière puisque c'est son cas d'étude dans [Moisdon, 2005]. Ainsi, l'institution hospitalière, le système médical français rentre parfaitement dans les organisations étudiées par les sciences de gestion et l'insertion des systèmes experts en médecine, déjà vu comme un outil de gestion, a pu être étudiée et l'échec de cette insertion expliquée par des chercheurs en gestion [Hatchuel & Weil, 1992]. C'est l'hypothèse, claire dans cette dernière étude, du système expert comme outil de gestion que nous allons essayer de réactiver au niveau des SBC⁵.

Depuis cette période, les outils de l'intelligence artificielle et maintenant de l'ingénierie des connaissances se positionnent différemment des SE dont l'échec est dû, entre autres, à une formalisation mal assumée et à une insertion organisationnelle pas toujours réfléchie. En effet, l'ingénierie des connaissances vise une « représentation » des connaissances nécessaires à un artefact informatique pour réaliser la tâche qui lui est dévolue par ses concepteurs. Cette représentation crée une tension entre une formalisation nécessaire au fonctionnement de l'artefact – *e.g.* un SBC – et l'obligation de conserver le contexte d'énonciation des connaissances nécessaire à l'interprétation du fonctionnement de l'artefact par l'utilisateur et à son appropriation. Cette tension trouve des amorces de solution dans des systèmes conservant le contexte d'énonciation des connaissances, le texte lui-même, et est rendue possible par le développement de l'ingénierie documentaire comme nous le verrons dans un exemple (*cf.* 4.1).

Dans cet article, nous voulons expliciter et analyser les méthodes de l'ingénierie des connaissances. Nous voulons aussi étudier la position des outils de l'ingénierie des connaissances par rapport à l'organisation et aux sciences de gestion. Pour cela, nous pensons qu'il est nécessaire de revenir sur les fondements des SBC pour tirer tous les bénéfices d'un travail de réflexion épistémologique, historique et pragmatique.

⁴« Candidat » est utilisé ici littéralement. Nous pensons que c'est une évolution inéluctable des SI que d'être pensés en termes de connaissances [Charlet *et al.*, 2001; Ermine, 1996] même si cette évolution n'est qu'esquissée à ce jour.

⁵Sur la question des indicateurs médico-économiques – *i.e.* le PMSI ou Programme de médicalisation du système d'information – et de leur usage comme outil de gestion (apprentissage, détournement, ...), au sein du système hospitalier, nous ne serions trop recommander la lecture de [Moisdon, 2005].

La 2^e section nous permettra d'aborder des problèmes épistémologiques de l'ingénierie des connaissances. La 3^e section, se fondera sur les hypothèses et résultats développés précédemment pour proposer un programme et des méthodes de travail pour l'ingénierie des connaissances. Dans la 4^e section, nous mettrons ce programme et ces méthodes à l'épreuve en essayant d'analyser, à l'aide d'exemples, la portée des propositions faites. Enfin dans la 5^e section, nous essaierons d'abord de situer l'ingénierie des connaissances dans une perspective historique pour comprendre son statut scientifique et la validité des connaissances qu'elle nous permet d'acquérir. Nous essaierons ensuite de positionner les outils de l'ingénierie des connaissances par rapport aux sciences de gestion pour étudier les apports réciproques de ces 2 disciplines dans la constitution des outils qu'elles ont vocation à mettre en œuvre.

2 Épistémologie et définitions

L'ingénierie des connaissances se situe au carrefour de plusieurs réflexions : la linguistique pour étudier la formulation linguistique des connaissances ; la terminologie et les recherches sur la genèse des ontologies pour dégager les concepts ; la psychologie pour élaborer les méthodes d'élicitation et les modèles d'assistance ; la logique pour élaborer les modèles formels ; l'informatique pour les opérationnaliser ; l'ergonomie pour interpréter et s'approprier le comportement du système ; la gestion pour concevoir et replacer les systèmes dans leur environnement organisationnel, etc. La plupart de ces disciplines ressortissent à la sémiotique⁶, la science des signes . Elles se fécondent réciproquement et conduisent à faire évoluer le paradigme. Nous allons donc mobiliser un certain nombre d'entre elles pour aborder successivement les concepts de connaissances, de SBC et de SI.

2.1 De l'information à la connaissance

Définir la connaissance en soi est une entreprise philosophique qui n'est pas de notre propos ici. Par contre, nous allons essayer de caractériser cette connaissance, d'un point de vue effectif, qui est celui qui nous intéresse. Nous l'aborderons d'abord d'un point de vue dit « épistémologique » puis d'un point de vue « systémique ». Il n'y a pas de séparation étanche entre les points de vues définis ici. Cela tient plus aux domaines qui ont développé et historiquement exploités les constats faits, ingénierie des connaissances (et donc intelligence artificielle) d'un côté, systèmes d'information de l'autre, qu'à une quelconque préséance ou qualité de réflexion. Enfin, si les caractéristiques attestées ici traduisent des fondements bien acceptés par les deux communautés, elles ne sont pas toujours en accord avec d'autres courants de l'intelligence artificielle qui font l'hypothèse d'une pensée structurée formellement ou computationnellement. Pour des débats et arguments à ce sujet, nous renvoyons le lecteur à [Bachimont, 1996].

2.1.1 Point de vue épistémologique

D'un point de vue épistémologique, nous retiendrons trois caractéristiques de la connaissance que l'on peut résumer ainsi [Bachimont, 1996; Ganascia, 1998; Kayser, 1997] :

⁶La sémiotique est comprise ici dans une acceptation très large et, dans cette hypothèse, il n'est pas étonnant de la voir englober des disciplines qui participent d'une science – ce que nous essayons de montrer dans cet article –, l'ingénierie des connaissances, dont le but est d'aider à construire des systèmes informatiques donc systèmes de symboles. Seule, parmi les disciplines citées ici, la gestion se situe dans un cadre applicatif. Elle apparaît ici comme l'aune à laquelle seront mesurés les résultats obtenus, dans l'aptitude qu'auront les SBC construits à s'insérer dans l'environnement organisationnel pour lequel ils ont été conçus [Charlet *et al.*, 2001].

1. La connaissance n'est possible que dans un environnement *technique* qui est le notre, qu'il apparaisse fortement technique (industrie, monde professionnel), ou non (monde quotidien, ...). L'informatique et les calculateurs sont un élément de notre environnement technique. Dans notre recherche et définition d'une ingénierie de la connaissance, l'informatique, le support qu'elle représente, les calculs qu'elle autorise est alors la technique qui permet la mémorisation de la connaissance⁷.
2. Il y a connaissance et représentation des connaissances quand les manipulations symboliques effectuées par la machine via des programmes, prennent un sens et une justification pour les utilisateurs interagissant avec ces programmes. Les utilisateurs interprètent alors le comportement de la machine.
3. La pensée repose sur la médiation externe du signe. La technique (ici l'informatique), par ses outils et ses capacités de mémorisation, permet alors à l'homme de se constituer des connaissances qui évoluent et s'accumulent et n'existeraient pas sans cela⁸.

Ainsi, les deux concepts clés de la connaissance sont l'interprétation humaine qui lui donne son existence et, dans notre domaine, l'outil informatique qui lui offre son support de mémorisation comme de genèse.

2.1.2 Point de vue systémique

D'un point de vue plus « systémique » et parce que les concepts de données, d'information ont été historiquement abordés par l'informatique et les Systèmes d'information, on va s'intéresser à ces concepts, les caractériser les uns par rapport aux autres et constater, qu'ici aussi, il est possible de proposer quelques caractérisations clés qui font consensus [Ermine, 1996; Ganascia, 1998; Le Moigne, 1973, 1995; Mèlèse, 1990; Poitou, 1996; Shannon & Weaver, 1971] et sont respectées par les travaux qui seront cités par la suite. Ainsi, les concepts de données, informations, processus, connaissances peuvent être considérés en interactions selon les critères suivants :

- L'information est un concept technique apparu pour les besoins des télécommunications. L'information fait appel aux concepts de codage, transmission, décodage et fait référence au nombre potentiel de messages que peut délivrer un système.
- Une donnée est toute information affectant un programme ou un système pour en modifier le comportement. Le processus d'utilisation est fixe, au contraire des données qui évoluent et se renouvellent. Mais un programme ou processus peut être lui-même la donnée d'un autre processus.

⁷Dans cette note, nous résumons très rapidement les arguments épistémologiques et philosophiques qui fondent ces affirmations. Nous reprenons par là les thèses de B. Bachimont, auquel nous renvoyons pour de plus amples discussions [1996, chap. 7, § 3]. La technique et les *outils* de la technique permettent à l'homme de mémoriser des connaissances de façon externe, sans plus avoir la nécessité de les mémoriser en interne. Il en est par exemple d'un couteau ou d'un marteau qui prescrivent, par leur existence et leur forme, leur usage [Leroi-Gourhan, 1964]. La technique est alors le support de la connaissance. En étendant la technique à « toute extériorité spatiale et matérielle dont la structure prescrit l'usage ou l'interprétation sans les déterminer » [Bachimont, 1996, chap. 7, § 3.2], l'informatique peut être vue comme une technique qui explicite du sens.

⁸Cette troisième caractéristique est une conséquence de la première (*cf.* note 7) : par les mises en correspondances qu'elle permet, l'informatique est une technique du calcul qui crée du sens, y compris sans intentionnalité (*Ibid.*). On sait que l'écriture a permis de classer les mots en listes, de combiner les listes en tableaux dans un paradigme appelé par J. Goody [1979] « la raison graphique ». L'informatique propose alors une nouvelle rationalité, celle de la « raison computationnelle » [Bachimont, 2000b].

- Il n’y a pas de frontière donnée/information/processus/connaissance. Nous sommes devant un continuum par rapport à un processus d’action et nous plaçons des étiquettes sur des concepts manipulés par ce processus en fonction des niveaux d’entrée dans celui-ci.

La connaissance peut alors être caractérisée de la façon suivante :

1. Il y a présomption de connaissances, si la faculté d’utiliser l’information à bon escient est attestée. Cette utilisation passe d’abord par une interprétation puis par une action. Ce qui fait dire à Ph. Lorino [1995] que « l’autonomie d’un acteur est une marge d’interprétation pour l’action ». Enfin, cette action peut être l’extériorisation de la connaissance par l’utilisateur d’un outil, ici une réécriture par l’informatique⁹.
2. Il y a connaissance quand il y a contexte d’utilisation de l’information. Cela veut dire qu’une information n’a pas de valeur isolément mais en fonction du contexte dans lequel elle est élaborée et interprétée.
3. La connaissance *a priori* n’existe pas : elle est construite à partir d’un projet propre au modélisateur. Étudier cette construction permet d’essayer d’en dégager des invariants méthodologiques. C’est le projet des épistémologies constructivistes [Le Moigne, 1995].

La prise en charge des connaissances par le SBC va alors être la capacité qu’il aura à proposer des données, sources d’interprétation par l’utilisateur, à prendre en compte et expliciter le contexte d’utilisation de ces données, à fournir à cet utilisateur les moyens – informatiques – d’agir et donc de réécrire les résultats de son interprétation. Cette capacité d’interaction homme-machine mettant alors l’accent sur l’utilisateur (savoir, savoir-faire, modes d’interaction) et la prise en compte du changement dans l’action personnelle et l’action collective des organisations.

2.2 Des systèmes à base de connaissances

Ces caractéristiques étant posées, il est alors possible de s’intéresser aux SBC et à leur place dans les organisations. Les SBC manipulent des représentations symboliques selon des prescriptions formalisées lors de la modélisation des connaissances. Ces représentations s’expriment à l’aide de primitives qui renvoient à des notions du domaine en leur empruntant leur libellé ou terminologie : les primitives sont les termes du domaine¹⁰. Ces primitives sont manipulées en respectant la grammaire du système formel dans lequel elles s’inscrivent. Ainsi, toutes les manipulations effectuées consistent dans la construction syntaxique de représentations mobilisant ces termes : ces représentations peuvent et doivent se rapporter à des expressions interprétables dans le domaine par tout spécialiste. Mais les règles de cette manipulation formelle ne sont pas celles de l’interprétation : le calcul produit des résultats que la rationalité de l’interprétation n’anticipe pas forcément. La combinatoire des expressions suggère alors la possibilité de formuler des inscriptions dont l’interprétation renvoie à de nouvelles connaissances. De même que l’ordinateur ne « voit » pas les images qu’il permet de construire, il ne « pense » pas les nouvelles inscriptions qu’il formule. Mais il permet de voir du nouveau comme de penser autrement .

Pour concevoir un SBC, il faut donc tenir compte de la manière dont des utilisateurs s’en approprient et lui attribuent du sens en interprétant, sur la foi des primitives empruntées à la terminologie

⁹Cette réécriture ne préjuge pas, quand on quitte la sphère informatique, d’une action réelle effectuée par un l’agent qui interprète le comportement du système informatique et les réécritures qu’il effectue.

¹⁰Cette affirmation doit être tempérée et surtout précisée : il ne s’agit pas de poser l’équivalence entre les termes du domaine et les primitives conceptuelles qui évoluent dans deux registres différents, linguistique et conceptuel. En particulier, les termes du domaine ne peuvent prétendre être instantanément des primitives conceptuelles [Bachimont, 2000a] (cf. note 13).

du domaine, les représentations comme des expressions linguistiques de connaissances. Cette interprétation repose sur le contexte d'usage du SBC et son intégration dans un système de pratiques où il prend sens et justification . *Ainsi, il faut voir les SBC comme des systèmes sémiotiques de manipulation d'inscriptions symboliques, dont le fonctionnement informatique doit permettre à un utilisateur d'interpréter et de comprendre le système dans le cadre de son activité et de ses usages, en utilisant les termes du domaine* [Charlet & Bachimont, 1998]¹¹.

Un SBC étant un système technique plongé dans un système d'usage, son élaboration est une ingénierie (au sens où un ingénieur élabore un système pour un usage), une ingénierie fondée sur la manipulation de l'inscription symbolique de connaissances. Nous pouvons alors avoir un point de vue opérationnel sur l'ingénierie des connaissances : *l'ingénierie des connaissances correspond à l'étude de modèles symboliques formels plongés dans des systèmes d'usage ; c'est l'ingénierie informatique et logique de modèles en fonction des usages qu'ils rendent possibles et des appropriations qu'ils permettent*¹². Ainsi, l'ingénierie des connaissances peut consister à mettre en perspective des outils formels ou techniques – *p. ex.* logiques de description, langages de description documentaire (SGML/XML), etc. – avec des concepts d'usage, issus ou non des sciences cognitives – *p. ex.* le raisonnement classificatoire, la navigation documentaire, l'exploration d'hypothèses dans un domaine donné, etc.

3 Un programme de travail pour l'ingénierie des connaissances...

Ces deux derniers points de vue, conceptuel et opérationnel, sont complémentaires et l'ingénierie des connaissances ne peut mobiliser l'un sans l'autre au risque de perdre sa spécificité, son efficacité et l'objet même de sa recherche : d'un point de vue conceptuel, l'ingénierie des connaissances trouve l'objet de sa recherche dans des systèmes dont le fonctionnement est interprétable par l'humain en termes de connaissances ; d'un point de vue opérationnel, l'ingénierie des connaissances intègre ces réflexions dans des systèmes informatiques – *e.g.* systèmes d'information – qu'il faut modéliser, concevoir, bâtir et faire évoluer.

Ces deux points de vue ne sont pas indépendants : le système informatique utilisé par un acteur, le système d'information utilisé par une collectivité modifient le comportement et l'organisation de l'acteur comme de la collectivité. Il nous semble alors nécessaire de penser la technique à travers les usages puisque un système est conçu pour une pratique, et les usages à travers la technique puisque les pratiques n'existent qu'à travers les techniques qui les rendent possibles, .

Ces réflexions doivent nous permettre de proposer un programme de travail pour l'ingénierie des connaissances. Ce « programme » doit être compris au sens large, c'est-à-dire avec ses corollaires que sont les méthodes de travail et les champs d'application. Enfin, ce sont les définitions et les caractérisations de la connaissance qui vont nous servir de jalons dans la description de ce programme .

¹¹La citation ici rend mal compte de l'origine de cette proposition due à B. Bachimont durant les réflexions et travaux communs qui ont amené la susdite publication.

¹²On pourrait dire « c'est l'ingénierie informatique "ou" logique de modèles [...] », les travaux sur l'ergonomie ou interfaces homme-machine (IHM) ne semblant pas ressortir à des problèmes formels. En pratique, on peut au contraire considérer que tous travaux étudiant des systèmes informatiques par rapport à leur usage ressortissent à une problématique formelle au sens large (c.-à-d. par rapport à une formalisation pas seulement mathématique) : ainsi, en étudiant l'ergonomie d'IHM on est obligé de caractériser les comportements du système informatique en fonction des régularités, des combinaisons et appropriations qu'il permet. Cela passe *in fine* par une caractérisation syntaxique ou formelle de ces systèmes y compris dans le cadre d'IHM pour le travail coopératif où le sens vient *p. ex.* de la succession, de la juxtaposition ou de la persistance, syntaxiquement organisées, d'écrans.

3.1 En guise de programme

Ainsi, une démarche d'ingénieries des connaissances peut se caractériser en fonction de 3 objectifs :

- L'objet de la recherche sont les systèmes dont le fonctionnement est interprétable par l'humain en termes de connaissances . Cela veut dire que l'on cherche à construire un SBC ayant un comportement compréhensible par l'utilisateur et acceptable, cognitivement parlant. Cela passe par une modélisation à un niveau d'abstraction pertinent qui fait sens pour les différents acteurs impliqués (ingénieur cognitivistes, experts métiers, utilisateurs, etc.) et leur permet de s'approprier le comportement du système et d'interagir avec lui. C'est le « niveau des connaissances » (*cf. supra*). On utilise pour cela l'une des principales caractéristiques des systèmes formels : les symboles primitifs sur lesquels ils reposent, sont des unités signifiantes emprunté à la langue naturelle [Bachimont, 2000b]¹³.
- Le but correspond alors à l'intégration de ces réflexions dans des artefacts informatiques à modéliser, construire, faire évoluer . L'artefact informatique doit être, ici, compris au sens large ; c'est-à-dire que cela peut-être aussi bien de la documentation non structurée que des descriptions de méthodologies ou de protocoles ou un système formalisant des raisonnements. . . *pour peu que dans cet artefact, nous ayons besoin et nous cherchions explicitement à modéliser des connaissances.*
- Selon sa nature et les buts poursuivis, l'artefact informatique peut modifier les usages singuliers comme collectifs, jusqu'aux organisations . Le fonctionnement et l'utilisation du SBC doivent alors être appréhendés par rapport à son intégration dans une organisation. Dans ce contexte, l'ingénierie des connaissances doit permettre de modéliser et d'explicitier le fonctionnement cognitif de l'utilisateur ou du collectif dans sa relation au SBC.

3.2 En guise de méthode

De la même façon, nous allons mobiliser les points de vues précédents sur la connaissance pour mettre l'accent sur 3 critères méthodologique cruciaux :

- La connaissance étant contextuelle, se donner les moyens de représenter ce contexte dans une démarche de représentation des connaissances. Cela peut amener à des choix de formalisation minimum comme dans les projets médicaux HOSPITEXTE (*cf. infra*) et ONCODOC [Bouaud *et al.*, 1998].
- La connaissance, toujours contextuelle, ne pouvant être indépendante de la tâche à réaliser, se donner les moyens d'assumer explicitement cette dépendance (réutilisation limitée). Cela peut se traduire, dans le cadre des ontologies, par des choix méthodologiques forts figeant les contextes d'utilisation.
- La connaissance étant liée à l'action, veiller à l'usage qui est fait des représentations ; en particulier qu'elles sont bien mémorisées dans un but, pour un usage prévu.

¹³ Cette confusions des formes entre signifiés linguistiques et labels utilisés dans les systèmes formels (comme les ontologies) sont à l'origine de bien des surprises et erreurs (*cf. note 10, p. 7*). Pour préciser, nous voulons dire par là que le concept formel a une signification qui vient de sa situation dans le système de concepts construit, de sa place dans la hiérarchie pour une conceptualisation ontologique. Son sens ne vient pas du label qui lui est donné même si par souci de lisibilité nous utilisons le mot de la langue le plus proche de la signification du concept. C'est cette lisibilité recherchée qui peut créer une confusion. Par exemple, le concept de sténose dans une ontologie médicale peut avoir une acceptation plus réduite (en général) que le sens du mot sténose dans des comptes rendus d'hospitalisation (*cf. 4*).

3.3 En guise de champ d'action

L'ingénierie des connaissances propose des concepts, méthodes et techniques permettant de modéliser, de formaliser, d'acquérir des connaissances dans les organisations dans un but d'opérationnalisation, de structuration ou de gestion au sens large. Ces mêmes connaissances sont des informations destinées à être, in fine, interprétées par un humain, dans son interaction avec l'artefact – *i.e.* le SBC – construit. Elles délimitent ainsi le champ d'action de l'ingénierie des connaissances qui se focalise sur les artefacts informatiques pour lesquels nous avons besoin et cherchons explicitement à modéliser des connaissances. Par ailleurs et parce que nous cherchons à construire des modèles et des méthodes aussi réutilisables que possible, il nous faut chercher à mettre en œuvre une réelle démarche d'ingénierie, reproductible et proposant des instrumentations techniques – *i.e.* des méthodes et des outils – favorisant la dynamique des connaissances dans l'organisation.

Les applications d'une telle démarche sont là où l'on veut modéliser les connaissances, les processus cognitifs, les processus de coopération et les savoir-faire d'un métier. C'est pourquoi l'ingénierie des connaissances a vocation à se trouver associée à de nombreuses disciplines :

- d'une part, par sa fonction d'ingénierie où elle trouve des échos dans des spécialités comme la conception de systèmes d'information, les méthodes d'analyse et de conception à objets, le raisonnement à base de cas, l'ingénierie documentaire ou l'ingénierie éducative ;
- d'autre part, en se rapprochant de disciplines abordant les mêmes objets de recherche et produisant des connaissances complémentaires aux siennes comme la sociologie, la gestion ou l'ergonomie qui peuvent concourir à une démarche d'explicitation de ce que sont les connaissances dans un contexte humain et organisationnel.

4 ... applicable et appliqué ?

Je voudrais illustrer ces propositions par 2 exemples de recherches qui illustrent justement le programme et la méthode de l'ingénierie des connaissances. Ces exemples sont pris parmi les projets auxquels j'ai collaboré puisque ce sont ceux que je connais le mieux. Faute de place, je ne décris pas d'autres travaux de l'ingénierie des connaissances en accord avec les réflexions développées ici ¹⁴.

4.1 HOSPITEXTE

Développé par B. Bachimont, V. Brunie et moi-même, le projet HOSPITEXTE [Brunie *et al.*, 2000] a consisté à réaliser un plate-forme d'expérimentation d'informatisation du dossier patient en suivant 2 objectifs : *a)* création d'un dossier patient virtuel en utilisant les capacités du Web à permettre la « reconstruction » d'un unique dossier patient à partir d'informations résidant dans d'autres services ou à l'extérieur d'un hôpital et *b)* élaboration d'une station de travail professionnelle où les capacités de calcul de l'informatique sont utilisées pour résoudre les problèmes de surcharge informationnelle et de désorientation. Les motivations et arguments épistémologiques d'un tel projet sont forts.

¹⁴Comme les ouvrages [Aussenac-Gilles *et al.*, 1996], [Charlet *et al.*, 2000b] et [Teulier *et al.*, 2004] en illustrent bien le développement.

4.1.1 Arguments

La plupart des expérimentations visant à informatiser le dossier médical du patient ont montré, par leurs échecs, que l'information médicale enregistrée dans des bases de données accessibles à distance conserve beaucoup moins bien ses caractéristiques contextuelles que l'information médicale supportée par le dossier papier qui reste dans son environnement d'origine. Une autre solution consiste à considérer que la forme linguistique et documentaire d'un dossier patient telle qu'elle se constate dans l'hôpital ou la clinique n'est pas gratuite et contribue à déterminer le contenu médical de l'information véhiculée dans le susdit dossier. Il est alors illusoire de vouloir répertorier les faits médicaux pertinents indépendamment du formatage linguistique et documentaire [Bachimont, 2001]. Ainsi, une information médicale n'a pas de sens médical par elle-même mais est fonction du contexte [Charlet *et al.*, 1999]. La langue naturelle et les genres textuels du dossier permettent d'exprimer l'information dans son contexte de création et de la proposer au praticien dans une forme qui permet l'interprétation ou la réinterprétation et la génération de nouveaux documents (p. ex. compte rendu de sortie). C'est pour cela que le dossier médical d'un patient n'est, en dehors des images, qu'une accumulation de textes comme des comptes rendus ou des notes de suivi ou des résultats d'exams complémentaires, sous des formes et avec des mises en page constantes. Pour conclure sur ce thème, il n'est possible de « servir » l'information médicale recherchée que dans le format documentaire pertinent et culturellement lisible et exploitable pour un utilisateur médical : *le document lui-même*.

4.1.2 Du support papier au support électronique

HOSPITEXTE consiste ainsi dans le passage du support papier au support informatique du dossier patient avec conservation de sa structure documentaire. On espère ainsi préserver le contenu médical du dossier papier. Ceci ne serait immédiatement possible que si la nature matérielle du support était neutre quant au contenu informationnel du dossier. Or cela n'est pas le cas. On a pu mettre en évidence que les caractéristiques physiques liées à la nature du support matériel, en général papier, sont exploitées par le praticien qui effectue la lecture et influencent par conséquent le contenu informationnel qu'il en retire. En effet, les anciennes hospitalisations correspondent à des documents papier jaunis qui sont systématiquement mis de côté, p. ex. lors d'une première lecture. Les liasses documentaires de fine épaisseur correspondent à des hospitalisations de routine, etc. Enfin, un dossier patient peut être étalé sur une table et sa disposition sur le plan de la table donne une vue synoptique du dossier permettant au praticien de construire son parcours de lecture quasi instantanément [Tange, 1995]. Par conséquent, si l'information médicale est d'une certaine manière neutre quant à son support informatique ou papier, il apparaît que l'accès à cette information est largement conditionné à la nature de ce support.

Par ailleurs, une consultation hypertextuelle est une navigation visant à atteindre un objectif de lecture [Nygren & Henriksson, 1992]. Au cours de sa lecture, l'utilisateur définit lui-même, par des techniques annotatives¹⁵, quelles sont les informations médicales et indique leur nature. En d'autres termes, ce qu'est une information (médicale) ne se détermine qu'à l'occasion d'une lecture, en fonction d'un objectif de lecture déterminé. En somme, il en est du dossier comme des textes en général : on ne sait ce qu'un mot veut dire, comment se délimitent les unités linguistiques, que lorsqu'on a lu et compris la phrase en son entier. Il résulte ainsi qu'une information médicale est l'aboutissement d'un « parcours lectorial » et non sa condition. Le dossier ne se réduit pas à une somme d'informations médicales que l'on aurait mises en forme pour en faciliter la consultation.

¹⁵Comme on le voit dans le dossier papier, où de nombreux documents sont émaillés de notes et surlignages divers.

HOSPITEXTE repose par conséquent sur une philosophie opposée à celle des approches instituant un modèle conceptuel de données fixe : la donnée médicale n'est pas une donnée *a priori* mais une donnée *a posteriori*, résultant de la consultation du dossier. Le dossier n'est pas un sac d'informations élémentaires, mais une structure documentaire et matérielle dont l'utilisation toujours contextuelle détermine la valeur et le contenu en fonction du contexte de lecture [Berg, 1998].

Un hyperdocument n'est pas un document dans la mesure où, contrairement à celui-ci, il n'existe pas de parcours canonique de lecture [Bachimont, 2001]. Puisqu'aucun parcours de lecture n'est proposé par défaut dans un hyperdocument de spécialité, il faut que le lecteur puisse construire son parcours comme une lecture en inscrivant lui-même ce que chaque nœud atteint signifie pour lui en fonction de son objectif. D'autre part, l'hyperdocument, écrit par une multiplicité d'auteurs, n'est pas conçu par un auteur unique pour un lecteur unique. Littéralement (!), le lecteur doit alors être l'auteur de sa lecture et doit posséder pour cela les moyens de l'écrire. Le dossier patient informatisé doit donc « servir » les documents suivants :

- les documents « originaux », c'est-à-dire ceux existant actuellement dans le dossier papier ; ce sont donc les comptes rendus, les résultats d'examen, etc. ;
- les documents « de navigation », qui correspondent à des écrans proposés à l'utilisateur pour lui permettre de se rendre à un point particulier du dossier ; ce sont par exemple des tables de matières, ou des listes chronologiques de documents, où « cliquer » sur un élément de la liste permet de visualiser le contenu associé à cet élément ;
- les documents « de lecture », qui correspondent aux annotations que l'utilisateur a effectuées au cours de sa navigation et qui représentent ce que le lecteur a retenu de son investigation ; c'est le contenu de sa lecture ; par exemple, cela pourra être la liste des facteurs de risque pour une pathologie donnée.

Les documents de navigation et de lecture sont, par leur nature, des documents dits de « synthèse ». Les documents originaux comme les documents de synthèse sont tous semi-structurés et décrits dans le langage XML et respectent donc chacun une DTD (définition de type de documents). En résumé, l'utilisateur *a*) se voit servir les documents originaux, *b*) navigue dans l'hypertexte grâce à des documents de navigation et *c*) construit son parcours (*i.e.* écrit sa lecture) grâce à des fonctionnalités d'annotation.

4.1.3 Un prototype et des perspectives

Un prototype du système HOSPITEXTE, implémentant ces principes, a été développé [Brunie *et al.*, 2000]. Ce prototype correspond à la mise en place d'une chaîne de traitement documentaire permettant de passer des documents originaux (comptes rendus, résultats d'examen, etc) sous format électronique (Word en général) à un dossier patient informatisé *navigable* sur le Web.

Les perspectives de développement d'un tel prototype peuvent être envisagées dans un contexte industriel. Ainsi, HOSPITEXTE sert de base technologique au projet DOCPATIENT¹⁶ auquel nous participons et qui vise à développer un dossier médical hypertextuel à l'hôpital d'Amiens, dans un premier temps, au service de néonatalogie.

Du point de vue de la recherche académique, la question des annotations, souvent citées mais pas encore mises en œuvre dans le prototype, doit être abordée pour instrumenter la lecture du dossier. Documents plus annotations, nous avons là 2 facettes d'une même problématique, *la sémantique hypertextuelle*. Alors que la sémantique textuelle étudie le sens des mots au regard de leur position dans une unité textuelle – *e.g.* la phrase –, la sémantique hypertextuelle va étudier

¹⁶<http://www.laria.u-picardie.fr/Equipes/ic/htsc/>

le sens des documents au regard de leur position dans le réseau hypertextuel. Les balises XML définissant une syntaxe pour cette sémantique, la sémantique hypertextuelle définit quelles sont les transformations structurelles qui ont du sens. Par exemple, construire dynamiquement une table de tous les éléments balisés « facteur de risque » dans un dossier médical peut avoir du sens. C'est en suivant cette direction qu'un premier travail a été fait sur les documents de navigation [Brunie *et al.*, 2000]. Les capacités d'annotation sont d'une importance primordiale et donc disponibles dans la plupart des logiciels traitant des textes. En effet, et c'est vrai pour les hypertextes, ces significations – *i.e.* lectures – personnelles sont la condition de la légitimité qu'acquiert le système vis-à-vis de son utilisateur et de la tâche pour laquelle il est prévu. Comme pour les documents de synthèse précédents, mais ici à la destination d'un seul utilisateur, il faut savoir quelles transformations structurelles (organisations hiérarchiques, agrégation d'annotations, etc.) sont légitimes. Ces transformations sont théorisées dans le cadre d'une *sémantique hypertextuelle annotationnelle*. C'est un travail que nous avons commencé [Bringay *et al.*, 2004].

4.1.4 *Le point de vue de l'ingénierie des connaissances*

Au regard des réflexions développées précédemment, HOSPITEXTE est en adéquation avec les caractérisations qui y sont défendues, principalement :

Prise en compte du contexte La médecine s'enseigne comme une science mais s'exerce comme une pratique contextuelle. Les textes – sur support papier jusqu'à maintenant, sur support électronique ensuite – sont considérés, au sein du projet, comme la seule façon de conserver le contexte d'élaboration des connaissances ;

Création de nouvelles connaissances Par sa capacité à générer des tables des matières en tous genre, des *navigateurs*, ou de nouveaux types de documents par agrégation de parties de documents originaux, l'informatique crée de nouvelles proximités entre informations médicales, de nouvelles connaissances ;

De la technique et des usages Le langage de description de la structure des documents comme du contenu médical qui y est repéré est XML/SGML et ses avatars. L'ingénierie documentaire et Internet, techniques informatiques, sont alors utilisés dans au regard de leur capacité à produire du sens au regard des usages, ici, professionnels, collectifs comme idiosyncrasiques.

4.2 *Des thésaurus à la construction d'ontologies*

4.2.1 *Des ontologies en Ingénierie des connaissances*

La construction d'ontologies est un des champs de recherche de l'ingénierie des connaissances. Les ontologies¹⁷ sont apparues dans la discipline, avec l'idée de construire mieux et plus rapidement des SBC en réutilisant le plus possible des composants génériques, que ce soit au niveau du raisonnement ou des connaissances du domaine. Ce qui nous a amené incidemment à une première définition simple qui prend déjà acte d'une décision personnelle quant au choix des objets :

Ontologie (déf. 1) : *Ensemble des objets reconnus comme existant dans le domaine. Construire une ontologie c'est aussi décider de la manière d'être et d'exister des objets.*

¹⁷On parle ici d'ontologies en informatique. Bien que les réflexions qui ont amené la définition d'un tel concept ne sont pas indépendantes des réflexions philosophiques sur les ontologies, en particulier avec Aristote, les objets dont

Pour avancer sur la question de ce qu'est une ontologie, il nous semble indispensable de rappeler que les travaux sur *les ontologies sont développés dans un contexte informatique où le but final est de spécifier un artefact informatique*. Ce contexte est important pour comprendre les buts poursuivis par les concepteurs d'ontologie et sur les contraintes qui se posent à eux et seront développées au long de ce chapitre. En particulier, la question de la conceptualisation devient centrale dans le but de construire un artefact puisqu'on a besoin, dans ce contexte, de définir et spécifier les concepts à prendre en compte. Cette recherche des définitions au moment de l'élaboration de l'ontologie nous situe au niveau du sens, au niveau de l'*intension*. De plus, nous sommes devant un objet qu'il a bien fallu réfléchir, conceptualiser et qu'il faudra finalement *opérationnaliser* – c'est-à-dire coder dans un langage opérationnel, exécutable. Enfin, pour prendre en compte la construction des ontologies, on peut revenir à un travail de N. Guarino et P. Giaretta [1995] qui ont essayé de trouver des définitions de l'ontologie durant son processus d'élaboration¹⁸, à savoir l'ontologie comme un *système conceptuel informel*, puis l'ontologie comme *la représentation d'un système conceptuel via une théorie logique et son vocabulaire*. On peut alors proposer la seconde définition de ce qu'est une ontologie, en reprenant les spécifications de Gruber [1993] et M. Uschold *et al.* [1996] :

Ontologie (déf. 2) : *Une ontologie implique ou comprend une certaine vue du monde par rapport à un domaine donné. Cette vue est souvent conçue comme un ensemble de concepts – e.g. entités, attributs, processus –, leurs définitions et leurs interrelations. On appelle cela une conceptualisation.*

[...]

Une ontologie peut prendre différentes formes mais elle inclura nécessairement un vocabulaire¹⁹ de termes et une spécification de leur signification.

[...]

Une ontologie est une spécification rendant partiellement compte d'une conceptualisation.

Cette définition, nous explicite ce qu'est une ontologie par rapport à une application informatique et nous permet d'aborder les contraintes qui s'imposent successivement au concepteur :

- une ontologie est bien une conceptualisation, entendons par là que l'on y définit des concepts ;
- devant être par la suite un artefact informatique dont on veut spécifier le comportement, l'ontologie devra également être une théorie logique pour laquelle on précisera le vocabulaire manipulé ;
- enfin, la conceptualisation étant spécifiée parfois de manière très précise, une théorie logique ne peut toujours en rendre compte de façon exacte. Elle le fait donc partiellement²⁰.

on parle ne sont pas identiques. Par ailleurs, ces mêmes ontologies ont été remises sur le devant de la scène dans le cadre des recherches sur le Web sémantique où elles jouent un primordial en tant que classification des métadonnées pour l'indexation des ressources. Pour plus de développement, voir [Charlet, 2002], en particulier le chap. 4 qui donne de nombreux points d'entrée et propose une synthèse sur la question.

¹⁸Les développements faits ici sont une libre interprétation des attendus des travaux de N. Guarino et P. Giaretta [1995]. Ils ont pour but de progresser vers une définition rendant compte d'un processus de construction évolutif. Nous renvoyons le lecteur qui veut approfondir la question aux articles de ces auteurs.

¹⁹« Vocabulaire » est utilisé ici tel qu'il apparaît dans le texte. Il peut être compris – c'est le cas ici – dans un sens logique et être le vocabulaire des expressions manipulées par une théorie logique ou compris dans le contexte des systèmes d'information où le vocabulaire des objets du domaine vise à fournir le référentiel des objets existant. En tout état de cause, ce vocable prête à confusion dans le domaine des ontologies où la question de la langue et l'opposition *termes versus concepts* est fondamentale (*cf.* § 4.2.4).

²⁰ La théorie logique représentant formellement l'ontologie est l'*engagement ontologique* de cette dernière [Guarino

4.2.2 Des taxinomies

Les définitions proposées ci-dessous ont bien montré le caractère formel de l'artefact informatique construit mais pour préciser ce qu'est une conceptualisation, début du processus de construction d'une ontologie, nous revenons sur des conceptualisations connues de tous que sont les taxinomies en sciences naturelles.

La science a toujours eu pour premier but de repérer et classer les objets du monde pour les comprendre, comprendre leur fonctionnement et leur genèse. La recherche s'est systématisée en sciences naturelles, d'abord en botanique et ensuite pour tout le règne animal. Les classifications ainsi construites sont des taxinomies. Elles comportent la classification elle-même et les critères d'icelle. Sa définition rend compte de cette nature.

Taxinomie : 1. DIDACT. *Étude théorique des bases, lois, règles, principes, d'une classification.* 2. *Classification d'éléments* (Le Petit Robert).

Les motivations des classifications ont évolué dans le temps. En prenant l'exemple de la botanique, on a, au début du XVI^e siècle des classifications qui dépendent de critères que l'on croyait efficaces pour prédire des actions bénéfiques sur l'homme. Durant ce siècle et le suivant, ces classifications évolueront vers des critères liés aux organes de reproduction puis des critères floraux distinguant les végétaux supérieurs et inférieurs.

Au XVIII^e siècle, Carl VON LINNÉ [1707-1778] départage les grands groupes de plantes sur des caractères sexuels, mène à bien une recension complète des espèces et invente une nomenclature « binomiale » qui désigne une espèce par un nom générique commun à plusieurs espèces voisines et un nom spécifique différent pour chaque espèce du groupe. Il décrit ainsi la place d'un nœud dans la taxinomie, en nommant son père.

Au même siècle, Georges BUFFON (Georges-Louis LECLERC, Comte de BUFFON – [1707-1788]), et d'autres scientifiques commencent à suggérer que les formes de vie ne sont pas figées. Georges BUFFON défend cette idée dans son Histoire naturelle où il classe le monde vivant et en esquisse une théorie de l'évolution.

Plus près de nous, au XX^e siècle, les classifications effectuées jusque-là en suivant le système de VON LINNÉ sont remises en cause (250 ans plus tard quand même) : les théories de DARWIN sont apparues un siècle après VON LINNÉ et les espèces sont considérées comme des entités provisoires. À l'opposé d'une conception statique du monde, les biologistes pensent que les espèces représentent des entités mouvantes qui évoluent dans le temps. Ils cherchent alors à les classer dans des hiérarchies qui tiennent compte de ces évolutions : la taxinomie phylogénétique.

À la lumière de ces considérations, on peut s'intéresser par exemple aux mammifères et aux oiseaux dans le groupe des vertébrés. « Allaiter ses petits grâce à des glandes lactéales » et « avoir un tégument à poils » est, pour les 3500 espèces qui constituent à ce jour la classe des mammifères, toujours vrai. On peut donc considérer que ces deux caractères sont l'essence même des mammifères et donc définitoires.

Dans ce contexte, les grands traits de la classification des mammifères sont fondés sur l'anatomie de l'appareil génital femelle et s'appliquent simplement aux mammifères actuels. Pourtant, au regard d'une taxinomie phylogénétique et en essayant donc de reconstituer l'histoire paléontologique des mammifères, les choses se compliquent avec la découverte de fossiles reptiles présentant des caractères prémammaliens. C'est ainsi que l'origine reptilienne des mammifères a été affirmée et qu'il a fallu rechercher, du côté du squelette, de nouveaux caractères de différenciation entre les

& Giaretta, 1995]. Cet engagement est d'autant plus important à noter que la théorie logique ne rendant qu'approximativement compte de la conceptualisation visée, cet engagement est aussi la description de l'écart qu'il y a entre l'ontologie visée et sa formalisation.

reptiles et les mammifères au cours de l'évolution. On a là un exemple illustratif de la remise en cause de la taxinomie des vertébrés au XX^e siècle [Manaranche, 1995].

De ce petit aperçu sur les taxinomies en général et les mammifères en particulier, on peut tirer quelques enseignements qui intéressent les ontologies :

- Il est clair qu'une réflexion conceptuelle n'est pas toujours motivée par la construction d'un artefact informatique... évidemment ! Mais ici la motivation est, comme pour les ontologies en Ingénierie des connaissances, la reconnaissance et la classification des objets du domaine. L'objet construit ici, la taxinomie, est bien une conceptualisation des objets du domaine et donc une ontologie.
- Le mode de classification, c'est-à-dire l'organisation des objets et les caractères pris en compte (*i.e.* les propriétés et attributs de nos ontologies dépendent des buts fixés : si une classification arborescente s'est imposée depuis ARISTOTE, la variation des buts des taxinomies en sciences naturelles change radicalement une classification. On verra que ce constat trouve des résonances pour les ontologies en Ingénierie des connaissances (*cf.* § 4.2.4).

4.2.3 Des thésaurus en médecine

La médecine a une longue tradition de thésaurus médicaux dont les attendus ont d'abord été l'indexation des textes scientifiques médicaux et sont maintenant la représentation des données médicales sur les patients. Une définition est : *Un thésaurus est un ensemble de termes normalisés fondé sur une structuration hiérarchisée. Les termes y sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques. Organisé alphabétiquement, il forme un répertoire alphabétique de termes normalisés.*

Le plus connu des thésaurus en médecine est le MeSH . C'est le thésaurus d'indexation de la base bibliographique MEDLINE²¹. Il est traduit en français par l'INSERM et sert aussi de thésaurus au site CISMef²². Le MeSH offre une organisation hiérarchique et associative et comprend jusqu'à neuf niveaux de profondeur. Il est important de noter que ontologies et thésaurus sont proches mais que les thésaurus ayant été développés dans un contexte linguistique – des termes pour indexer ou pour exprimer la médecine – ils n'ont pas les propriétés formelles des ontologies qui permettent leur utilisation par un SBC. Inversement, le sens des termes d'un thésaurus est accessible par quelqu'un du domaine, là où les concepts d'une ontologies le sont plus difficilement.

Des travaux récents sur les *serveurs de terminologies* ou *thésaurus sémantiques* visent à représenter les termes d'un thésaurus avec les concepts d'une ontologie [Le Moigno *et al.*, 2002a,b; Roussey *et al.*, 2002]. Les thésaurus médicaux sont tellement gros et complexes, qu'il n'est plus possible d'assurer leur maintenance et cohérence dans un domaine en perpétuelle évolution. Une ontologie formant l'armature conceptuelle de ces thésaurus est une solution envisagée et mise en œuvre par certaines équipes [Rodrigues *et al.*, 1999] : c'est un type d'approche qui nous semble extrêmement prometteur, séparant explicitement les conceptualisations des façons d'en parler.

4.2.4 Une méthodologie de construction d'ontologies

La méthode que B. Bachimont *et al.* ont mis au point sur le projet MENELAS [Bouaud *et al.*, 1995] puis qui a été théorisée [Bachimont, 2000a], mobilise un certain nombre de réflexions philosophiques et épistémologiques. Si la dimension linguistique de la construction des ontologies y

²¹Le MeSH est accessible à <http://www.nlm.nih.gov/mesh/meshome.html>

²²<http://www.chu-rouen.fr/cismef/>

était rapidement reconnue [Bachimont, 1996], elle était peu ou pas outillée. Cet outillage est venue de réflexions sur les corpus initiées par D. Bourigault [Bourigault, 1994; Charlet *et al.*, 2000a]. Réflexions qui se sont encore enrichies en replaçant la problématique au sein de la construction de ressources terminologiques et ontologiques [Bourigault *et al.*, 2004]²³.

En résumé, la méthode propose 4 étapes :

1^{re} étape : la primauté du corpus et son analyse Nous sommes, par hypothèse, dans des domaines où les connaissances s'expriment en langue – la médecine en étant un des meilleurs exemples (*cf.* HOSPITEXTE). Il est alors possible de construire un corpus textuel qui devient la source privilégiée qui permettra de caractériser les notions utiles à la modélisation ontologique et le contenu sémantique qui leur correspond.

Pour ce faire, on utilise une « démarche de corpus » et des outils terminologiques pour commencer à modéliser le domaine. Ces outils, pour la plupart, reposent sur la recherche de formes syntaxiques particulières manifestant les notions recherchées comme des syntagmes nominaux pour des candidats termes, des relations syntaxiques marqueurs de relations sémantiques, ou des proximités d'usage – *e.g. contextes partagés* – pour des regroupements de notions. Ils font ce qu'on appelle de l'extraction terminologique et permettent d'obtenir des signifiés linguistiques avec une organisation plus ou moins structurées, souvent sous forme de réseaux.

2^e étape : la normalisation sémantique L'étape précédente nous a fourni des candidats termes [Bourigault, 1994] dont les libellés ont un sens pour le lecteur, souvent spécialiste du domaine. Mais rien n'assure que ce sens soit unique : au contraire, nous sommes dans un fonctionnement linguistique où les significations sont ambiguës, les définitions circulaires et dépendant en particulier du contexte interprétatif des locuteurs. Or, dans la modélisation ontologique, on cherche à construire des primitives dont le sens ne dépend pas des autres primitives et est surtout non contextuel. Il est nécessaire, pour prendre le chemin du formel, de *normaliser* les significations des termes pour ne retenir, pour chacun d'eux, qu'une seule signification, qu'une seule interprétation possible par un être humain. C'est ce que propose la *sémantique différentielle* de B. Bachimont. On dégage ainsi un ensemble de termes en construisant un système de différences entre ces termes. Dans ce contexte, la structure construite est, comme pour ARISTOTE, un arbre [Bachimont, 2000a].

À la fin de cette étape, l'ontologie construite n'est pas formelle : c'est un arbre de signifiés linguistiques normés – ou concepts linguistiques – par les principes différentiels appliqués. Ayant fixé le contexte d'interprétations de ces signifiés linguistiques, on a fabriqué une ontologie qui n'est valable que pour un contexte particulier, c'est-à-dire localement, c'est une *ontologie régionale* [Bachimont, 2000a].

3^e étape : l'engagement ontologique À cette étape, nous avons un arbre de primitives qui vont pouvoir être modélisées de façon formelle en définissant une sémantique formelle. Celle-ci va permettre de créer des concepts formels à partir et par opposition aux concepts linguistiques. Cette sémantique ne considère plus des notions sémantiques mais des extensions, c'est-à-dire l'ensemble des objets qui vérifient des propriétés définies en intension dans l'étape précédente, propriétés ayant une définition formelle à ce niveau. La structure de l'ontologie est ici un treillis. En effet, au sein de cette ontologie formelle, le treillis des concepts doit être compris comme la possibilité de créer des concepts dits définis en combinant les

²³Ces réflexions se sont en particulier développées au sein du groupe TIA – « terminologie et intelligence Artificielle » – et s'y poursuivent encore.

concepts primitifs : par exemple, une *personne* qui a pour rôle social d'être un *médecin*, cet « objet » défini en extension héritant des caractéristiques des personnes et des médecins alors qu'au niveau précédent, l'intension des personnes et des médecins étaient irréductibles.

4^e étape : l'opérationnalisation Dernière étape de la méthodologie (et de beaucoup d'autres), l'opérationnalisation consiste en la représentation de l'ontologie dans un langage de représentation des connaissances permettant de surcroît des services inférentiels de type classification des concepts ou généralisation, etc. Selon les langages considérés, les calculs possibles et donc les services inférentiels ne sont pas identiques et, à ce niveau aussi, il y a un engagement qui est pris avec de nouvelles contraintes et possibilités, justifiant l'existence d'une *ontologie computationnelle*.

L'ensemble des étapes et la nature des objets élaborés est résumé dans la figure 1.

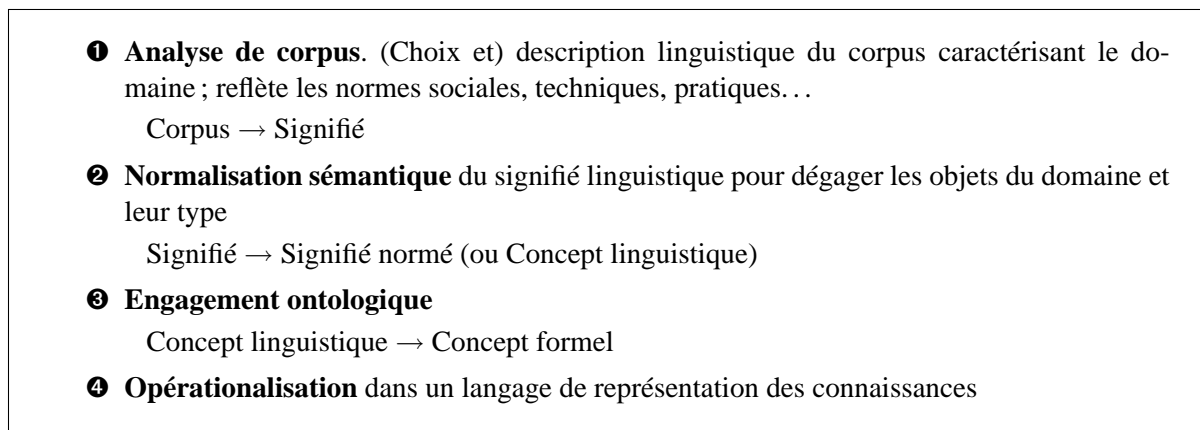


FIG. 1 – Les étapes de la méthodologie.

4.2.5 *Le point de vue de l'ingénierie des connaissances*

De cette rapide présentation des ontologies, on peut en retenir 2 idées fortes :

Le changement de paradigme. Sous les hypothèses rapidement citées, la construction d'ontologies à partir de textes implique de mettre au point une méthodologie globale qui s'organise selon différentes phases, chacune décrivant la façon de construire un produit terminologique intermédiaire – *i.e. ontologie régionale, conceptuelle* et finalement *computationnelle* –, correspondant à des changements de paradigme. Ainsi, la construction d'ontologie conceptuelle pour un SBC ou de référentiel « métier » pour un SI passe par ces changements qu'il faut savoir assumer, du linguistique au computationnel, en passant par le sémantique et le formel. Les réécritures que l'informatique permet dans un tel processus doivent être théorisées au risque d'aboutir à des thésaurus métiers dont le statut non clairement affirmé, permet toutes les interprétations et toutes les erreurs. La médecine, encore, fournit des exemples d'élaboration de thésaurus médicaux dont le paradigme, linguistique, est inadéquat à une utilisation conceptuelle comme référentiel « métier » [Zweigenbaum & Charlet, 1999].

La dépendance de la tâche. La construction d'ontologies « réutilisables » a été et est encore le but affiché d'un certain nombre de travaux. Cette hypothèse qui sous-tend l'affirmation d'une séparation forte entre les « raisonnements » et les connaissances du domaine, appelée *hypothèse d'interaction limitée* est battue en brèche par ses propres instigateurs [van Heijst *et al.*,

1997]. Il ressort finalement que si il y a généralité dans une ontologie, elle est dans la tâche – promise au SBC en construction – qui a présidé à l’élaboration de l’ontologie.

De plus, les travaux sur corpus ont montré la dépendance forte qu’il y avait entre l’élaboration du susdit corpus et la construction de l’ontologie du futur SBC : représentatif de la pratique à l’origine de la modélisation, le corpus est « porteur », via leurs expressions linguistiques, des futurs concepts de l’ontologie élaborée dans une démarche constructiviste²⁴.

5 Réflexions et discussion

5.1 Introduction

L’argumentaire développé jusque ici a visé à définir l’ingénierie des connaissances en tant que science et à proposer pour cela ses méthodes et son programme. Ceci fait, cela mérite critique et approfondissement. Ensuite, les discussions du séminaire de Cerisy m’ont permis de faire avancer ma réflexion et de proposer des rapprochements entre l’ingénierie des connaissances et les science de gestion. C’est donc dans ces 2 directions que je vais aller dans la suite de cette section.

5.2 Un domaine scientifique, un enseignement ?

Arrivé à ce point de l’exposé et des champs d’action de l’ingénierie des connaissances et avant de la comparer aux sciences de gestion, il me semble indispensable de se poser la question de savoir si l’argumentaire épistémologique mis en place ici est celui d’un domaine scientifique et de quel domaine. En effet, faire la preuve de l’existence d’un domaine scientifique ne va pas de soi et même si l’argumentaire développé jusque ici repose sur des travaux validés, d’un point de vue philosophique (*cf.* § 2) comme pratique (*cf.* § 4), il est nécessaire de replacer l’ingénierie des connaissances dans une perspective historique pour permettre de situer ce domaine – si c’en est un – par rapport aux autres et proposer des caractérisations.

Avant de comparer l’ingénierie des connaissances aux sciences de gestion, il m’a semblé indispensable de se poser la question de savoir si l’argumentaire épistémologique mis en place jusque ici est celui d’un domaine scientifique et de quel domaine. En effet, faire la preuve de l’existence d’un domaine scientifique ne va pas de soi et même si l’argumentaire développé repose sur des travaux validés, d’un point de vue philosophique (*cf.* § 2) comme pratique (*cf.* § 4), il est nécessaire de replacer l’ingénierie des connaissances dans une perspective historique pour permettre de situer ce domaine – si c’en est un – par rapport aux autres²⁵ et proposer des caractérisations.

La réponse à cette question a été développée dans [Charlet, 2002]. Nous n’en reprendrons que les conclusions et renvoyons le lecteur aux références citées.

Avant cela, un premier type de réponse quelque peu simpliste, est de regarder « l’état » du domaine en termes de personnes impliquée – *e.g.* le nombre d’abonnés à la liste de l’ingénierie des connaissances²⁶ –, la *bonne* santé de la conférence « IC » ou le nombre de thèses passées dans le domaine. Si les 2 premiers points sont favorables, le 3^e nécessiterait des études que nous avons

²⁴Une autre question est la construction d’une ontologie « partageable ». C’est une condition forte leur utilisation par une communauté comme le suppose par exemple le Web sémantique. Mais la dépendance du corpus donc d’une activité limite ce caractère partageable à des communautés de points de vue et d’intérêt.

²⁵De façon « différentielle ».

envisagées sans jamais les finaliser²⁷.

5.2.1 Des épistémologies positivistes aux épistémologies constructivistes

Pour placer l'ingénierie des connaissances dans une perspective historique, nous nous fondons sur les travaux de J.-L. Le Moigne sur les *épistémologies constructivistes* [1995]. Dans son livre, J.-L. Le Moigne étudie la constitution des connaissances « valables » – l'épistémologie –, à savoir : quel est le statut, la méthode et la valeur de la connaissance ? Il se place dans une perspective historique et analyse et critique les *épistémologies positivistes* qui ont jusqu'à récemment constitué le fondement du statut social de la connaissance, en particulier dans le milieu de la recherche. Il n'est pas de notre propos de développer cette critique. Nous allons plutôt nous restreindre à une présentation qui sert nos desseins, c'est-à-dire caractériser l'ingénierie des connaissances²⁸.

Les 2 principales hypothèses des épistémologies positivistes sont l'*hypothèse ontologique* et l'*hypothèse déterministe* : la première postule que *la connaissance que constitue la science est la connaissance de la réalité, une réalité postulée indépendante des observateurs qui la décrivent (même si leurs descriptions n'en sont pas indépendantes !)* ; la seconde postule une *détermination interne propre à la réalité connaissable. Sous une forme simplifiée c'est la causalité* (*Ibid.*, p. 18). Cette épistémologie « traditionnelle » est celle qui a justifié les activités scientifiques depuis le XIX^e siècle et jusqu'à récemment. On comprend bien, en première approximation, son adéquation à des sciences comme la physique classique²⁹ mais l'incapacité de telles hypothèses à rendre compte de la physique quantique ou l'idée que la connaissance d'un sujet est son expérience du réel (il est alors *modélisateur* de son expérience) ou encore l'incapacité du positivisme à rendre compte des sciences de « l'artificiel » comme la cybernétique, la gestion ou l'Intelligence artificielle a amené le développement des hypothèses constructivistes. Au regard du positivisme, les épistémologies constructivistes sont fondées sur 2 hypothèses :

- *L'hypothèse phénoménologique*³⁰. Sous ce vocable, on retrouve d'abord *le primat absolu du sujet connaissant capable d'attacher quelque valeur à la connaissance qu'il constitue : la connaissance implique un sujet connaissant et n'a pas de sens ou de valeurs en dehors de lui* (*Ibid.*, p. 67). Corollaire, il est impossible de séparer la connaissance du processus (intelli-

²⁶La liste « info-ic » <info-ic@biomath.jussieu.fr> est la liste de diffusion des informations de la communauté Ingénierie des connaissances (cf. http://www.irit.fr/ACTIVITES/EQ_SMI/GRACQ/index-act.html). Elle comporte à ce jour 370 abonnés.

²⁷De telles études ne sont pas originales et ont été effectuées dans d'autres domaines comme la gestion pour en citer une récente [Favier, 1998]. Il n'empêche qu'une telle étude sur le nombre de thèses passées en Ingénierie des connaissances reste à faire. Elle aurait une force de conviction extrêmement probante.

²⁸Pour argumenter son propos, J.-L. Le Moigne mobilise évidemment de très nombreux auteurs que nous ne citerons pas nous-mêmes. Nous renvoyons directement aux références de l'auteur [Le Moigne, 1995]. Par ailleurs, on trouve chez Nonaka [1997, p. 39-47] un point de vue oriental et un développement plus court qui reprend les mêmes arguments en opposant les *épistémologies rationalistes* et les *épistémologies empiriques* mais l'argumentaire développé par J.-L. Le Moigne, pour similaire qu'il soit, mobilise des concepts différents et propose des conclusions sur la validité des domaines scientifiques qui sont les problématiques qui nous intéressent ici.

²⁹Nous n'entrerons pas ici dans le débat sur la place des mathématiques par rapport aux épistémologies positivistes et nous nous en tiendrons à l'exemple de la physique que nous connaissons mieux en notant que la physique n'est pas « bêtement » positiviste : les principes de relativité ont pour fonction de neutraliser l'influence de l'observateur sur les phénomènes observés et par conséquent d'en affirmer l'objectivité : ainsi, la relativité galiléenne dit que les lois de la mécanique sont les mêmes dans tous les repères en translation uniforme les uns par rapport aux autres. C'est donc qu'il y a des lois qui sont objectives à un principe de relativité près, c'est-à-dire à une équivalence entre observateurs près. La physique quantique, quant à elle, force à admettre la non séparabilité des phénomènes et de leur observateur et remet en cause un simple principe d'objectivité de l'observation.

gence) qui la produit. Deux conséquences de cette hypothèse, parmi d'autres, son intéressantes à développer, (1) la dialectique de la cognition et (2) la récursivité de la cognition : La première exprime *le caractère dialectique que le sujet connaissant attribue à ses perceptions* (*Ibid.*, p ; 73), caractère dialectique qui implique une vérité non absolue, argumentée. La seconde conséquence *rend compte de l'interdépendance assumée entre le phénomène perçu et sa connaissance construite* (*Ibid.*, p 74). Interdépendance qui implique que les représentations que nous construisons d'un phénomène sont des modèles plus ou moins actifs – en termes d'interactions cognitives avec l'utilisateur – qui transforment notre perception.

- *L'hypothèse téléologique*³¹. En attribuant au sujet connaissant le rôle décisif dans la construction de la connaissance, l'hypothèse phénoménologique oblige en quelque sorte à prendre en compte *l'intentionnalité ou la finalité de ce sujet connaissant*. C'est-à-dire qu'on interprète plus volontiers le comportement cognitif du sujet connaissant en termes de « causes finales » plutôt qu'en termes de « causes efficientes ». Mais surtout la détermination et la transformations de ces finalités semblent très souvent devoir être interprétées en termes endogènes, produits par le sujet lui-même. Ainsi l'homme est le seul animal capable (ne serait-ce qu'un peu) de se finaliser. Il n'existe plus de vérité objective mais une connaissance à laquelle le sujet attribue une propre valeur, *liée à sa faculté de juger, à ses intentions, et plus à une quelconque norme du vrai*.

5.2.2 L'ingénierie des connaissances

Après avoir rapidement abordé ce passage historique des épistémologies positivistes aux épistémologies constructivistes, on peut maintenant en synthétiser les conséquences sur les caractérisations de l'ingénierie des connaissances, en empruntant à H.A. Simon ses réflexions sur le paradigme des « sciences de l'artificiel » [Simon, 1981]. Ces caractérisations seront au nombre de 3 et porterons sur (1) les hypothèses, (2) la méthode et enfin (3) la validité de la connaissance et son enseignement.

Notre étude historique et épistémologique nous a donc amené à caractériser l'ingénierie des connaissances, en empruntant à H.A. Simon ses réflexions sur le paradigme des « sciences de l'artificiel » [Simon, 1981]. Ces caractérisations seront au nombre de 3 et porterons sur *a)* les hypothèses, *b)* la méthode et enfin *c)* la validité de la connaissance et son enseignement :

Les hypothèses La connaissance n'est pas objective. Elle n'existe pas *a priori*. Elle est construite à partir d'un projet propre à un sujet connaissant. Ce sujet est un *modélisateur*. On retrouve ici des arguments pour alimenter le débat sur la réutilisation des ontologies au sein de l'ingénierie des connaissances : la modélisation dépend du *modélisateur*, de son projet et donc de la tâche visée pour l'artefact.

La connaissance est accessible par la médiation d'une représentation, que cette représentation soit un système de symboles ou tout autre langage. Le support de cette représentation, parce qu'il fournit un système de signifiants supplémentaire – *e.g.* l'interface d'un

³⁰La phénoménologie est la méthode philosophique qui se propose de considérer les choses en elles-mêmes en dehors de toute construction conceptuelle. Une phénoménologie de la connaissance consiste alors à penser ce qu'est la connaissance et comment elle se constitue. La phénoménologie n'est pas unique et si nous relayons les hypothèses dont J.-L. Le Moigne se fait le porte-parole, d'autres phénoménologies sont mentionnées par rapport aux sciences de l'artificiel et donc de l'Intelligence artificielle. On trouve dans [Bachimont, 1996] des développements plus approfondis sur la question, en particulier au sujet de la phénoménologie husserlienne.

³¹La téléologie est l'étude des finalités.

programme d'ordinateur (*cf.* § 4.1) – est aussi un système de médiation. Il y a ainsi non séparabilité de la connaissance et de sa représentation.

La méthode La *modélisation systémique* ou analyse des systèmes s'est développée de façon concomitante au constructivisme. Les sciences de l'artificiel qui se développaient en même temps – *e.g.* sciences de l'information, de l'organisation, gestion, etc. – y ont fait rapidement appel pour pallier les déficiences de la modélisation analytique ignorant l'utilisateur, le modélisateur, l'environnement. . . La modélisation systémique privilégie ainsi le projet – *i.e.* la téléologie – sur l'objet. L'ingénierie des connaissances, dans ce paradigme, construit des modèles, des SBC, modélise des interactions entre l'homme et l'ordinateur, se donne pour projet le développement de SBC, abandonnant la recherche sur les connaissances de l'expert qu'il fallait extraire pour « alimenter » un SE. C'est ce type de démarche qui amène maintenant l'ingénierie des connaissances à rencontrer d'autres disciplines – *e.g.* terminologie, ergonomie, gestion, etc. – pour mener à bien son projet : développer des SBC, des artefacts, dans des environnements de plus en plus complexes (entreprises, SI, . . .), pour des utilisateurs effectuant de nombreuses tâches. La gestion des connaissances, symptôme plutôt que solution [Le Masson, 2000], interroge ainsi nos capacités de modélisation et d'appréhension de la complexité.

La validité de la connaissance et son enseignement Étant assuré qu'il n'y a pas de connaissance objective, non construite, H.A. Simon propose *le principe d'action intelligente* qui décrit comment le sujet connaissant, mettant en œuvre toutes les ressources du raisonnement dialectique dès lors qu'elles sont reproductibles, construit, par des systèmes de symboles, des connaissances « faisables », des connaissances qui peuvent dès lors être tenues pour « possibles ».

Le modélisateur, aussi enseignant, est alors responsable des connaissances qu'il communiquera ou enseignera. *Il ne pourra plus assurer que ces connaissances sont « démontrées » et devra montrer qu'elles sont « argumentées » et reproductibles, de façon à permettre leur intelligibilité pour son interlocuteur* [Le Moigne, 1995, p. 83-6]. Dans le contexte de l'ingénierie des connaissances, on retrouve cela au niveau de la modélisation où la seule preuve de la vérité est l'efficacité et la généralité d'une méthode ou d'un modèle : ainsi, les différents composants et étapes de modélisation du modèle conceptuel, pour discuter qu'ils soient se décrivent de façon générale – et très souvent – par le schéma proposé en 1992 par N. Aussenac-Gilles, J.-P. Krivine et Ph. Laublet [1992].

Dans ce cadre, l'ingénierie des connaissances doit être enseignée et développée, ce à quoi nous (et ce n'est pas un nous de circonstance mais bien un nous représentant un certain nombre d'enseignants exerçant au sein de DEA ou DESS) nous attelons³².

On peut remarquer que l'ingénierie des connaissances est une science de l'artificiel qui travaille sur des artefacts pour lesquels, avantage sur les sciences du naturel, nous connaissons souvent le projet intentionnel du sujet connaissant qui a construit cet artefact, quand ce n'est pas nous-mêmes. Pour terminer, il est intéressant de retrouver une bonne adéquation entre les conclusions de cette étude historique et des résultats sortis de lectures plus philosophiques ou plus pragmatiques (*cf.* § 2) ; adéquation assez évidente pour qu'il ne soit pas nécessaire de la reprendre ici. Plus productif, semble-t-il, est la recherche de complémentarités. Ainsi, cet historique nous a surtout permis de compléter notre approche sur le statut de l'ingénierie des connaissances et sur la validité

³²Nous avons, depuis quelques temps, proposé des supports de cours à disposition sur le Web <<http://www.irit.fr/GRACQ/cours.html>> .

des connaissances que l'on peut acquérir et donc enseigner. Il n'est pas inutile de réfléchir au fait que nous n'enseignons pas des vérités absolues mais que notre faculté de juger, les expériences que nous avons mené dans le domaine – et c'est bien un « nous » pluriel, au nom des « praticiens » de l'ingénierie des connaissances – nous permet de proposer des méthodes reproductibles et argumentées et que c'est cela qui justifie un tel domaine scientifique. Pour conclure, en accord avec J.-L. Le Moigne et H. Simon, nous pouvons alors dire que l'ingénierie des connaissances sciences du génie parmi d'autres, ou science de l'ingénierie par excellence peut se présenter en des termes épistémologiques argumentés et quitter le champ des « sciences physiques pour l'ingénieur » auquel les épistémologie positivistes et les institutions la cantonnaient et rejoindre le *club des concepteurs de projets constructibles*³³.

5.3 Une ingénierie des inscriptions numériques des connaissances

Dans cet article, il est souvent utilisé le vocable « représentation des connaissances » avec parfois des guillemets, en tous cas souvent des précautions. Cet usage vient d'abord du fait que ce vocable dénote une discipline de l'intelligence artificielle qui s'occupe des langages dans lesquels on pourrait « représenter » des connaissances. Cette discipline voyant ses réflexions réactivés par le Web sémantique et ses besoins de standardisation des langages. Ensuite, les guillemets sont évidemment liés à la possibilité de réellement représenter des connaissances dans une machine. C'est cette difficulté – même impossibilité – que nous allons discuter en reprenant les arguments de B. Bachimont [2005].

Dans son argumentation, B. Bachimont propose d'abord de réfléchir au statut des modèles de l'ingénierie des connaissances. Il argumente ainsi que ces modèles ne sont ni des modèles formels que l'ingénierie des connaissances utilise pour implémenter ses outils mais pas dans sa réflexion modélisatrice, ni des modèles cognitifs de raisonnement que l'ingénierie des connaissances ne prétend plus développer depuis qu'elle se concentre sur les modèles des artefacts à construire.

Si modèles il y a, ce sont les modèles des outils que l'ingénierie des connaissances développe. Dans cette hypothèse, ces modèles ne peuvent être validés formellement ou cognitivement pour eux-mêmes : la validation ne peut être effectuée dans l'absolue ; elle ne se fait que dans le contexte de l'application développée³⁴. Ainsi, l'ingénierie des connaissances cherche à instrumenter un travail sur la connaissance et sa légitimité va venir des organisations intellectuelles qu'elle propose et matérialise dans ces outils, modifiant l'organisation où ils s'insèrent.

³³Le développement du Web et des NTIC – *i.e.* nouvelles technologies de l'information et de la communication – et les argumentations des chercheurs ont finalement amené la création du département STIC – *i.e.* sciences et technologies de l'information et de la communication – du CNRS et par là, nous pouvons l'imaginer, la reconnaissance institutionnelle espérée.

³⁴Je précise bien ici que que la validation formelle du modèle logique sous-jacent est possible et souvent souhaitable – *e.g.* la formalisation et l'opérationnalisation des opérations de subsomption dans les ontologies à l'aide de « logiques de description » types OWL <<http://www.w3.org/2004/OWL/>>. Mais cette formalisation, si elle garantit la fondation logique du modèle, n'est en rien une garantie de sa pertinence sémantique, une garantie que son comportement sera correctement interprété par un utilisateur. Cette pertinence doit être mesurée au regard du contexte d'usage. Il en est de même de la validité cognitive de l'outil développé : la validité formelle acquise, cette dernière doit être appréciée dans le contexte organisationnel d'usage de l'outil. On notera que ces validations apparaissent successives pour des raisons de facilité alors qu'elles sont liées : la validation formelle dans le contexte d'usage ne peut s'apprécier qu'en se servant de l'outil. C'est toute la difficulté de l'ingénierie des connaissances : faire des outils « génériques » dont la validation n'est réellement faisable qu'au niveau des produits finis et qui ne s'apprécient que dans des situations toujours uniques.

L'ingénierie des connaissances peut être alors comprise comme une technique manipulant des inscriptions numériques devant s'interpréter comme des connaissances – proposant des machines *qui donnent à penser et non des machines qui pensent*. Ces inscriptions numériques doivent être valides d'un double point de vue : a) du point de vue des techniques informatiques et formelles qui permettent de les manipuler et b) et par rapport au contexte d'utilisation de l'outil développé, c'est-à-dire par rapport à un système de normes et conventions du domaine de pratique dans lequel s'insère l'outil. *l'ingénierie des connaissances est une technique des inscriptions formelles et une critique de leur interprétation* [Bachimont, 2004].

Caractérisés ainsi, l'ingénierie des connaissances et ses outils convergent en plusieurs points avec la gestion et ses outils :

- Il n'y a pas « représentation des connaissances » dans le sens où l'outil véhiculerait une connaissance. Il y a un outil qui propose des interprétations privilégiées par rapport à un système de normes (*e.g.* l'usage du marteau en fonction de sa forme).
- Le modèle de l'outil n'est pas la représentation du sens, il est lui-même un instrument pouvant intervenir dans l'interprétation des situations d'action au regard du système de normes pratiques (l'organisation) dans lequel il s'insère.
- La question de la place des outils dans les processus d'apprentissage collectif nous semble théorisable de la même façon dans les 2 disciplines à partir du moment où le statut des outils – guide pour l'interprétation des situations – est potentiellement identique. Mais cette convergence de positionnement reste à réfléchir et argumenter.

5.4 La généricité des outils

Science des inscriptions numériques, ne validant ses modèles que par rapport à des situations toujours particulières, l'ingénierie des connaissances subit une tension quant à la généricité des outils développés : développant des artefacts informatique, elle partage avec d'autres domaines comme le génie logiciel, le besoin de standardiser ses modèles et ses modèles pour plus d'efficacité ; d'un autre côté, elle partage avec les sciences de gestion, le besoin que les outils développés soient utilisés par un maximum de gens (un collectif) d'une organisation et que cette utilisation passe par des adaptations à l'idiosyncrasie des utilisateurs qui soient anticipables dans l'outil lui-même.

Cette tension se réifie dans les approches mises en œuvre dans les outils de l'ingénierie des connaissances. En prenant un point de vue formel, on remarque qu'une information est une inscription dont l'appropriation s'effectue de manière déterministe, c'est-à-dire indépendamment du contexte. La forme de l'information prescrit son usage, comme en témoigne de manière extrême la théorie de l'information. À l'inverse, la connaissance, inexistante en tant que telle, l'inscription d'une connaissance donc plutôt, est une inscription possédant un flottement interprétatif qu'il faut lever en faisant appel au contexte, à l'action, au but. Deux tendances apparaissent alors³⁵ :

- Une *tendance technique* est de formaliser les inscriptions pour les émanciper du contexte, de manière à leur épargner l'interprétation par un utilisateur. Les approches formelles à base d'ontologies rentrent évidemment dans ce paradigme. Dans cette tendance, l'ingénierie des connaissances cherche à produire des méthodes rendant compte de comportements invariants. C'est toute la difficulté de l'entreprise : figer un contexte d'interprétation pertinent pour un temps suffisamment long au regard de l'entreprise et du système d'information mais prévoir les évolutions qui ne manqueront pas de se produire.
- Une *tendance épistémologique* est de rechercher une façon de conserver ce contexte et donc la capacité d'interprétation de l'utilisateur. Les approches à base de textes semi-structurés (*cf.* HOSPITEXTE infra) rentrent dans cet autre paradigme et doivent s'affirmer en permettant le

développement d'outils suffisamment génériques pour qu'ils valent d'être développés mais respectant les interprétations personnelles des utilisateurs pour qu'ils s'en servent. Dans le domaine de la médecine – *i.e.* annotations au sein des dossiers médicaux – comme dans d'autres – *i.e.* station de lecture/écriture critique –, cela veut dire inscrire l'idiosyncrasie de l'utilisateur dans des classes d'usage des outils, des fonds de chaînes opératoires communs à un domaine [Chahuneau *et al.*, 1992; Stiegler, 1993].

6 Conclusion

Nous avons voulu, ici, étudier le domaine de l'ingénierie des connaissances, la valeur de la connaissance en son sein, ses méthodes et son statut scientifique. À la conclusion – temporaire – de ce travail, il nous semble être arrivé à des caractérisations *constructives* en ce sens que nous pouvons proposer un programme – une téléologie – des méthodes et... des exemples.

Par ailleurs, il nous a semblé utile, dans la suite du colloque de Cerisy, de réinterroger la position de l'ingénierie des connaissances parmi les sciences de l'artificiel, en particulier les sciences de gestion que ce soit par rapport aux attendus des disciplines ou aux outils construits.

Si les attendus des modèles construits ne sont pas tous identiques, en particulier, les outils de l'ingénierie des connaissances ne sont, *a priori*, pas liés au pilotage, leur mise en œuvre et leur statut d'instruments inséré dans les processus d'interprétation des acteurs créent des communautés de réflexion qui nous semblent intéressantes de développer que ce soit par rapport aux situations d'actions précédemment citées ou à l'apprentissage collectif. Dans ce contexte, la gestion peut et doit, par des réflexions déjà avancées, apporter beaucoup. À l'inverse, l'ingénierie des connaissances peut proposer ses méthodes aux sciences de gestion, dans les modes de développement des outils d'inscription des connaissances. Beaucoup (tout ?) reste évidemment à faire dans l'activation de cette pluridisciplinarité.

Références bibliographiques

- AUSSENAC-GILLES N., KRIVINE J.-P. & SALLANTIN J. (1992). Editorial. *Revue d'Intelligence Artificielle*, 6(1-2), 7-18. *Introduction to the special issue on Knowledge Acquisition*.
- N. AUSSENAC-GILLES, P. LAUBLET & C. REYNAUD, Eds. (1996). *Acquisition et ingénierie des connaissances : tendances actuelles*. Cepadue-éditions.
- AUSSENAC-GILLES N., REYNAUD C., TCHOUNIKINE P. & TRICHET K. (1997). Associer un type de raisonnement à un domaine : une question de rôle ? In *Actes des 1^{es} Journées Ingénierie des Connaissances*, Roscoff, France.
- BACHIMONT B. (1996). *Herméneutique matérielle et Artéfacture : des machines qui pensent aux machines qui donnent à penser ; Critique du formalisme en intelligence artificielle*. Thèse de doctorat d'épistémologie, École Polytechnique.
- BACHIMONT B. (2000a). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In Charlet *et al.* [2000b], chapitre 19.

³⁵Cette opposition technique *versus* épistémologique nous a été suggérée par B. Bachimont. Par ailleurs, elle est à rapprocher de l'opposition technique *versus* conceptuelle de A. David [2001].

- BACHIMONT B. (2000b). L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle. In J. PETITOT & R. CASATI, Eds., *Au nom du sens*. Grasset.
- BACHIMONT B. (2001). Dossier et lecture hypertextuelle : problématique et discussions. *Les cahiers du numérique*, p. 105–23. numéro spécial sur l'information médicale numérique sous la direction de Pierre Le Beux et Dominique Boullier.
- BACHIMONT B. (2004). Pourquoi n'y a-t-il pas d'expérience en ingénierie des connaissances ? In N. MATTA, Ed., *Actes des 8^{es} Journées Ingénierie des Connaissances*, Lyon, France : Presses universitaires de Grenoble. ISBN 2 7061 1221 2.
- BACHIMONT B. (2005). Ingénierie des connaissances, ingénierie de la contingence : la technique entre le nécessaire et le singulier. In Teulier & Lorino [2005], chapter 18. Actes du colloque de Cerisy « Activité, connaissance, organisation », *À paraître*.
- BERG M. (1998). Medical work and the computer-based patient record : A sociological perspective. *Methods of Information in Medicine*, **37**(3).
- BOUAUD J., BACHIMONT B., CHARLET J. & ZWEIGENBAUM P. (1995). Methodological principles for structuring an “ontology”. In *Proceedings of the IJCAI'95 Workshop on “Basic Ontological Issues in Knowledge Sharing”*, Montréal, Canada.
- BOUAUD J., SÉROUSSI B., ANTOINE E.-C., GOZY M., KHAYAT D. & BOISVIEUX J.-F. (1998). Hypertextual navigation operationalizing generic clinical practice guidelines for patient-specific therapeutic decisions. *Journal of the American Medical Informatics Association*, **5**(suppl), 488–492.
- BOURIGAULT D. (1994). *LEXTER, un logiciel d'extraction de terminologie. Application à l'acquisition de connaissances à partir de textes*. Thèse de doctorat, École des hautes études en sciences sociales.
- BOURIGAULT D. (2000). Construction de ressources terminologiques. In J.-M. PIERREL, Ed., *Ingénierie des langues*, Traité IC2, chapter 9. Paris : Hermès.
- BOURIGAULT D., AUSSENAC-GILLES N. & CHARLET J. (2004). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas. Numéro spécial *Techniques informatiques et structuration de terminologies* de la *Revue d'intelligence artificielle*, 18(1).
- BRINGAY S., BARRY C. & CHARLET J. (2004). *Les documents et les annotations du dossier patient hospitalier*, In J.-M. SALAÜN & J. CHARLET, Eds., *Le document numérique*, p. 191–211. Cépaduès : Toulouse. Numéro thématique *Le document numérique* de la revue *Information - Interaction - Intelligence*, 4(1), ISBN 2.85428.634.0.
- BRUNIE V., BACHIMONT B. & MORIZET-MAHOUDEAUX P. (2000). Modélisation des connaissances structurelles documentaires pour la conception d'un dossier médical hypertextuel. In Charlet *et al.* [2000b], chapter 25, p. 407–22.
- CHAHUNEAU F., LECLUSE C., STIEGLER B. & VIRBEL J. (1992). Prototyping the ultimate tool for scholarly qualitative research on texts. In *Proceedings of the 8th Annual Conference of the UW Center for the New Oxford English Dictionary and Text Research*, Waterloo.

- CHARLET J. (2002). *L'Ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales*. Habilitation à diriger des recherches, Université Paris 6. Disponible à http://tel.ccsd.cnrs.fr/documents/archives0/00/00/69/20/index_fr.html le 20.09.04.
- CHARLET J. & BACHIMONT B. (1998). De l'acquisition à l'ingénierie des connaissances : applications et perspectives. In *Actes des assises nationales 1998 du PRC-I3*.
- CHARLET J., BACHIMONT B., BRUNIE V., EL KASSAR S., ZWEIGENBAUM P. & BOISVIEUX J.-F. (1999). L'ingénierie documentaire au service du dossier patient électronique. In A. VENOT & H. FALCOFF, Eds., *L'informatisation du cabinet du futur*, Informatique et Santé. Springer Verlag. Présenté au colloque "L'Informatisation du Cabinet Médical du Futur", 29-30 janvier 1999.
- CHARLET J., REYNAUD C. & TEULIER R. (2001). Ingénierie des connaissances pour les systèmes d'information. In C. CAUVET & C. ROSENTHAL-SABROUX, Eds., *Ingénierie des systèmes d'information*, chapitre 10. Paris : Hermès. ISBN 2-7462-0219-0.
- CHARLET J., ZACKLAD M., KASSEL G. & BOURIGAULT D. (2000a). Ingénierie des connaissances : recherches et perspectives. In Charlet *et al.* [2000b], chapitre 1, p. 1–22.
- J. CHARLET, M. ZACKLAD, G. KASSEL & D. BOURIGAULT, Eds. (2000b). *Ingénierie des connaissances : évolutions récentes et nouveaux défis*. Paris : Eyrolles.
- CONDAMINES A. & AUSSENAC-GILLES N. (2001). Entre textes et ontologies formelles : les bases de connaissances terminologiques. In M. ZACKLAD & M. GRUNDSTEIN, Eds., *Ingénierie et capitalisation des connaissances*. Paris : Hermès. ISBN 2-7462-0234-4.
- DAVID A. & PALLEZ F. (2001). Les systèmes d'information à l'épreuve de l'organisation. In C. CAUVET & C. ROSENTHAL-SABROUX, Eds., *Ingénierie des systèmes d'information*, chapitre 2, p. 23–60. Paris : Hermès. ISBN 2-7462-0219-0.
- ERMINE J.-L. (1996). *Les systèmes de connaissance*. Hermès.
- FAVIER M. (1998). Système d'information. les cinquantes prochaines années pour les SI ! In *Journées nationales des IAE*, p. 101–119 : LRSG, Univ. Nantes.
- GANASCIA J.-G. (1998). *Le Petit Trésor. Dictionnaire de l'informatique et des sciences de l'information*. Édition Flammarion.
- GOODY J. (1979). *La raison graphique*. Éditions de Minuit.
- GRUBER T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, **5**, 199–220.
- GUARINO N. & GIARETTA P. (1995). Ontologies and knowledge bases. In *Towards Very Large Knowledge Bases*. Amsterdam : IOS Press.
- HATCHUEL A. & WEIL B. (1992). *L'expert et le système. Gestion des savoirs et métamorphoses des acteurs dans l'entreprise industrielle*. Economica. 263 p.
- KAYSER D. (1997). *La représentation des connaissances*. Hermès.

- LE MASSON P. (2000). *La gestion des connaissances : au-delà des effets de mode, une notion essentielle pour comprendre et affronter les évolutions de l'entreprise*. Rapport interne, École des Mines de Paris.
- LE MOIGNE J.-L. (1973). *Les systèmes d'information dans les organisations*. Paris : PUF. P. 15-16.
- LE MOIGNE J.-L. (1995). *Les épistémologies constructivistes*. PUF.
- LE MOIGNO S., CHARLET J., BOURIGAULT D., DEGOULET P. & JAULENT M.-C. (2002a). Terminology extraction from text to build an ontology in surgical intensive care. In *Actes du AMIA Annual Symposium 2002*, San Antonio, Texas.
- LE MOIGNO S., CHARLET J., BOURIGAULT D. & JAULENT M.-C. (2002b). Construction d'une ontologie à partir de corpus : expérimentation et validation dans le domaine de la réanimation chirurgicale. In B. BACHIMONT, Ed., *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 229–38, Rouen, France.
- LEROI-GOURHAN A. (1964). *Le geste et la parole*. Albin Michel.
- LORINO P. (1995). *Comptes et récits de la performance. Essai sur le pilotage de l'entreprise*. Les éditions d'organisations. ISBN : 2-7081-1833-1.
- LORINO P. (2005). Théories des organisations, sens et action : le cheminement historique des idées. In Teulier & Lorino [2005], chapitre 2. Actes du colloque de Cerisy « Activité, connaissance, organisation », *À paraître*.
- LÉPINE P. & AUSSENAC-GILLES N. (1996). *Modélisation de la résolution de problèmes : comparaison de KADS et MACAO sur une application juridique*, chapitre 7, p. 131–48. In Aussenac-Gilles et al. [1996].
- MANARANCHE R. (1995). Mammifères. In *Encyclopædia Universalis*, vol., p. 405–13.
- MOISDON J.-C. (2005). Comment apprend-on par les outils de gestion ? retour sur une doctrine d'usage. In Teulier & Lorino [2005], chapitre 15. Actes du colloque de Cerisy « Activité, connaissance, organisation », *À paraître*.
- MÉLÈSE J. (1990). *Au sujet de l'information*. Paris : Les éditions de d'Organisation.
- NEWELL A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87–127.
- NONAKA I. & TAKEUCHI H. (1997). *La connaissance créatrice. La dynamique de l'entreprise apprenante*. De Boeck University. 303p. ISBN 2-7545-0034-8.
- NYGREN E. & HENRIKSSON P. (1992). Reading the medical record. I. Analysis of physicians' ways of reading the medical record. *Computer Methods and Programs in Biomedicine*, **39**, 1–12.
- POITOU J.-P. (1996). La gestion des connaissances comme condition et résultat de l'activité industrielle. *Intellectica*, **1**, 185–202.

- RODRIGUES J.-M., TROMBERT-PAVIOT B., RECTOR A., BAUD R., CLAVEL L., ABRIAL V., IDIR H. & VERY J.-M. (1999). GALEN, il existe quelque chose après les mots : leur signification et au delà le savoir médical. *Innovation Stratégique en Information de Santé*, (2-3), 48-62.
- ROUSSEY C., CALABRETTO S. & PINON J.-M. (2002). Le thésaurus sémantique : contribution à l'ingénierie des connaissances documentaires. In B. BACHIMONT, Ed., *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 209-20, Rouen, France.
- SHANNON C.-E. & WEAVER W. (1971). *Information*. University of Illinois Press. 12e édition (édition originale, 1949).
- SIMON H. A. (1981). *The sciences of the artificial*. Cambridge : The MIT Press. (trad. Française 1990).
- STIEGLER B. (1993). Lecture et édition savante assistée par ordinateur : l'hypertraitement de texte. In G. DUPOIRIER, Ed., *Actes du congrès Afcet*, volume 4, p. 37-45, Versailles : AFCET. tome 4 : Bureautique, document, groupware et multimedia.
- TANGE H. J. (1995). The paper-based patient record : Is it really so bad ? *Computer Methods and Programs in Biomedicine*, **48**, 127-131.
- R. TEULIER, J. CHARLET & P. TCHOUNIKINE, Eds. (2004). *Ingénierie des connaissances*. Paris : L'Harmattan. *À paraître*.
- R. TEULIER & P. LORINO, Eds. (2005). *Entre la connaissance et l'organisation, l'activité collective*. La découverte. Actes du colloque de Cerisy « Activité, connaissance, organisation », *À paraître*.
- USCHOLD M. & GRUNINGER M. (1996). Ontologies : Principles, methods and applications. *Knowledge Engineering Review*.
- VAN HEIJST G., SCHREIBER A. T. & WIELINGA B. J. (1997). Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, **45**(2/3), 183-292.
- ZWEIGENBAUM P. & CHARLET J. (1999). *Information, connaissance et langue médicales*. Rapport interne 99-221, DIAM, SIM/DSI/AP-HP. Document de travail pour le groupe de travail INSERM « Sciences de l'Information ».