



HAL
open science

L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales

Jean Charlet

► **To cite this version:**

Jean Charlet. L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. domain_stic.gest. Université Pierre et Marie Curie - Paris VI, 2002. tel-00006920

HAL Id: tel-00006920

<https://theses.hal.science/tel-00006920>

Submitted on 20 Sep 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.

Mémoire d'Habilitation à diriger des recherches

présenté à l'Université Pierre et Marie Curie
Spécialité : Informatique

par

Jean Charlet

Mission de recherche en
sciences et technologies de l'information médicale
Assistance Publique-Hôpitaux de Paris

L'INGÉNIERIE DES CONNAISSANCES DÉVELOPPEMENTS, RÉSULTATS ET PERSPECTIVES POUR LA GESTION DES CONNAISSANCES MÉDICALES

Soutenu le 10 décembre 2002 devant la commission d'examen composée de :

MM	Jean-Charles	POMEROL	Président
	Régis	BEUSCART	Rapporteur
	Yves	CHIARAMELLA	Rapporteur
	Philippe	LORINO	Rapporteur
	Catherine	GARBAY	Examineur
	Thomas	SIMILOWSKI	Examineur
	Bruno	BACHIMONT	Examineur

Version finale du 28 janvier 2003

À Florian, Corentin et Véronique.

Remerciements

Mes remerciements vont d'abord au professeur Jean-Charles Pomerol qui a accepté de diriger le jury de cette HDR. Ses conseils m'ont d'abord été utiles sur le plan scientifique où il a œuvré, par ses critiques, pour une plus grande clarté de ce mémoire. Il m'a ensuite permis de gérer l'aspect institutionnel, ce qui, dans des conditions parfois complexes, m'a permis d'arriver serein ou autant que faire se peut à la soutenance.

Je remercie le Professeur Régis Beuscart. Professeur en informatique médicale, il a accepté de se plonger dans mes digressions sur l'Ingénierie des connaissances et m'a fourni des retours riches d'enseignements à la frontière entre l'Ingénierie des connaissances et l'informatique hospitalière.

Je remercie le Professeur Yves Chiaramella. Il a accepté de rapporter sur un domaine éloigné de ses sujets de préoccupation. Ses critiques m'ont été d'une grande utilité et les discussions qui s'en sont suivies m'ont permis de préciser encore ma démarche.

Je remercie le Professeur Philippe Lorino. Pendant que je m'essayais à la pluridisciplinarité entre l'Ingénierie des connaissances et les sciences de gestion, il faisait une partie du chemin dans le sens inverse et ses critiques et discussions durant les séminaires Ingénierie des connaissances/Sciences de gestion m'ont conforté dans mes recherches. Je lui suis reconnaissant d'avoir accepté de rapporter sur ce travail.

Je remercie le Professeur Thomas Similowski. Médecin clinicien en réanimation pneumologique, il a été notre fournisseur de données et notre référence durant le projet HOSPITEXTE. Plus, la qualité de ses commentaires et sa faculté à expliquer les problèmes des relations entre la médecine et l'informatique font de ses analyses des sources de réflexions inépuisables.

Je tiens à remercier Catherine Garbay, sous-directeur du département STIC du CNRS. Durant ses années de direction du GdR-I3, elle a poussé et organisé la pluridisciplinarité du groupe. Son action en ce domaine puis au département STIC m'a encouragé à poursuivre mes réflexions pluridisciplinaires. Je suis ravi qu'elle ait accepté de juger ce travail.

Bruno Bachimont a accepté de relire ce mémoire en tant qu'ami mais, encore plus, critique intransigeant. Je l'en remercie encore.

Je partage avec mon ami Jacques Bouaud des communautés de vue et d'intérêt sur de nombreux sujets allant de la recherche à la typographie en passant par \LaTeX ou le Mac et Unix. Ce sont des « détails » qui n'en sont pas et qui, dans le travail, immunisent contre le syndrome de l'écran bleu du chapelier fou.

Enfin, je remercie Catherine Maksud pour ses relectures attentives.

Crédits

Ce document se veut une synthèse sur l'ingénierie des connaissances d'un côté, ses conséquences méthodologiques sur la gestion des connaissances médicales de l'autre.

Cela m'amène à décrire des travaux menés en collaboration avec des collègues, des travaux menés en collaboration avec des chercheurs extérieurs à la mission de recherche sur les sciences et technologies de l'information médicale (STIM), ainsi que des travaux sur lesquels je n'ai pas travaillé mais qui me permettent d'illustrer mon propos.

Il y a donc, à la fin de chaque chapitre, une section « crédits » où je précise le nom des différentes personnes impliquées dans les travaux décrits.

Dans cette optique et après les habituels remerciements, je voudrais citer ici les quelques personnes sans qui ce travail ne serait pas ce qu'il est.

Bruno Bachimont a d'abord été un collègue de travail. Il est ensuite devenu un ami puis finalement, pour moi, une référence en philosophie. Sans sa rigueur épistémologique que j'ai tenté de reprendre à mon compte, ce document ne serait pas ce qu'il est.

Régine Teulier et Nathalie Aussenac-Gilles ont toujours été à mes côtés durant ces années d'animation de la communauté Ingénierie des connaissances. L'une comme l'autre m'ont beaucoup apporté, que ce soit à l'occasion de l'organisation de séminaires, de décisions à prendre sur la conduite ou la pluridisciplinarité au sein de la communauté Ingénierie des connaissances, de projets communs et, plusieurs fois, de l'écriture d'articles scientifiques permettant de s'enrichir de points de vues différents et, pour moi, de développer une réflexion soumise à leur critique amicale.

Didier Bourigault, inventeur et développeur de l'outil LEXTER/SYNTAX a toujours été un collaborateur et un partenaire précieux, d'abord au sein du groupe TIA et depuis quelques années dans les collaborations que nous avons eu sur les problèmes d'indexation ou de constitution d'ontologies.

Chantal Reynaud et Philippe Laublet ont animé le GRACQ avant même que l'on parle d'Ingénierie des connaissances. Je les ai rejoints un peu plus tard. Depuis, nous sommes associés à des actions d'animation permettant de développer et de confronter nos réflexions. La dernière est l'action spécifique « Web sémantique » du département STIC du CNRS. Elle est toujours en cours et me permet, en leur compagnie, d'aborder le Web sémantique et les rapports d'icelui avec l'informatique médicale.

Jacques Bouaud, Brigitte Séroussi et Pierre Zweigenbaum sont mes collègues du STIM. Malgré des sujets de recherche souvent séparés et un environnement pas toujours propice, nous avons développé ensemble une communauté de vue sur les applications « textuelles » en médecine que nous avons pu tester et valider sur un certain nombre de projets. De nombreuses discussions avec l'un ou l'autre m'ont convaincu du bien fondé de ces approches, ont nourri et nourrissent encore ma réflexion sur la gestion des connaissances médicales.

Michel Daigne m'a initié au milieu, complexe, des réseaux de santé, sources de problèmes et de réflexions inépuisables. Avec d'autres membres du groupement Centrale-Santé, il m'a fait découvrir les réflexions de la communauté industrielle de Centrale-Santé dans le domaine de la bioinformatique et de l'informatique médicale.

Jérôme Euzenat est à l'origine de ce texte. C'est dans le mémoire de son habilitation à

diriger des recherches que j'ai pioché l'idée de ces crédits qui permettent, ici comme dans chaque chapitre, de rendre à César ce qui est à César.

Enfin, nombreuses sont les personnes que je n'ai pas nommées et à qui je dois de nombreuses interactions, d'intéressantes discussions durant des séminaires, des discussions au sein du bureau du GRACQ, des points de vue d'autres disciplines. Je sais ce que je leur dois et les remercie ici sans les nommer, sauf une mention spéciale à Pierre Tchounikine dont les réflexions épistémologiques sur les environnements informatiques pour l'apprentissage humain me confortent dans mes réflexions au sujet de l'Ingénierie des connaissances.

Le but de ce mémoire est de faire le point sur le domaine de l'Ingénierie des connaissances et de voir ses apports, présents et futurs, à la gestion des connaissances médicales. L'Ingénierie des connaissances ne peut pas résoudre à elle seule les problèmes de la gestion des connaissances médicales. Un certain nombre d'années passées à travailler dans ces domaines fondamentaux – pour l'Ingénierie des connaissances – et appliqués – pour la gestion des connaissances médicales – nous ont convaincu qu'il était nécessaire de mobiliser d'autres domaines dans une approche pluridisciplinaire. C'est que nous avons fait en entreprenant des recherches qui ressortissent à l'Ingénierie des connaissances mais aussi au domaine des systèmes d'information et à la gestion. C'est cette démarche que nous allons essayer de relater dans ce mémoire, en sollicitant les disciplines les unes après les autres et en concluant sur les axes de recherche qui nous semblent indispensables de développer.

Le 1^{er} chapitre est une introduction à l'acquisition des connaissances et à la façon dont le domaine rencontre les applications médicales. Nous y abordons les concepts et les modèles de l'acquisition des connaissances – *e.g.* modèle conceptuel, connaissances du domaine, connaissances de raisonnement, langages de modélisation, ontologies, etc. – et y rappelons rapidement les résultats que les travaux ayant concouru à ces recherches nous ont permis d'obtenir. Nous illustrons cela avec des travaux sur les modèles causaux et d'autres sur la validation. Nous terminons sur les questions laissées en suspens ou les sujets de réflexion vis-à-vis de l'ingénierie des connaissances comme des applications médicales.

Dans le 2^e chapitre, nous développons une analyse épistémologique de l'Ingénierie des connaissances, discipline jeune, au sein d'un domaine, l'Intelligence artificielle, à peine plus ancien. Cette analyse est nécessaire pour argumenter les hypothèses de travail de l'Ingénierie des connaissances et aborder les problèmes qu'elle permet de résoudre. Accessoirement – vis-à-vis de ce mémoire – la question du domaine scientifique et de l'enseignement de l'Ingénierie des connaissances est abordée.

Dans le 3^e chapitre, nous essayons de comprendre la genèse de quelques problèmes et l'apport des réflexions précédentes pour avancer dans un certain nombre de développements et d'applications. Ainsi, le caractère contextuel de l'information et ses rapports avec son support (ici informatique) nous amène à mettre l'accent sur l'intérêt du document textuel comme support et véhicule de la connaissance. Le projet de dossier médical informatisé HOSPITEXTE (comme d'autres développés au sein du service, *e.g.* ONCODOC) est présenté dans ce paradigme et la problématique de l'indexation des documents est abordée en décrivant le projet CISMED dans le contexte du Web sémantique.

Dans le 4^e chapitre, le plus long du mémoire, nous abordons la notion d'ontologie de façon approfondie en justifiant et prenant une position affirmée sur ce que nous

pensons que c'est et que ça n'est pas. Cette position nous semble nécessaire à la compréhension de ce que l'on peut faire ou non avec les ontologies comme avec de nombreux « produits » terminologiques tels que les thésaurus ou les bases de données lexicales. Nous prenons appui sur les travaux développés au sein de la communauté Ingénierie des connaissances, s'appuyant eux-mêmes sur des travaux plus anciens comme la terminologie ou la recherche d'information, puis nous développons la conception des ontologies en nous fondant sur des travaux et exemples récents, en médecine comme dans d'autres domaines. Nous défendons, à l'occasion de la description de travaux développés en collaboration, l'utilisation d'outils du traitement du langage naturel comme mode d'accès privilégié aux concepts d'une ontologie. Des ontologies en médecine, développées durant ces travaux servent d'exemples à nos propos. Nous terminons ce chapitre sur le problème de l'indexation en lien avec les ontologies.

Dans le 5^e chapitre, nous abordons le contexte de la mise en œuvre d'une gestion des connaissances médicales, à savoir celui des systèmes d'informations. Systèmes dont le champ d'action grandit, originellement limité à un même établissement de soins – hôpital ou clinique –, concernant maintenant des communautés de soins plus larges que sont les « réseaux de santé » qui s'étendent à la médecine de ville et impliquent l'ensemble des professionnels de santé pour une pathologie ou une situation donnée. Cette situation crée de nouvelles contraintes pour la mise en œuvre de ces larges systèmes d'informations des réseaux – des *systèmes d'information de santé* –, en particulier au niveau d'un de leurs buts premiers qui est d'assurer l'interopérabilité des différents acteurs du réseau. Cette interopérabilité passe et doit passer par des standards qui se construisent et s'imposeront lentement. Cette lenteur est nécessaire à la reconnaissance du bien fondé de ces standards et à leur acceptation. Cette acceptation étant, comme nous le verrons, très dépendante des contextes organisationnels et techniques des différents systèmes de soins et des systèmes d'information de santé sous-jacents.

Dans le 6^e chapitre, nous revenons aux origines de la gestion des connaissances en tant que problématique. Ces origines sont à retrouver au sein des entreprises et, pour le côté académique, chez les chercheurs en sciences de gestion. Cela nous permet de rechercher un certain nombre de points communs et de convergences avec l'Ingénierie des connaissances et de proposer un certain nombre d'indications, venues de la gestion comme de l'Ingénierie des connaissances, sur l'élaboration d'un système de gestion des connaissances en santé. À partir de là, nous repositionnons les enjeux de la gestion des connaissances médicales ainsi que leur spécificité au regard des problématiques développées dans les chapitres précédents. Les projets comme les développements normatifs sont resitués dans un contexte d'intégration. Enfin, nous essayons de tirer les conséquences de ces travaux dans le contexte des développements futurs attendus, en particulier au regard des nouvelles structures de soins qui se font jour comme les réseaux de santé.

Finalement, nous concluons ce mémoire sur la façon dont ont évolué nos recherches, centrées au début sur l'Ingénierie des connaissances et s'ouvrant, dans une approche pluridisciplinaire, à d'autres domaines.

Table des matières

Prologue	ix
1 De l'acquisition à l'ingénierie des connaissances	1
1 L'acquisition des connaissances	1
2 Modélisation des connaissances	2
2.1 Nature et contenu des modèles	2
2.2 Les connaissances du domaine des SESG	3
2.3 Les modèles de raisonnement	4
2.4 La combinaison réfléchie des modélisations domaine et raisonnement	4
3 Le processus de modélisation	5
3.1 Un processus en quatre étapes	5
3.2 Des architectures spécialisées	6
3.3 Des architectures génériques	7
3.4 Des méthodes descendantes <i>versus</i> ascendantes	8
3.5 Des langages de modélisation <i>versus</i> opérationnalisation	10
4 Réflexions et questions	12
4.1 Nature et rôles des systèmes développés	12
4.2 Les ontologies	13
4.3 Les langages de représentation	13
4.4 L'organisation et la gestion	14
4.5 Pour terminer	14
2 L'ingénierie des connaissances	15
1 Introduction	15
2 Épistémologie et définitions	17
2.1 De l'information à la connaissance	17
2.2 Des systèmes à base de connaissances	19
3 Programme de travail	20
3.1 En guise de programme	21
3.2 En guise de méthode	21
3.3 En guise de champ d'action	22
4 Exemples	22
4.1 HOSPITEXTE	22
4.2 Des thésaurus à la construction d'ontologies	23
4.3 Du contexte partagé	24
5 Un domaine scientifique, un enseignement?	25

5.1	Des épistémologies positivistes aux épistémologies constructivistes	25
5.2	L'Ingénierie des connaissances	26
6	Conclusion	28
3	Des réflexions aux applications "textuelles"	29
1	L'information médicale et son contexte ou la primauté du texte	29
1.1	La spécificité textuelle de l'information médicale	29
1.2	XML et la représentation des connaissances médicales	30
2	Le projet HOSPITEXTE	31
2.1	Du support papier au support électronique	31
2.2	Des objectifs	33
2.3	Un prototype	34
2.4	Perspectives	35
3	Une aide à la décision « textuelle »	37
4	CISMeF	38
5	Au sujet du Web sémantique	40
4	De nécessaires ontologies ?	43
1	Introduction et définitions	43
1.1	Les ontologies en IC	43
1.2	Repérer, organiser les objets du monde	45
1.3	Que représente-t-on dans une ontologie?	48
1.4	Quelle utilité?	48
2	Quelle méthodologie de construction?	49
2.1	Au commencement était la langue	49
2.2	Une méthodologie constructiv(ist)e	50
2.3	Quelques bons principes	54
2.4	Une ontologie en médecine : MENELAS	55
3	Quelques questions autour des ontologies	56
3.1	Deux exemples pas trop « jouets »	56
3.2	Ontologies <i>versus</i> base de connaissances	59
3.3	Ontologies <i>versus</i> thésaurus	59
3.4	Quelle utilité en médecine?	62
4	Acquérir des ontologies à partir de corpus	64
4.1	Introduction épistémologique	64
4.2	L'usage d'un outil de TAL	65
4.3	Une ontologie en réanimation chirurgicale	65
5	De la réutilisation aux ontologies génériques	67
5.1	La réutilisabilité des ontologies	67
5.2	Des ontologies génériques	68
6	L'indexation	68
6.1	Introduction	68
6.2	Une question de normalisation	69
6.3	Difficultés et perspectives	70
6.4	Indexer la médecine	71
7	Des ontologies aux textes : le web sémantique	72
8	Conclusion	73
5	Des systèmes d'information de santé et de l'interopérabilité	75
1	L'interopérabilité	75
2	L'information médicale et son contexte	76
3	De l'échange de données	77
3.1	XML et l'EDI	77
3.2	HL7/CDA : a « Clinical Document Architecture »	77

4	Un standard d'enveloppe	80
4.1	Contexte de la proposition	80
4.2	La notion d'enveloppe d'échange médical	81
5	Comparaison des approches	81
5.1	Une information textuelle disponible sur un serveur Web.	82
5.2	La normalisation de l'information <i>versus</i> normalisation des messages.	83
5.3	L'échange d'extraits de dossier multimédia	84
6	Discussion et perspectives	84
6.1	De l'interopérabilité sémantique	84
6.2	Du Web sémantique	84
6.3	Des standards pour l'information médicale	85
7	Conclusion	86
6	La gestion des connaissances médicales	89
1	Les connaissances de la gestion	89
1.1	Les connaissances organisationnelles	89
1.2	Une solution <i>versus</i> un symptôme	91
1.3	De la gestion à l'ingénierie : convergences épistémologiques	92
2	L'hôpital et la gestion	93
2.1	Les affres du pilotage organisationnel hospitalier	93
2.2	Enseignements : des SBC dans une organisation	95
2.3	Quelle gestion des connaissances médicales?	96
3	La gestion des connaissances médicales	97
3.1	Les enjeux de la gestion des connaissances médicales	97
3.2	Le dossier médical, objet focal de la gestion des connaissances	99
3.3	La transformation et le cycle de vie des connaissances médicales	100
4	Des perspectives de recherche	101
4.1	Perspectives dans le domaine de l'ingénierie des connaissances	101
4.2	Perspectives dans le domaine des systèmes d'information	103
4.3	Perspectives organisationnelles	104
5	Les réseaux de santé : de nouveaux modèles de fonctionnement	105
6	Conclusion et perspectives	107
	Épilogue	109
	Références	111
	Liste des figures	121
	Index	126

De l'acquisition à l'ingénierie des connaissances

Ce chapitre est une introduction à l'acquisition des connaissances et à la façon dont le domaine rencontre les applications médicales. Nous y abordons les concepts et les modèles de l'acquisition des connaissances – e.g. modèle conceptuel, connaissances du domaine, connaissances de raisonnement, langages de modélisation, ontologies, etc. – et y rappelons rapidement les résultats que les travaux ayant concouru à ces recherches nous ont permis d'obtenir. Nous illustrons cela avec des travaux sur les modèles causaux et d'autres sur la validation. Nous terminons sur les questions laissées en suspens ou les sujets de réflexion vis-à-vis de l'ingénierie des connaissances comme des applications médicales.

1 L'acquisition des connaissances

L'acquisition des connaissances est apparue comme une discipline avec un objet de recherche à la fin des années 80. À la suite des développements des systèmes experts durant la décennie, la question de la modélisation et de l'acquisition des connaissances pour ces systèmes était apparue comme cruciale et problématique : la question du « goulet d'étranglement »¹ de l'acquisition des connaissances justifia de nombreux travaux de thèses, que ce soit avec des problématiques très cognitives (Aussenac, 1989) ou plus orientés vers des questions de niveau de représentation (Reynaud, 1989; Charlet, 1989, 1992)².

Ces travaux étaient le début d'une réflexion au sujet des systèmes experts de seconde génération (SESG). Dans ce nouveau paradigme, le système expert devenait un système à bases de connaissances (SBC) et les modèles de connaissances de ces bases devenaient multiples et, surtout, un sujet de réflexion intense. Problématique que nous allons développer ici.

¹Ce goulet correspondait au fait que le système expert devait permettre de « capturer » l'expertise d'un expert et la représenter au sein du système expert de façon à ce que le système se comporte comme l'expert sollicité dans la même situation. Les difficultés à capturer précisément cette expertise, ont amené à parler de goulet d'étranglement.

²Il n'est pas de notre propos de recenser les travaux en acquisition des connaissances à cette époque, raison pour laquelle nous n'en citons que quelques un ici parmi des travaux personnels et des collaborations. L'ensemble du chapitre fournira une bibliographie plus complète.

2 Modélisation des connaissances

2.1 Nature et contenu des modèles

Les travaux en Ingénierie des connaissances sont donc fondés sur la nécessité de modéliser de manière explicite les connaissances³. Il s'agit de construire des modèles adaptés à la nature des connaissances à décrire pour pouvoir ensuite les représenter dans des formalismes adéquats. Dans le cadre des SBC, trois types de connaissances font l'objet de modèles distincts avec des primitives de modélisation propres : les connaissances du domaine, les tâches et les méthodes⁴. L'ensemble de ces modèles forme alors ce qu'on appelle communément en Ingénierie des connaissances, le *modèle conceptuel*. L'analyse du domaine et l'analyse des tâches d'un SBC sont à rapprocher des analyses statiques et fonctionnelles effectuées en génie logiciel. En revanche, l'analyse de la dynamique en génie logiciel consiste à étudier les conséquences d'événements extérieurs sur une application alors que pour un SBC, il s'agit de rendre compte du contrôle interne du comportement du système à implémenter. En effet, le problème de l'efficacité d'un SBC ne peut être posé uniquement en termes d'efficacité algorithmique mais va également se poser par rapport aux connaissances utilisées ; celles-ci devant alors être décrites de manière déclarative en phase d'analyse.

La façon dont sont décrites les connaissances conditionne la construction du SBC et surtout la compréhensibilité de son fonctionnement. L'acquisition et l'Ingénierie des connaissances ont repris à leur compte les travaux de Newell (1982) qui, le premier, a différencié les connaissances à représenter dans un système et l'implémentation de ce système. Newell a fait apparaître la nécessité d'un niveau de description des systèmes qui ne soit pas celui des symboles et langages informatiques, le *niveau des connaissances*. Ce niveau de description devait permettre de décrire le comportement du système observé indépendamment de son implémentation formelle. Le système y était de plus considéré comme un agent rationnel qui dispose de connaissances, doit atteindre des buts, sait effectuer des actions, et est rationnel, c'est-à-dire qu'il choisit (avec ses connaissances) l'action suivante qui va le mener le plus directement au but.

Sans discuter les questions cognitives que posent des systèmes, agents rationnels, notons que ces hypothèses de travail ont été extrêmement productives pour l'Ingénierie des connaissances puisqu'elles ont amené les chercheurs à réfléchir en particulier à la façon de représenter les connaissances et à faire l'hypothèse d'un modèle conceptuel indépendant du système opérationnel (cf. § 2.4 et § 3).

Dans les deux sections suivantes⁵(cf. § 2.2 et § 2.3), nous allons discuter des connaissances du domaine et des connaissances de raisonnement de façon séparée puis en reliant les problématiques (cf. § 2.4). Les travaux cités à cette occasion abordent généralement les deux problématiques mais ils sont cités selon que leur axe de réflexion concerne un type de connaissance ou l'autre. Ainsi, la question de la modélisation du domaine des SESG puis les travaux sur les modèles causaux, s'ils ressortissent à des questions de raisonnement, restent dans un paradigme où les connaissances du domaine et du raisonnement ne sont pas explicitées de façon séparée et où l'accent est mis sur la modélisation du domaine sans toujours d'explicitation spécifique d'un modèle de raisonnement. Ils sont pour cela abordés au niveau des connaissances du domaine.

³Ces méthodes sont des méthodes de raisonnement, enchaînement d'opérateurs s'appliquant sur des données. Ce ne sont pas les méthodes des langages à objets.

⁴Nous parlons d'acquisition des connaissances puis passons à l'Ingénierie des connaissances. Cela vient évidemment du fait qu'il n'y a pas une frontière nette entre deux problématiques distinctes mais une évolution d'une problématique centrée sur les seuls systèmes experts à une problématique qui élargit son champ méthodologique comme nous allons le voir par la suite. Élargissement qui s'accompagne, comme nous le verrons au chapitre 2 de nouvelles réflexions et applications. En attendant, nous sommes passés « discrètement » de l'acquisition à l'Ingénierie des connaissances.

⁵La question des connaissances et de leur représentation sera discutée plus précisément au chapitre 2.

2.2 Les connaissances du domaine des SESG

La notion de modèle du domaine trouve son origine dans les travaux sur le raisonnement à partir de modèles développés dans les années 80 (De Kleer, 1987; De Kleer & Williams, 1989; Genesereth, 1982; Reiter, 1987). Les connaissances du domaine d'un SBC sont les connaissances relatives au domaine de l'application et nécessaires pour que les méthodes de raisonnement puissent s'exécuter. Les travaux qui ont porté sur ces connaissances depuis une décennie ont eu un double impact : d'une part, ils ont montré l'intérêt de distinguer les connaissances du domaine selon leur nature et de raisonner sur des modèles de connaissances multiples en exploitant les spécificités de chacun d'eux ; d'autre part, ils ont montré l'intérêt de disposer de modèles de connaissances structurés, exprimés à l'aide de langages ayant une sémantique bien définie et exprimés à différents niveaux de granularité.

La coopération entre plusieurs modèles du domaine a été étudiée dans le cadre des travaux sur les SESG (Steels, 1985; Charlet, 1992)⁶ dont l'objectif premier était de combiner raisonnement heuristique et raisonnement à partir de modèles (ou raisonnement fondé sur des connaissances de surface versus raisonnement fondé sur des connaissances profondes selon Kayser (1988)). C'est ainsi que l'idée d'accorder de l'importance à la modélisation des connaissances du domaine s'est imposée et que le concept « d'ontologie », *ensemble des objets reconnus comme existant dans le domaine*, est apparue en Ingénierie des connaissances avec, dès le départ, l'idée que ces connaissances seraient réutilisables d'une application à l'autre.

Nous avons approfondi cette question des modèles dans un travail sur les modèles causaux que nous allons décrire maintenant.

2.2.1 Les modèles causaux

En collaboration avec Ch. Reynaud et J.-P. Krivine, nous avons réalisé un travail qui visait à comparer les modèles causaux utilisés dans nos propres expériences d'acquisition des connaissances pour des SESG. Les outils analysés et comparés étaient ACTE (Charlet, 1993a), ADELE (Reynaud, 1993) et DIVA (David & Krivine, 1989).

Dans les modèles conceptuels de ces systèmes, nous avons des connaissances heuristiques, dites « de surface », et des connaissances causales censées exprimer des connaissances génériques, capturant les lois qui gouvernent un domaine. Elles représentent, pour cela, la *théorie du domaine* et sont alors dites « profondes ».

Les principaux résultats obtenus ont été l'identification de deux grandes catégories de modèles causaux avec des insertions spécifiques dans le processus d'acquisition des connaissances (Charlet *et al.*, 1996b) :

Bottom-Up designed Causal Models (ou modèles causaux conçus selon une démarche ascendante). Ces modèles sont construits indépendamment du processus d'acquisition des connaissances et, issus d'ouvrages, ils décrivent comment les choses fonctionnent ou tombent en panne. Le système ADELE comporte ce type de modèle dans le domaine de l'électromyographie et permet ainsi, dans le cadre du processus d'acquisition des connaissances, de valider les connaissances heuristiques de l'expert avec le modèle causal.

Top-Down designed Causal Models (ou modèles causaux conçus selon une démarche descendante). Ces modèles sont développés durant le processus d'acquisition des connaissances. Ils sont limités aux concepts de l'application et apparaissent comme des justifications des connaissances heuristiques qu'ils aident à recueillir. Dans le processus d'acquisition, le modèle causal et les connaissances heuristiques sont

⁶Voir aussi le livre « Second Generation Expert System » (David *et al.*, 1993) où sont développés la plupart des travaux à ce sujet, que ceux-ci soient plus axés sur les connaissances du domaine ou sur les connaissances de raisonnement.

acquis simultanément même si, comme nous l'avons déjà dit, le premier apparaît comme une justification ou le cadre conceptuel du second. C'est comme cela qu'on été acquises les connaissances dans le cadre de ACTE appliqué aux douleurs aiguës de l'abdomen : à partir d'un modèle causal entre processus physiopathologiques et d'heuristiques d'évocations de diagnostics (signes, syndrome), l'outil permet de construire des connaissances heuristiques complètes, consistant en un ensemble de règles de productions.

Ce travail nous a ainsi permis de mieux appréhender les types de modèles mis en œuvre dans l'acquisition des connaissances et d'avancer vers ce que sera le modèle conceptuel et son acquisition (*cf.* 3).

2.3 Les modèles de raisonnement

Les modèles de raisonnement décrivent de façon abstraite le processus de résolution à mettre en œuvre dans un SBC en termes de tâches et de méthodes, les tâches étant réalisées par des méthodes. Une tâche est une description de ce qui doit être fait dans l'application en termes de buts et de sous-buts. Elle se définit par des connaissances de sortie obtenues à partir des connaissances d'entrée, et ce en fonction de contraintes et ressources disponibles. Pour décrire une résolution de problèmes, on peut mettre en évidence des tâches de différents niveaux, les tâches de plus bas niveau poursuivant des sous-buts pour les tâches plus générales. Les méthodes décrivent comment un but peut être atteint en termes d'une série d'opérations et d'un ordre de réalisation. Deux types de méthodes sont généralement distinguées : celles qui consistent à décomposer une tâche en sous-tâches et celles qui mettent en œuvre une procédure élémentaire qui atteint directement un but. La distinction explicite faite entre les concepts de tâches et de méthodes n'a pas toujours été adoptée. Elle a été proposée initialement par L. Steels dans son approche « componentielle » (1990). Cette distinction, aujourd'hui reconnue par l'ensemble des travaux en Ingénierie des connaissances (Klinker *et al.*, 1991; Puerta *et al.*, 1992; Schreiber *et al.*, 1994; Tu *et al.*, 1995), présente l'avantage de décrire séparément le but visé de la façon de l'atteindre. Elle rend ainsi possible la définition explicite de différentes façons d'atteindre un même but par l'association de plusieurs méthodes à une même tâche.

2.4 La combinaison réfléchie des modélisations domaine et raisonnement

À la suite des travaux sur les SESG, l'intérêt de construire un modèle conceptuel d'une application en combinant réutilisation de composants de modèles de raisonnement et abstraction de connaissances du domaine a été reconnu. Une analyse des connaissances du domaine devint alors nécessaire pour mettre en correspondance les connaissances du domaine et leurs rôles dans le raisonnement (Aussenac-Gilles *et al.*, 1997). Les connaissances du domaine sont alors décrites en un noyau, l'ontologie du domaine regroupant les entités du domaine et les relations entre ces entités, et une couche extérieure, le modèle du domaine décrivant des connaissances du domaine plus complexes en combinant les connaissances répertoriées dans l'ontologie et en précisant un certain nombre de propriétés.

L'importance accordée à la modélisation des connaissances du domaine n'a pas eu toutefois un impact immédiat dans les travaux en Ingénierie des connaissances. L'accent a été d'abord mis sur la construction du modèle de raisonnement d'une application, une étape majeure du processus de modélisation. Les connaissances du domaine à acquérir ont alors été considérées comme étant celles nécessaires au modèle de raisonnement pour s'exécuter. Elles étaient désignées par leur rôle dans le raisonnement, ce qui guidait leur acquisition.

Afin de disposer de descriptions de raisonnement réutilisables, les modèles de raisonnement ne réfèrent aux connaissances du domaine qu'indirectement. Seuls les rôles joués par les connaissances du domaine dans le raisonnement sont spécifiés. Les termes utilisés à cet effet (symptômes, hypothèses, diagnostic, etc.) appartiennent au vocabulaire d'ontologies dites de méthodes ou de modèles (cf. CommonKADS – Schreiber *et al.*, 1994).

Ces modèles, comme tout modèle, sont réducteurs. Ils simplifient la vision que l'on peut avoir du système, un des buts étant justement de réduire la complexité en fournissant une représentation synthétique permettant de se focaliser sur les aspects importants dans le développement d'un SBC. Ces séparations dans les descriptions offrent trois points de vue différents qui interagissent pour donner une spécification du système au sein d'un modèle cohérent.

3 Le processus de modélisation

3.1 Un processus en quatre étapes

Plusieurs analyses décrivent la construction du modèle d'une application sous la forme d'un processus qui comporte quatre phases et sert plus ou moins de référence (cf. fig. 1.1 – Aussenac-Gilles *et al.*, 1992; Linster, 1992) :

1. le recueil de données brutes,
2. la construction d'un schéma de modèle conceptuel,
3. la définition du modèle conceptuel complet,
4. l'implémentation de ce dernier dans une base de connaissances opérationnelle.

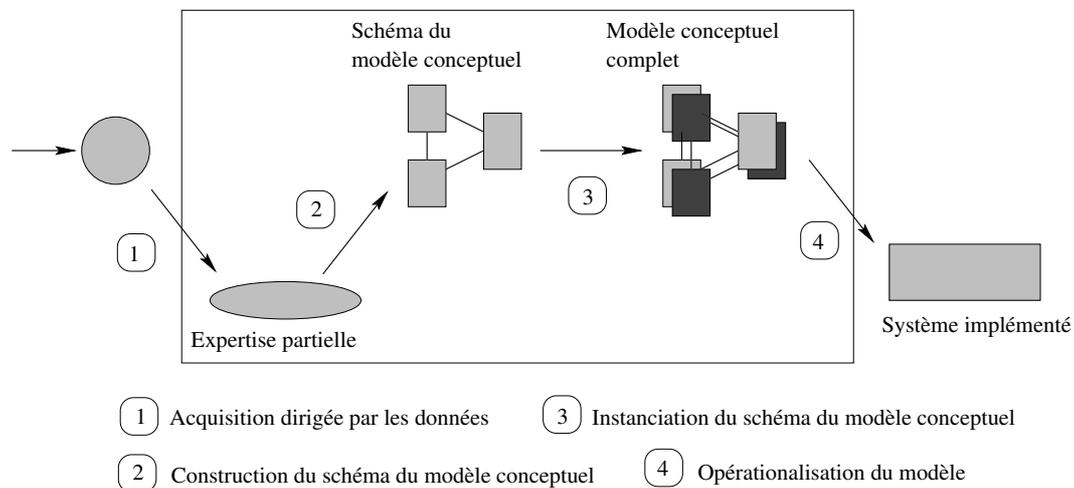


FIG. 1.1 – L'acquisition des connaissances conduite par les modèles (tiré de (Aussenac-Gilles *et al.*, 1992)).

La phase 1 a été étudiée à travers des travaux de recherche en psychologie, ergonomie et IA. Ces travaux ont pu proposer des méthodes informelles comme des interviews et d'autres plus formelles comme les grilles répertoires (Boose & Bradshaw, 1987). Toutes correspondent à des démarches dites ascendantes (Aussenac, 1989). Plus récemment, des outils d'analyse de corpus permettent de proposer des méthodes performantes, en particulier pour recueillir des données brutes en vue de l'élaboration de connaissances ontologiques (cf. § 4).

Les phases 2 et 3 concourent à l'élaboration du modèle conceptuel : La phase 2 se préoccupant d'éliciter un schéma de ce modèle – *i.e.* un schéma des connaissances du domaine et un schéma des connaissances de raisonnement ; la phase 3 consistant à enrichir ces schémas avec les données du domaine d'application. Si ce découpage est intéressant, il n'est pas toujours exactement suivi dans la réalité dans la mesure où le schéma peut se construire durant l'élaboration du modèle conceptuel, en particulier dans les méthodes ascendantes (*cf.* § 3.4).

Par ailleurs, au sein de ce processus, il est clair que nous ne sommes plus dans le paradigme d'un modèle permettant de représenter l'expertise de l'expert (*cf.* § 1) : le modèle conceptuel se veut en fait une spécification fonctionnelle de la partie résolution de problème de l'artefact à construire. On passe ainsi de *l'observation* à *l'interprétation*. Dans ce contexte, le modèle conceptuel est un langage d'expression des connaissances au sens de Newell (*cf.* § 2.1) : c'est d'un côté un langage partagé par l'expert, le cognitif, de l'autre un langage « compréhensible » par l'artefact.

La phase 4 correspond au choix d'un langage pour l'implémentation du système final. Les différentes recherches dans ce domaine ont porté et portent sur la nature de ce langage et ses capacités inférentielles. Cette opérationnalisation est souvent l'occasion de réaliser des validations sur le comportement et l'utilisation des connaissances du système réel (*cf.* § 3.5.1).

Ce rapide phasage du processus de modélisation effectué, un certain nombre de points doivent être approfondis pour mieux comprendre le contexte de développement des travaux en Ingénierie des connaissances. C'est ce à quoi nous nous attelons maintenant en commençant par l'architecture des outils d'aide à l'élaboration des SBC, selon qu'elle met l'accent sur l'obtention d'un modèle exécutable – *i.e.* opérationnel – ou sur l'élaboration du modèle conceptuel.

3.2 Des architectures spécialisées

Les travaux qui relèvent de la première approche considèrent que la méthode mise en œuvre dans une application conditionne les types de connaissances nécessaires et que le rôle de ces connaissances dans la méthode contraint la façon dont elles doivent être représentées et recueillies. Une telle approche suppose alors d'identifier au préalable la tâche à résoudre et la méthode de résolution de problèmes à mettre en œuvre – *e.g.* *propose-and-revise* ou *heuristic classification* (Marcus & McDermott, 1989) – avant de choisir l'architecture adaptée et d'acquérir les connaissances du domaine nécessaires à cette mise en œuvre.

Rentrent dans cette catégorie, l'approche des Tâches Génériques de Chandrasekaran (Chandrasekaran, 1986), les travaux sur les méthodes à limitation de rôles de McDermott (Marcus & McDermott, 1989). De telles approches mettent l'accent sur le concept de méthode de résolution de problème, laquelle doit être identifiée et pilote la tâche d'acquisition.

D'autres travaux comme ceux de Musen (1989) s'appuient plus directement sur les composants de l'application elle-même et plus particulièrement sur l'ontologie du domaine et sur l'ontologie des méthodes de résolutions de problèmes pour générer les outils et les interfaces permettant, d'un côté, de créer les instances des concepts de l'ontologie et les valeurs de leurs attributs et, d'un autre côté, de créer les méthodes de résolutions de problèmes. Par exemple, dans le domaine médical, PROTÉGÉ génère une interface graphique d'élaboration de guides de bonnes pratiques – *i.e.* un diagramme de branchement – à partir de l'ontologie des objets constituant ces guides – *i.e.* opérateurs de branchement, conditions, etc. – (Musen, 98).

De telles architectures facilitent grandement la phase 3 d'instanciation du modèle conceptuel dans la mesure où, le schéma étant fixé par l'architecture, il ne reste plus qu'à acquérir les connaissances du domaine nécessaires à la construction complète du

modèle conceptuel. Mais ces architectures ne sont utilisables que si il a été possible d'identifier précisément le raisonnement (la tâche) à mettre en œuvre et que l'architecture adaptée à ce raisonnement existe.

3.3 Des architectures génériques

Les architectures qui relèvent de cette seconde approche proposent principalement de construire un schéma de modèle conceptuel par assemblage de briques génériques. Schéma qu'il ne restera plus qu'à instancier avec les connaissances du domaine. La généralité est assurée par le niveau d'abstraction et de description de ces briques qui leur permettent de s'appliquer quels que soient la tâche, le domaine, et le raisonnement mis en œuvre dans l'application à développer.

Le projet européen CommonKADS est l'exemple le plus connu de ce type d'approche. Faisant suite au projet KADS (Wielinga *et al.*, 1992), il avait pour objectif d'intégrer d'autres approches développées en Ingénierie des connaissances, en particulier les travaux de Luc Steels (1990; 1993) sur les composants de l'expertise. La structuration des connaissances qu'il propose reste une référence par rapport à laquelle la plupart des autres travaux du domaine se sont comparés. Les deux principales caractéristiques de ce projet vis-à-vis de la modélisation des connaissances sont les suivantes (Schreiber *et al.*, 2000) :

Une architecture en trois couches. Les connaissances sont répertoriées dans un modèle à trois couches dit *modèle de connaissances*. La première couche correspond aux connaissances du domaine ; la deuxième couche répertorie les connaissances d'inférence : y sont décrites les étapes d'inférences basiques nécessaires à la construction du modèle de raisonnement. Les connaissances du domaine y sont utilisées par rapport aux rôles qu'elles jouent dans le raisonnement. Deux exemples dans le domaine du diagnostic médical sont l'inférence *générer une hypothèse* qui associe des symptômes avec une maladie possible et l'inférence *vérifier* qui identifie les tests qui peuvent être utilisés pour confirmer ou infirmer qu'une maladie a causé ou non les symptômes observés. La troisième couche décrit la tâche, c'est-à-dire les buts poursuivis au sein de l'application. Toujours dans le domaine médical, une tâche peut être le *diagnostic systématique* qui définit ce qui peut être réalisé à travers une séquence cyclique de générations d'hypothèses et de vérification (*cf.* fig. 1.2).

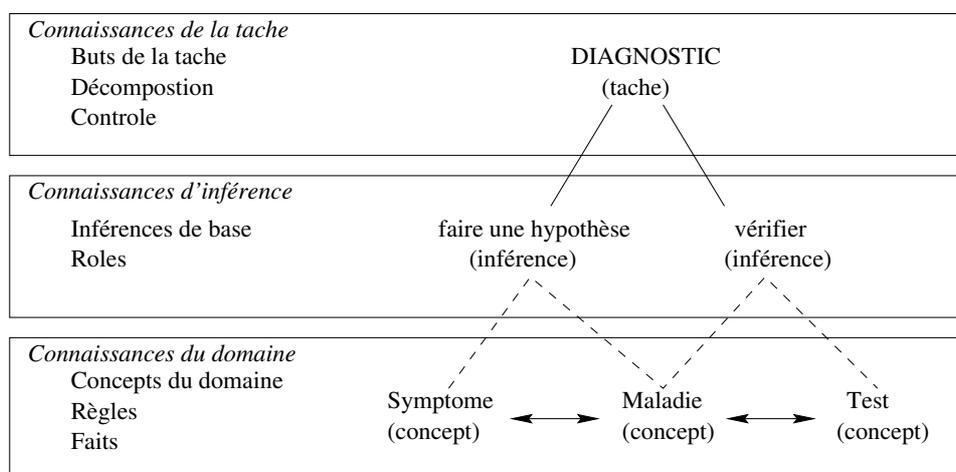


FIG. 1.2 – Le modèle en couche de KADS avec des exemples médicaux. Tiré de (Schreiber *et al.*, 2000).

Une bibliothèque de tâches. CommonKADS propose une bibliothèque de tâches génériques structurées sous forme de *patrons de tâches*. Les grands types de tâches proposées à la réutilisation sont les tâches d'analyses – *e.g.* classification, diagnostic, prédiction – ou de synthèse – *e.g.* conception, planning, assignement – parmi d'autres. Ces patrons sont organisés dans une taxinomie de tâches au sein de laquelle on peut les sélectionner en fonction de la définition de la tâche correspondante, de son but, de ses entrées et sorties. De plus, ils proposent une décomposition en sous-tâches et parfois une proposition de structure d'inférences typique ou nécessaire pour mener la tâche à bien si le découpage en tâches primitives est suffisamment précis. Le modèle en couches est une des réponses à la question de la généralité des modèles et donc de leur réutilisation. En particulier, la description des méthodes d'inférence dans des termes indépendants des connaissances du domaine permet leur description de façon indépendante et leur utilisation au sein des patrons de tâches.

En dehors de la question de la réutilisation sur laquelle nous reviendrons plus loin, la principale critique que l'on peut faire à ce projet, du point de vue de la modélisation, est d'offrir en fait peu de guides pour construire le schéma du modèle conceptuel, notamment pour sélectionner et instancier les tâches à partir de la bibliothèque (*cf.* 3.4).

Enfin, ce type d'approche inclut des préoccupations communes avec le génie logiciel en proposant des guides méthodologiques pour gérer les cycles de développement du système, en particulier au niveau des différents modèles à concevoir⁷. De façon générale, on constate une convergence entre l'Ingénierie des systèmes d'information et l'Ingénierie des connaissances : la première cherchant dans la seconde des moyens de traiter des informations de plus en plus difficiles à interpréter et donc des moyens de construire des modèles adaptés à cette nouvelle situation (Cauvet *et al.*, 2001), la seconde cherchant auprès de la première des méthodes de développement industrielles : c'est patent dans l'évolution du projet CommonKADS où les auteurs présentent maintenant une méthodologie intégrée de gestion des connaissances, que ce soit du point de vue de l'Ingénierie des connaissances ou du génie logiciel (Schreiber *et al.*, 2000).

3.4 Des méthodes descendantes *versus* ascendantes

Un autre point de vue sur les propositions d'architecture de l'Ingénierie des connaissances est d'étudier la façon dont est élaboré le schéma du modèle conceptuel : de façon *descendante*, par sélection d'un modèle générique (*cf.* 3.3) *versus* de façon *ascendante*, par abstraction à partir des données brutes de l'expertise (van Heijst *et al.*, 1997).

La première approche que nous venons de décrire, et dont CommonKADS est paradigmatique, est censée permettre la construction du schéma du modèle conceptuel en sélectionnant sa description dans une bibliothèque de modèles qui vont être ainsi réutilisés. L'acquisition des connaissances n'est ensuite plus qu'un *remplissage de rôles* où le modèle est instancié avec les connaissances du domaine. La seconde approche suppose l'acquisition du schéma du modèle conceptuel sans *a priori* sur la tâche à effectuer qui se dégagera de l'analyse des données brutes ni sur les données elles-mêmes nécessaires à ce travail. MACAO est un exemple d'outil proposant une telle méthode (Aussenac, 1989) qui est *a priori* plus complexe à mettre en œuvre qu'une méthode descendante.

En pratique, les choses sont plus nuancées, chacune des approches souffrant de défauts et qualités opposées, comme étudié dans (Lépine & Aussenac-Gilles, 1996) :

- Une méthode descendante suppose ainsi une bibliothèque où le « bon » modèle sera disponible et possiblement sélectionnable en fonction d'un certain nombre de caractéristiques de la tâche à effectuer (indexation et sélection des composants

⁷Les patrons de tâches sont aussi à rapprocher des patrons de conception au sein des systèmes d'information (Cauvet *et al.*, 2001)

génériques). Elle suppose aussi que le modèle sera adaptable à la tâche exacte (adaptation du composant). Par ailleurs, le niveau d'abstraction du composant est tel qu'il est inaccessible à un expert du domaine et doit être retraduit dans ses termes.

- Une méthode ascendante suppose un long travail d'analyse initiale et un travail de modélisation aussi long en raison d'une absence de modèle d'origine auquel se conformer. L'abstraction du modèle élaboré doit permettre d'exprimer l'expertise en des termes plus généraux et de construire un modèle de l'application (Lépine & Aussenac-Gilles, 1996). Mais, en pratique, le niveau d'expression des connaissances est proche du domaine et donc de l'expert.

Par rapport aux modèles génériques, on peut constater que la caractérisation des approches, ascendante *versus* descendante, est prototypique et que les approches réelles se conforment plus ou moins à l'un ou l'autre modèle sur un axe où, à une extrémité, la réutilisation est fortement recherchée (approche descendante) et, à l'autre extrémité, la réutilisation est moins recherchée au bénéfice d'une meilleure caractérisation de la tâche à effectuer (approche ascendante). En particulier, il est possible de mettre en œuvre une approche ascendante en ayant un type de modèle générique en tête et d'avoir ainsi une idée plus précise des connaissances recherchées dans l'analyse des données brutes. De la même manière, il est possible de mettre en œuvre une approche descendante où les modèles sélectionnés sont de très haut niveau et où l'instanciation demande des processus d'élicitation des connaissances du domaine.

Enfin, la question de la réutilisation traverse la caractérisation des approches descendantes *versus* ascendantes :

- Soit elle est recherchée dans les approches descendantes dans la description de modèles génériques réutilisables *par hypothèse*. Dans ce cas, l'on rencontre un certain nombre de problèmes liés au niveau d'abstraction des tâches génériques et de leur instanciation dans la réalité (*cf. supra*) mais aussi des problèmes liés au fait qu'une application met souvent en œuvre plusieurs raisonnements auxquels correspondent plusieurs modèles d'interprétation et qu'il faut alors pouvoir distinguer explicitement ces différents raisonnements.
En effet, si la réutilisation a pour but d'abaisser les coûts de développement, l'expérience montre que la réutilisation de méthodes générales, très largement applicables et donc peu contraignantes par rapport à la structure des connaissances du domaine manipulées, a un coût important, alors que les méthodes, d'application plus limitée, sont plus facilement réutilisables (Talbi & Laublet, 2000). Cette constatation a ouvert la voie à des approches plus flexibles, au sein desquelles les éléments réutilisés et adaptés sont de granularité plus fine. Il ne s'agit plus de réutiliser des modèles génériques complets de raisonnement mais des éléments de raisonnement. Les critères de sélection des composants génériques restent proches de ceux permettant la sélection des modèles d'interprétation. Ils portent sur les buts (ou sous-buts) que l'on cherche à atteindre et les caractéristiques des connaissances du domaine. Les buts sont rapprochés des compétences de méthodes ou mécanismes de résolution. Les caractéristiques des connaissances du domaine sont rapprochées de celles que doivent posséder les connaissances nécessaires à l'exécution du raisonnement (Reynaud, 1999).
- Soit elle est recherchée dans les approches ascendantes dans la mise en œuvre d'une méthodologie de travail avec l'expert. Dans ce cas, la réutilisation sera à rechercher dans la méthode de travail plutôt que dans le produit fini. Question qui sera abordée et débattue dans la communauté Ingénierie des connaissances, en particulier au niveau des ontologies (*cf. chap. 4*).

3.5 Des langages de modélisation *versus* opérationnalisation

Ces dernières années, de nombreux langages de description des connaissances⁸ ont été proposés pour décrire les modèles construits aux différentes étapes du cycle de développement d'un SBC. Certains ayant des propriétés et une couverture précise d'une ou deux étapes. D'autres cherchant à couvrir l'ensemble du processus. On a ainsi des langages de modélisation des connaissances, des langages de représentation ou d'opérationnalisation des connaissances.

La motivation de ces langages est liée aux connaissances que l'on veut décrire au sein d'un SBC. On peut ainsi vouloir : (1) acquérir, valider, mettre à jour le modèle conceptuel, (2) déboguer le système opérationnel construit dans les termes de l'expertise et non au seul niveau du code ou (3) fournir des explications du comportement du système. Sans chercher à faire des développements exhaustifs, nous allons caractériser les grandes classes de langages développés, en fonction des opérations qu'ils permettent (ou ne permettent pas) et de leurs propriétés.

Simuler, valider le modèle conceptuels. Des langages comme MODEL-K (Karbach *et al.*, 1991), ZOLA (Tchounikine *et al.*, 2000), OMOS (Linster, 1992) cherchent à simuler et donc valider le modèle conceptuel. Cela permet alors d'envisager de tracer le raisonnement en termes de buts et méthodes, etc. et de valider fonctionnellement des parties du modèle conceptuel (tâche, méthode, ...).

Inversement, ces langages ne permettent que de représenter certains modèles, en général ceux de KADS (*cf.* § 3.3) et des méthodes à limitation de rôles (*cf.* § 3.2). De plus, un certain nombre de propriétés sont perdues au moment du passage à l'implémentation définitive.

Formaliser le modèle conceptuel. Des langages comme KARL (Fensel *et al.*, 1991), FORKADS (Wetter, 1990), (ML)² (Harmelen & Balder, 1992) proposent une description formelle du modèle conceptuel pour faire des vérifications (formelles) de complétude ou de cohérences. Ils sont dédiés à des méthodes manipulant les types de connaissances définis dans CommonKADS, ce qui représente leur principale limite.

Coder directement le modèle opérationnel. Au regard des limites des précédentes approches, en particulier le fait que les propriétés et services proposés disparaissent au moment de l'opérationnalisation, des recherches ont abouti au développement de langages qui codent directement le modèle conceptuel dans un langage opérationnel comme LISA (Delouis, 1993; Jacob *et al.*, 2000), KARL (Fensel *et al.*, 1991), DEF-* (Kassel *et al.*, 2000). Une telle approche permet de tracer en permanence le raisonnement en termes de buts et méthodes utilisés. Cela permet aussi de mettre en œuvre une réflexivité permettant la mise en place d'un contrôle défini en termes de buts, méthodes (LISA) ou la mise en place de primitives de représentation des connaissances manipulées par le langage (DEF-*).

Les contraintes de ces langages sont évidemment qu'ils ne permettent la représentation que de certains modèles. Leur défaut, au regard des ambitions affichées est que ce ne sont pas toujours des langages réellement opérationnels en termes de développement, gestion des erreurs ou vitesse d'exécution.

Le langage K développé en sur-couche de LISP par J. Bouaud (1989) permet de coder directement un modèle opérationnel de haut niveau qui, même s'il ne code pas explicitement un modèle conceptuel, permet de représenter des structures de programme de haut niveau facilement reliables aux différentes composantes d'un modèle conceptuel.

⁸« Description » est sciemment utilisé à la place de « représentation » dans la mesure où les langages de représentation des connaissances habituellement appelés ainsi couvrent généralement exclusivement l'opérationnalisation du modèle conceptuel dans la description du processus d'acquisition des connaissances (*cf.* 3.1) alors que nous voulons parler des différents langages cherchant à couvrir les différentes étapes au sein de l'ensemble du processus.

Il a permis le développement du système AMD (Cavazza & Zweigenbaum, 1994) dans le domaine de la compréhension des comptes rendus médicaux du cancer de la thyroïde. La question de la validation d'un tel système a été spécifiquement étudiée par le travail de thèse de C. Haouche-Gingins (1996) que nous résumons dans le paragraphe suivant.

3.5.1 De l'opérationnalisation à la validation

Ce travail se situe résolument dans le contexte de l'opérationnalisation et de la validation d'un modèle conceptuel et porte, comme annoncé précédemment, sur la validation d'un système écrit dans un langage qui permet un codage directement opérationnel d'un modèle conceptuel.

La question de la validation d'un SBC s'est rapidement posée d'une façon différente de celle du génie logiciel en raison du type d'informations et de leur niveau de représentation traités par ces systèmes. Puisqu'il s'agissait des connaissances, représentées au sein d'un modèle conceptuel (cf. 2.1), il semblait intéressant d'utiliser ce même modèle conceptuel dans le processus de validation. Le travail réalisé consistait donc à développer une méthode de test pour le système AMD par rapport à son *comportement*. Les caractéristiques et hypothèses principales de cette méthode étaient alors les suivantes :

Le modèle conceptuel comme spécification. Utilisation du modèle conceptuel – c'est-à-dire comprenant une description de haut niveau des connaissances du domaine et de la méthode de résolution de problème – du SBC (AMD) comme spécification pour identifier l'origine de dysfonctionnements du système, de mauvais résultats mais aussi de mauvaises spécifications.

Le modèle conceptuel comme description valide. Une première hypothèse qu'il a fallu poser dans ce travail est que la méthode de résolution de problème décrite dans le modèle conceptuel spécifie le(s) comportement(s) valide(s) du SBC. Cette hypothèse est d'ailleurs implicite dans un projet comme KADS (cf. 3.3) où la structure d'inférence est censée décrire les comportements valides potentiels du SBC.

Le principe de correspondance structurelle. Reinders *et al.* ont introduit le *principe de correspondance structurelle* (1991) pour souligner la nécessité d'adopter une architecture de système reflétant le modèle conceptuel. Krivine et David (1991) ont souligné les avantages que l'on pouvait retirer de l'adoption d'un tel principe pour l'évolution et la maintenance des SBC. C'est ce qui a été posé au niveau de la validation de AMD en plaçant dans le code – une description de *bas niveau* – des liens vers les inférences du modèle conceptuel – une description de *haut niveau* – auquel chaque règle implémentée correspondait.

Ces caractéristiques étant posées, le principe général (simplifié) de validation est alors d'utiliser la description de la tâche et des méthodes d'inférences du système pour décider si son comportement est *normal* ou pas. Comme l'indique la figure 1.3, des traces de bas niveau sont récupérées lors de l'exécution du système sur des données de test. Ces traces sont constituées de séquences des noms des règles qui se sont déclenchées. Un mécanisme d'abstraction permet ensuite de disposer de traces de haut niveau d'abstraction. Exprimées au même niveau d'abstraction que la spécification de la méthode de résolution de problème, décrite dans le modèle conceptuel, elles lui sont comparables.

Sans reprendre les discussions approfondies de ce travail (Haouche-Gingins & Charlet, 1997, 1998), on peut constater qu'il se place dans un courant qui a cherché et cherche encore à appliquer les principes de développement d'applications du génie logiciel aux SBC⁹..

⁹On pourra lire à ce sujet (Charlet *et al.*, 2001) et (Cauvet *et al.*, 2001).

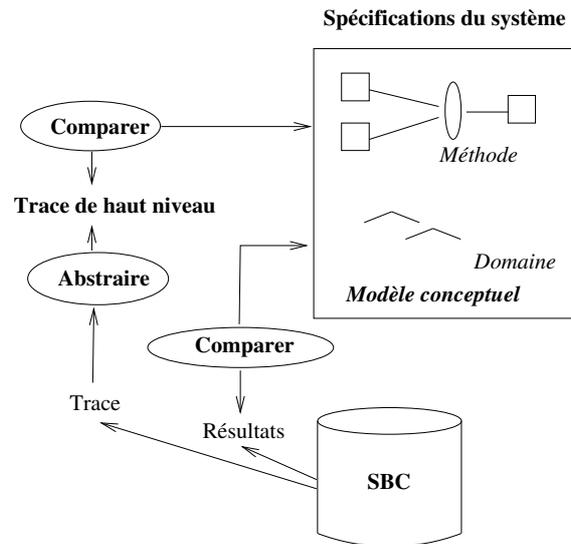


FIG. 1.3 – Principe général de la méthode de validation d'AMD.

4 Réflexions et questions

Ce chapitre nous a permis de partir de l'acquisition des connaissances et de balayer le contexte de réflexion de l'Ingénierie des connaissances. Nous voulons terminer par les questions ou réflexions laissées en suspens vis-à-vis de l'Ingénierie des connaissances et des applications médicales – qu'elles soient encore des problématiques de recherche ou que l'évolution du milieu médical ait modifié le contexte de mise en œuvre des SBC – et auxquelles nous nous attacherons à répondre ultérieurement.

4.1 Nature et rôles des systèmes développés

On peut constater, à l'usage, que les systèmes développés dans le milieu médical ne sont majoritairement plus les SBC de diagnostics sur lesquels nous avons travaillé jusqu'à récemment¹⁰. Deux raisons principales nous semblent pouvoir expliquer ce constat.

Premièrement, pour performants qu'ils soient, les systèmes ne convenaient que très rarement à ce qu'en attendaient les médecins en routine. Il n'attendaient pas un système qui fasse un diagnostic à leur place mais un système qui les aide dans cette tâche, que ce soit en leur fournissant les bonnes informations au bon moment sur un malade ou, plus récemment, en leur donnant les moyens de vérifier qu'ils respectent bien les procédures ou bonnes pratiques à mettre en œuvre¹¹. Les réponses à ces attentes peuvent être trouvées dans de nouvelles manières de mettre l'information médicale à disposition des praticiens via le document textuel comme mode d'échange privilégié de l'information médicale (cf. chap. 3). Par ailleurs, les systèmes construits devant tenir compte de l'interaction langagière avec l'être humain, la question des thésaurus, de leur maintenance, de leur interaction avec des représentations plus conceptuelles – *i.e.* les ontologies – amène de nouvelles réflexions et permet de proposer de nouveaux services

¹⁰Les dossier « IA & médecine » du bulletin de l'AFIA que nous avons coordonnés en 1993 et 2002 donnent des indications en ce sens (Charlet, 1993b; Corruble & Charlet, 2002) : si, en 1993, la problématique des SBC « classiques » était bien présente, leur usage en routine apparaissait comme problématique. En revanche, en 2002, le point de vue des contributeurs au dossier s'est complètement élargi à des problématiques de communications, d'indexation, etc. , au point que le dossier aurait aussi bien pu s'appeler « STIC & médecine ».

aux médecins (*cf.* chap. 4).

Secondement, l'avènement des réseaux, et du Web en particulier, a pointé le déficit d'interopérabilité et de communication en générale du milieu médical. Les recherches se sont alors concentrées sur la façon de combler ce déficit par des travaux sur la normalisation (*cf.* chap. 5) ou sur le travail coopératif (*cf.* chap. 6). Le « Web sémantique », s'il n'est pas encore une réalité, encore moins en médecine, devient un paradigme de réflexion en parfait accord avec le point précédent, c'est-à-dire une approche documentaire de l'information médicale (*cf.* chap. 6).

4.2 Les ontologies

Les ontologies ont été rapidement abordées dans ce chapitre comme noyau des connaissances du domaine, mais avec un faible rôle dans le processus d'acquisition des connaissances. Cela traduit une réalité dans la mesure où les démarches d'Ingénierie des connaissances faisaient, jusqu'au milieu des années 90, l'hypothèse que s'il y avait des objets existant dans le domaine, ils seraient faciles à repérer et que le raisonnement de façon générale et plus particulièrement la tâche du SBC devait être le cœur des démarches d'acquisition (*cf. infra*, en particulier les approches génériques). C'est donc ce relatif intérêt que nous avons voulu pointer ici mais pour affirmer maintenant qu'avec les nouveaux systèmes qui apparaissent (*cf.* § 4.1), la question des ontologies devient centrale en médecine et, qu'à l'usage, leur définition et construction se révèlent non triviales.

Des question épistémologiques – *e.g.* sur la nature et la réutilisabilité des ontologies – se posent et ont alimenté de longs débats (Charlet *et al.*, 1996a; Guarino, 1997; van Heijst *et al.*, 1997). La communauté française est particulièrement active dans ce domaine, en particulier au niveau du groupe TIA (« Terminologie et IA¹² »), au sein duquel ontologies, terminologies, bases de connaissances terminologiques, entre autres « produits » de la recherche, sont caractérisées par rapport aux besoins de la Terminologie et de l'Ingénierie des connaissances. Il en ressort que la réutilisation simple – voire simpliste – des ontologies espérée au début des années 90 a disparu et que des réflexions et outils nouveaux sont apparus nécessaires pour aborder la question de la construction des ontologies et de leur réutilisation (*cf.* chap. 4).

4.3 Les langages de représentation

Liés à la question de nouveaux services et systèmes qui voient le jour, les langages de description et de représentation des connaissances évoluent aussi dans le domaine médical. Ainsi, les connaissances du domaine, en particulier les ontologies, sont représentées en utilisant le formalisme des Graphes conceptuels de Sowa (1984) ou les logiques de description. L'un comme l'autre permettent de représenter des ontologies, des propriétés et de faire des inférences dessus, en particulier de la classification (Nobécourt, 2000). Depuis l'avènement du Web et des travaux autour du Web sémantique, la représentation des ontologies est en passe d'être normalisée dans un langage, OWL (*Ontologies Web Language*), fondé sur des sur-couches de XML, et les logiques de descriptions ont pris le pas sur les graphes conceptuels (*cf.* chap. 4, § 7).

Par rapport à la problématique de la représentation des connaissances au sein de documents textuels, XML devient le standard de fait de l'ingénierie documentaire : lan-

¹¹Cela ne veut pas dire que les systèmes d'aide à la décision n'existent plus. Il ont parfois pris une autre forme comme indiqué ici ou se sont cantonnés à quelques endroits précis dans le suivi de maladies chroniques ou vers des services d'aide à la décision dans des tâches plus proches de la gestion (*cf.* chap. 6). Enfin, notre propos n'est pas de fustiger une approche qui a apporté et apporte encore des avancées mais plutôt de proposer à la réflexion un nouveau contexte de développement des SBC qui est le nôtre et, semble-t-il, celui de bien d'autres (*cf.* note 10).

¹²<http://www.biomath.jussieu.fr/TIA/>

gage de balisage de texte, il se place là en héritier direct de SGML. De plus, il fournit avec des sur-couches comme RDF la façon de décrire des métadonnées permettant l'indexation et l'accès à l'information médicale (cf. chap. 3).

Enfin, dans le contexte de l'interopérabilité, XML devient aussi l'espéranto de l'échange de données informatisées (EDI) : les balises y sont utilisées pour repérer de l'information dans des messages structurés ou même pour décrire des protocoles d'échange et les comportements des programmes qui pilotent ces échanges (format d'enveloppe fondé sur ebXML, SOAP – cf. chap. 5).

4.4 L'organisation et la gestion

Les organisations, les entreprises, découvrent, avec le concept de gestion des connaissances, que les savoirs et les savoir-faire sont cruciaux dans leur développement. Ces connaissances se transmettant entre les individus, la question de l'apprentissage devient centrale. Au niveau de la santé, nous sommes dans une situation où les connaissances et les savoir-faire se transmettent assez bien entre les médecins dans les services, dans les phases de formation et d'apprentissage. En revanche, l'usage de ces savoirs, portés par le corps médical au sens large ne sert pas l'institution, c'est-à-dire ne permet pas une meilleure organisation, une meilleure gestion de l'hôpital. Le PMSI (programme de médicalisation du système d'information), censé aller dans cette direction a de nombreuses difficultés à imposer une démarche gestionnaire, en raison de sa nature d'indicateur plus économique que médical. L'Ingénierie des connaissances a des champs de proposition à explorer dans ce sens, pour permettre une meilleure transmission des savoirs, en particuliers entre les différents métiers impliqués dans la santé (hôpital, réseau de santé, ...). Cela passe par une modélisation globale de l'information et des connaissances partagées au sein d'un établissement de soin et par l'appréhension de l'action et de la cognition collective (cf. chap. 6, § 3.1).

4.5 Pour terminer

Ces questions ne pourront pas être traitées dans leur ensemble si nous ne nous interrogeons pas sur la nature de l'Ingénierie des connaissances, à savoir son objet d'études et ses méthodes pour étudier quand et comment elle pourra intervenir, en collaboration avec d'autres disciplines, pour aider à l'élaboration des réponses à fournir (cf. chap. 2).

Crédits

L'historique et les définitions des modèles en acquisition et Ingénierie des connaissances exposés ici doivent beaucoup à un travail réalisé en commun avec Chantal Reynaud et Régine Teulier (Charlet et al., 2001). Les cours sur l'Ingénierie des connaissances développés en coopération avec Nathalie Aussenac-Gilles m'ont aussi fourni de la matière première pour l'élaboration de ce chapitre ainsi que la synthèse faite par Chantal Reynaud dans son mémoire d'habilitation à diriger des recherches (Reynaud, 1999). Les travaux décrits au § 2.2.1 sont le fruit d'une collaboration étroite avec Chantal Reynaud et Jean-Paul Krivine. Les travaux sur la validation décrits en § 3.5.1 sont le fruit du travail de thèse de Corinne Haouche-Gingins (Haouche, 1996; Haouche-Gingins & Charlet, 1998) sous ma direction.

L'ingénierie des connaissances

Développé à l'origine pour justifier un certain nombre de travaux personnels et d'autres en ingénierie des connaissances, ce chapitre a été transformé en article pour des besoins d'évaluation (Charlet, 2001). Si il peut apparaître quelque redondance par rapport au reste du mémoire, en particulier au niveau des exemples, il nous a paru intéressant de conserver la structure argumentative de l'article qui donne une autonomie de compréhension au chapitre. Nous l'avons donc légèrement modifié pour citer d'autres parties du mémoire. La section cinq sur la question du domaine scientifique et de l'enseignement de l'Ingénierie des connaissances est conservée à dessein ici mais pourra être évitée par le lecteur non intéressé : elle nous semble cependant boucler l'argumentaire développé ici, à savoir qu'une validation d'une « science de l'artificiel » passe entre autres par le fait qu'elle est enseignée même si, comme nous le verrons, ce n'est pas une raison suffisante.

Nous abordons donc ici quelques considérations épistémologiques et pratiques pour discuter et justifier l'extension du champ d'action de l'acquisition des connaissances – i.e. les systèmes experts – à un champ beaucoup plus large des systèmes à base de connaissances. Cela nous amène à réfléchir sur la notion de connaissances et à redéfinir ce qu'est un système à base de connaissances. Enfin, nous concluons ces réflexions en proposant ce qui pourrait être un cadre de programme pour l'Ingénierie des connaissances.

Les idées défendues ici, pour personnelles qu'elles soient, ont été confrontées aux travaux de certains, aux critiques d'autres, dans le but d'atteindre un certain consensus si ce n'est une certaine « universalité » à leur niveau. Elles n'engagent pourtant que leur auteur et ont vocation à être discutées.

1 Introduction

L'acquisition des connaissances a constitué 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 im-

plique 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, sans se réduire seulement à cela, l'acquisition des connaissances était focalisée sur une modélisation psychologique ou empirique des connaissances d'un expert dans le but de les coder dans un système expert (SE – cf. chap. 1).

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 plus ou moins 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.

Dans tous les cas, il s'agit de définir quelles connaissances sont pertinentes, pour les exploiter en formulant un modèle du problème. Classiquement, l'acquisition des connaissances propose des méthodes permettant de modéliser des raisonnements en faisant abstraction du domaine d'activité et en se concentrant sur la façon dont est conduit ce raisonnement. 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.

Les recherches en acquisition 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. Ainsi, tout système 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'acquisition des connaissances⁴. Les motivations et le cadre de travail décrit ici ont fait évoluer les attendus de l'acquisition des connaissances vers une « Ingénierie des connaissances » où l'apport de méthodes et techniques reproductibles est devenu central.

Ces constats faits, il nous semble indispensable d'aller un peu plus loin dans la caractérisation de l'Ingénierie des connaissances, justement pour mieux comprendre l'évolution du domaine, en termes de types de travaux mis en œuvre et, évidemment, de résultats méthodologiques possibles ou attendus. Pour ce faire, nous pensons qu'il est nécessaire de revenir sur les fondements des SBC pour tirer tous les bénéfices d'un

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).

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) et renverrons le lecteur au chapitre 1.

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).

travail de réflexion épistémologique, historique et pragmatique.

La 2^e section de ce chapitre 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'exemple, la portée des propositions faites. Enfin dans la 5^e section, nous essaierons 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.

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'appropriier le comportement du système ; les sciences de gestion pour concevoir et replacer les systèmes dans leur environnement organisationnel, etc. Ces réflexions, développées au sein de disciplines, 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 système d'information (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 développés 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) :

1. La connaissance n'est possible que dans un environnement *technique* qui est le nôtre, 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

⁴« Candidat » est utilisé ici littéralement. Nous pensons que c'est une évolution inéluctable de nombreux systèmes, en particulier systèmes d'information, 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.

qu'elle autorise est alors la technique qui permet la mémorisation et la genèse de la connaissance⁵. Il n'est pas de notre propos ici de nier toute connaissance interne à l'utilisateur ni d'engager un débat (*cf.* note 5) mais de situer notre réflexion dans un environnement technique puisque c'est celui de notre intervention et de toute intervention d'un chercheur cherchant à développer ou installer un SBC quel qu'il soit.

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, c'est-à-dire qu'ils lui donnent un sens et tendent même, au regard de l'usage qu'ils ont de la machine, à lui conférer une certaine rationalité.
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 cela⁶.

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'information⁷, 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.
- 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.

⁵Dans 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, par les mises en correspondance qu'elle autorise, peut être vue comme une technique qui crée du sens, y compris sans intentionnalité.

⁶Cette troisième caractéristique est une conséquence et ré-affirmation de la première : par les mises en correspondances temporelles qu'elle permet, l'informatique est une technique du calcul qui fait émerger du sens. 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).

⁷Ici, nous nous intéressons en particulier à l'information telle que l'a théorisée l'informatique.

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 (cf. 2.1.1 n° 2) 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 est d'abord et potentiellement, l'extériorisation de la connaissance par l'utilisateur d'un outil, ici une réécriture par l'informatique⁸ (cf. 2.1.1, n° 3).
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 modélisateur⁹. Étudier cette construction permet d'essayer d'en dégager des invariants méthodologiques. C'est le projet des épistémologies constructivistes (Le Moigne, 1995 – cf. 5.1).

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 admises, il est alors possible d'étendre le champ de nos investigations 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 (cf. 2.1.1 n° 3).

Pour concevoir un SBC, il faut donc tenir compte de la manière dont des utilisateurs se l'approprient et lui attribuent du sens en interprétant, sur la foi des primitives empruntées à la terminologie du domaine, les représentations comme des expressions linguistiques de connaissances. Cette interprétation repose sur le contexte d'usage du

⁸Cette réécriture ne préjuge pas, quand on quitte la sphère informatique, d'une action réelle effectuée par l'agent qui interprète le comportement du système informatique et les réécritures qu'il effectue.

⁹Avec cette expression, nous ne voulons pas dire qu'il n'y a pas de connaissance avant le projet mais qu'elle n'est pas instanciée, ne sert que de réservoir, avant un but et une action qui lui donneront toute sa dimension et son utilité.

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 14).

SBC et son intégration dans un système de pratiques où il prend sens et justification (cf. 2.1.1 n° 2). 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), raisonnement à partir de cas, 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 Programme de travail

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, (cf. 2.1.1 n° 3 et (Charlet & Bachimont, 1998)).

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 (cf. section 2)¹³.

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 : 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.

¹³Ce programme de travail est en droite filiation de celui proposé à l'Intelligence artificielle par B. Bachimont (1996), *l'Artefacture*. Il essaie d'avancer et de développer plus précisément un programme de travail pour la « partie » de l'Intelligence artificielle se préoccupant le plus de connaissances, l'Ingénierie des connaissances. Par ailleurs et puisque l'Ingénierie des connaissances a toujours cherché à développer des méthodologies pour résoudre de façon autrement que *ad hoc* les problèmes qui se posaient à elle, nous cherchons nous

3.1 En guise de programme

Ainsi, une démarche d'Ingénierie 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 (cf. 2.1.1 n° 2). 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 cogniticiens, 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 ». On utilise pour cela l'une des principales caractéristiques des systèmes formels : les symboles primitifs sur lesquels ils reposent, sont des unités significatives emprunté à la langue naturelle (Bachimont, 2000b)¹⁴.
- Le but correspond alors à l'intégration de ces réflexions dans des artefacts informatiques à modéliser, construire, faire évoluer (cf. 2.1.1 n° 1). 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 (cf. 2.1.1 n° 3). 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 méthodologiques cruciaux :

- La connaissance étant contextuelle, se donner les moyens de représenter ce contexte dans une démarche de représentation des connaissances (cf. 2.1.2 n° 2). Cela peut amener à des choix de formalisation minimum comme dans les projets médicaux HOSPITEXTE et ONCODOC (cf. § 4.1 et chap. 3, § 3).
- 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 (cf. 2.1.2 n° 1).
- La connaissance, contextuelle, ne pouvant être indépendante de la tâche à réaliser, se donner les moyens d'assumer explicitement cette dépendance (réutilisation limitée) (cf. 2.1.2 n° 3). Cela peut se traduire, dans le cadre des ontologies, par des choix méthodologiques forts figeant les contexte d'utilisation (cf. chap. 4, § 2.2.2).

aussi, à proposer une méthodologie.

¹⁴ 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. 19). 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. chap. 4, § 2.2.2).

3.3 En guise de champ d'action

Le champ d'action de l'Ingénierie des connaissances conclut ce chapitre et reprend pour partie les axes défendus lors de la conférence IC 2001.

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 possible¹⁵, 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 ;
- d'autre part, en se rapprochant de disciplines abordant les mêmes objets de recherche ou des objets proches et produisant des connaissances complémentaires aux siennes comme la sociologie, les sciences de gestion, l'ergonomie ou l'ingénierie éducative qui peuvent concourir à une démarche d'explicitation de ce que sont les connaissances dans un contexte humain et organisationnel.

4 Exemples

Nous voudrions conclure ce « beau » programme par quelques exemples de recherches qui illustrent justement le programme et la méthode de l'Ingénierie des connaissances. Ces exemples sont d'abord et évidemment pris parmi les projets auxquels nous avons collaboré puisque ce sont ceux que nous connaissons le mieux. Ce sont aussi des méthodes développées par l'un ou l'autre et utilisées ou mise en œuvre par suffisamment de personne pour qu'on puisse y voir une validation pragmatique. Enfin, il est difficile de décider *a posteriori* de la façon dont les gens ont pu attaquer tel ou tel problème et quelles réflexions ont fondé leur démarche. Si nous pouvons avoir un avis dessus, surtout pour les projets dans lesquels nous étions impliqués, il est plus facile de rechercher dans ces exemples des illustrations sur la façon d'appréhender et de solutionner tel ou tel problème – *e.g.* la représentation de connaissances médicales – et de laisser les intéressés (*cf.* § 6) s'exprimer sur leur propre démarche et la méthodologie retenue.

4.1 HOSPITEXTE

Développé par B. Bachimont, V. Brunie et nous-même, le projet HOSPITEXTE (*cf.* chap. 3, § 2 – Brunie *et al.*, 2000) a consisté à réaliser un plate-forme d'expérimentation d'informatisation du dossier patient en suivant 2 objectifs : (1) création d'un dossier patient virtuel en utilisant les capacités du Web à permettre la « reconstruction » d'un unique

¹⁵On a vu (*cf.* chap. 1, § 3.4), que cette réutilisabilité est plus à chercher du côté des méthodes que de celui des modèles.

dossier patient à partir d'informations résidant dans d'autres services ou à l'extérieur d'un hôpital et (2) 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.

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. Par exemple, un médecin utilisant un dossier médical à la HOSPITEXTE pourrait avoir des facilités d'annotation et de regroupements d'informations facilitant la création de nouvelles connaissances comme l'écriture d'un compte rendu de sortie.

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, technique informatique, est alors utilisée au regard de ses capacités à produire du sens dans un système d'usages.

Bien que le projet ait pu fournir un certain nombre d'avancées, des recherches doivent être poursuivies sur le sens des différents documents générés en fonction de l'ensemble des documents disponibles et de leur place dans l'hypertexte étudié, vers ce qu'on pourrait appeler une *la sémantique hypertextuelle*. Par ailleurs et en complément logique, il est nécessaire d'explicitier et de travailler la fonction *d'annotations* c'est-à-dire d'envisager comment donner à une personne les moyens de construire sa propre lecture des documents qui lui sont proposés et de générer sa propre vision du dossier. Ce que l'on peut théoriser dans une *sémantique hypertextuelle annotationnelle* (Charlet *et al.*, 1999). Comme précédemment, ces 2 problématiques ressortissent aussi à la façon dont la technique permet de créer du sens au regard des usages, ici, professionnels, collectifs comme idiosyncrasiques (*cf.* chap. 3, § 2.4.1).

4.2 Des thésaurus à la construction d'ontologies

La construction d'ontologies est un des champs de recherche de l'Ingénierie des connaissances. 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 était rapidement reconnue (Bachimont, 1996), elle était peu ou pas outillée (*cf.* chap. 1, § 5.2). Cet outillage est venu 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 (Bourigault, 2000 – *cf.* chap. 4, § 4)¹⁶. Si un certain nombre de travaux montrent tout l'intérêt de ces réflexions (Bourigault & Charlet, 1999), on peut en retenir deux 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

¹⁶Ces réflexions se sont en particulier développées au sein du groupe TIA – « terminologie et Intelligence Artificielle » – et s'y poursuivent encore (*cf.* note 12, page 13).

de construire un produit terminologique intermédiaire – *i.e. ontologie régionale, conceptuelle* et finalement *computationnelle* –, correspondant à un changement de paradigme. Ainsi, la construction d'ontologie conceptuelle pour un SBC ou de référentiel « métier » pour un SI passe par des changements de paradigme, 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 et *cf.* chap. 4).

La dépendance de la tâche La construction d'ontologies « réutilisables » a 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 (*cf.* chap. 4, § 4) 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 constructiviste.

4.3 Du contexte partagé

Le travail présenté par M. Zouinar *et al.* sur le contexte partagé (2000), présente un cadre conceptuel et méthodologique pour analyser les effets de la conception des environnements de travail coopératif sur le *contexte partagé*, vu comme un ensemble d'événements contextuels mutuellement accessibles à plusieurs agents. Cette recherche est appliquée au contrôle du trafic aérien et aux collaborations entre contrôleurs.

Dans ce travail, le modèle informatique développé se veut un modèle de simulation et un *outil heuristique qui offre des pistes de réflexion* pour l'analyse des situations de travail. Les 2 idées fortes d'un tel travail (toujours au regard des réflexions développées ici) sont :

La prise en compte du contexte Ici, le contexte de travail est visuel et auditif avec ou sans acte de communication volontaire et il est formellement représenté. Remarquons que pour ce type de travaux ressortissant à ce qu'on appelle la *cognition située*, la prise en compte du contexte est fondamentale, constitutive de la discipline ;

L'utilisation d'un modèle informatique au regard des appropriations qu'il permet

Le programme informatique utilisé est un programme de simulation qui permet de générer le contexte partagé pour un scénario donné dans un environnement de travail donné. Un peu différemment de précédemment (*cf.* § 4.1) – disons, avec avec un niveau d'indirection supplémentaire –, le modèle formel proposé l'est ici en fonction de ses capacités à (re)créer des scénarios pour tester d'autres programmes – *e.g.* des outils de CSCW (*computer supported cooperative work*) – qui seront évalués à leur propre capacité à permettre une bonne appropriation de la situation par un agent humain.

5 Un domaine scientifique, un enseignement ?

Arrivé à ce point de l'exposé et des champs d'action de l'Ingénierie des connaissances, il nous reste à nous 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 jusqu'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 autres¹⁷ et proposer des caractérisations.

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 connaissances¹⁸ –, 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 envisagées sans jamais les finaliser¹⁹.

5.1 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 « variables » – 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 connaissances²⁰.

Les deux 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* (Op. Cit., 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 classique²¹ mais l'incapacité de telles hypothèses à

¹⁷De façon « différentielle ».

¹⁸La 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, plus de 400 abonnés.

¹⁹De 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.

²⁰Pour argumenter son propos, J.-L. Le Moigne mobilise évidemment de très nombreux auteurs que nous ne reciterons pas nous-mêmes. Nous renvoyons directement aux références de l'auteur (Le Moigne, 1995). Par ailleurs, on trouve chez Nonaka et Takeuchi (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.

²¹Nous 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

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, les sciences de 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*²². 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* (Op. Cit., p. 67). Corollaire, il est impossible de séparer la connaissance du processus (intelligence) qui la produit. Deux conséquences de cette hypothèse, parmi d'autres, sont 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* (Op. Cit., 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* (Op. Cit., 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*²³. 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 les 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 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) et ainsi argumenter que l'ingénierie des connaissances est une science. Ces caractérisations seront au nombre de trois et porteront sur (1) les hypothèses, (2) la méthode et enfin (3) 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

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.

²²La 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.

²³La téléologie est l'étude des finalités.

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 programme d'ordinateur (*cf.* § 4.1 et § 4.3) – 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 (*cf.* § 4.3), 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. (*cf.* § 3.3) – 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 discutés 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 attelons²⁴.

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

²⁴Nous avons récemment commencé à proposer des supports de cours à disposition sur le Web <<http://www.irit.fr/GRACQ/cours.html>>.

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 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, 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* (Le Moigne, 1995, p. 96)²⁵.

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 finalement 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.

Maintenant, dans une perspective constructiviste, ce programme, ces méthodes sont liés à nos facultés de jugement de sujet connaissant. Bien qu'elles nous semblent partagées par quelques autres et, nous espérons, par ceux sur le travail desquels nous nous sommes appuyés, elles demandent à être discutés, « dialogués », affinés et (re)mise à l'épreuve de l'expérience pour aboutir à prouver, plutôt argumenter, le statut épistémologique de l'Ingénierie des connaissances comme réelle science située, d'un point de vue institutionnel, au sein des STIC²⁶.

Crédits

Le travail développé ici doit beaucoup à Bruno Bachimont pour la partie épistémologique et à et Régine Teulier qui situe notre débat au niveau organisationnel.

²⁵Le 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.

²⁶L'argumentaire développé ici est sciemment limité à l'Ingénierie des connaissances alors que les « sciences de l'artificiel » englobent évidemment l'Intelligence artificielle. Mais comme indiqué précédemment (cf. note 13, p.20), nous essayons de limiter la portée de nos réflexions à ce que nous connaissons le mieux, l'Ingénierie des connaissances.

Des réflexions aux applications “textuelles”

À la suite du chapitre précédent, nous essayons ici de comprendre la genèse de quelques problèmes et comment les réflexions développées précédemment permettent d'avancer dans un certain nombre de développements et d'applications. Ainsi, le caractère contextuel de l'information et ses rapports avec son support (ici informatique) nous amène à mettre sur l'accent sur l'intérêt du document textuel comme support et véhicule de la connaissance. Le projet de dossier médical informatisé HOSPITEXTE (comme d'autres développés au sein du service, e.g. ONCODOC) est présenté dans ce paradigme et la problématique de l'indexation des documents est abordée en décrivant le projet CISMED dans le contexte du Web sémantique.

1 L'information médicale et son contexte ou la primauté du texte

1.1 La spécificité textuelle de l'information médicale

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. Ainsi, ces expérimentations considèrent le dossier médical et donc l'information médicale comme un ensemble de faits bien déterminés qui, pour des raisons de culture et de traditions médicales, sont exprimés linguistiquement. Ici, le fait de s'exprimer en langue et de suivre des formats textuels documentaires n'ajoute rien au contenu exprimé. Une informatisation du dossier passe alors par (et seulement par) la définition *a priori* des faits médicaux pertinents à retenir dans le cadre d'un schéma conceptuel des données. Cette approche, si elle a pu donner d'intéressants résultats dans le cadre de recherches épidémiologiques ou de suivi des patients souffrant de pathologies chroniques comme l'hypertension ou le diabète ne permet pas de « travailler » avec le dossier.

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

sus-dit dossier. Il est alors illusoire de vouloir répertorier les faits médicaux pertinents indépendamment du formatage linguistique et documentaire (Bachimont, 2001a). En effet :

- La mise en forme linguistique confère à l’information médicale la contextualité que possède la langue naturelle. Et ce n’est que parce que les informations lues dans un compte rendu de sortie (p. ex.) sont immergées dans le texte qu’elles sont réellement compréhensibles et réinterprétables par le praticien. Par exemple, quel que soit le mode de codage proposé et aussi complexe soit-il (thésaurus hiérarchique de spécialité, d’actes, de diagnostics, etc.), il y a toujours une situation dans laquelle le praticien considérera le codage proposé comme insuffisant à décrire la situation clinique du patient. La solution choisie la plupart du temps consistera dans la rédaction d’une note, d’un *addendum*, sur support papier ou numérique¹.
- La mise en forme documentaire, sur le fond – organisation logique, sectionnement... – comme sur la forme – titres, mise en page, taille des polices... – est un reflet des contraintes pragmatiques s’exerçant sur le praticien lors de la création et de la consultation des documents médicaux conçus pour être lus dans des circonstances particulières et pour narrer des événements particuliers : lorsqu’un praticien lit un tel document, sa lecture est conditionnée par le fait que le genre textuel – e.g. un compte rendu d’examen – renvoie à un contexte particulier – e.g. l’examen, ses motivations, etc. (Berg, 1998). Par exemple, la phrase « suspicion de dosage de phénobarbital » n’a pas la même signification dans un rapport d’examen clinique – elle y exprime une hypothèse – que dans la conclusion d’un dosage sanguin où elle exprime une affirmation.

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’examen 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*.

1.2 XML et la représentation des connaissances médicales

Sans revenir sur la syntaxe de XML qui commence à être bien connue, il suffit de noter que c’est un langage de balisage qui satisfait un certain nombre des critères exprimés précédemment :

- XML permet de distinguer, pour un document, d’une part, sa structure et, d’autre part, son allure. La structure renvoie au fait que l’organisation du document n’est pas quelconque et qu’il faut par exemple distinguer un titre, un résumé et des sections. L’allure renvoie au fait que la structure du document ainsi que son contenu doivent être visualisés : par exemple, les titres sont en gras et centrés et les sections en maigre et justifiées.
- De plus, alors que les *balises* affectées aux éléments du texte par les logiciels de traitement de texte définissent des procédures (par exemple « mot à présenter en

¹Il n’est pas de notre propos ici de rejeter tout système de codage dont nous connaissons la nécessité en médecine, que ce soit à finalité épidémiologiques ou gestionnaires; mais plutôt d’insister sur le fait que tout système de codage est une projection et donc une *réduction* d’une information médicale selon un certain point de vue. Ce codage effectué, les autres dimensions de l’information, les autres points de vues sont irrémédiablement perdus.

police Times, 12 points, gras, aligné à droite ». . .), les balises en XML sont des *balises descriptives de l'information* elle-même et permettent d'envisager des usages liés à la nature de cette information : identifier le mot comme renvoyant au concept de « diagnostics », identifier la liste des adresses URL auxquelles les liens hypertextes de ce mot renvoient (pour p. ex. pointer sur des guides de bonne pratique au sujet de ce diagnostic ou faire des substitutions sémantiques de mots ayant la même signification), afficher la date de modification de chacun des éléments du texte, etc.

- XML permet de définir la nature et la structure de *balises* qui peuvent être insérées dans un document. Ces balises délimitent des *éléments* du document, donc permettent de catégoriser des segments du document tels que le titre de la section, le paragraphe ou, dans une optique nous intéressant davantage, le nom d'une personne (patient, médecin) ou un *diagnostic*. Ce balisage ne modifie pas le document et ne présage en rien de son usage². Ceci dit, pour ce qui nous intéresse, le fait de repérer un diagnostic permet d'envisager des usages tels que proposer des codes s'y référant ou des liens vers des protocoles de soins, etc. L'ensemble des *balises* autorisées pour une classe de documents est spécifié dans une *Définition de type de document* ou DTD.
- Enfin, le document représenté en XML peut contenir, en sus du texte lui-même, des métadonnées qui renseignent sur les conditions de création du document, son contexte d'élaboration, etc., toute information que l'on peut juger nécessaire à la compréhension du contexte dans lequel le document s'insère. La figure 3.1 montre un exemple de compte rendu de radiologie balisé en XML.

2 Le projet HOSPITEXTE

Sur ces considérations, nous avons construit le projet HOSPITEXTE. Ce projet a été élaboré à la suite du projet DOME , financé par l'union européenne (MLAP #63-221 – Séroussi *et al.*, 1996). Le principal résultat de DOME a été d'élucider l'intérêt, pour des médecins hospitaliers, d'avoir à leur disposition un dossier patient électronique multimédia avec des fonctionnalités de navigation hypertextuelle. Ces spécifications avaient été accompagnées de la mise en œuvre d'une maquette utilisant les technologies du Web (Bouaud & Séroussi, 1996). Dans le fil de ce travail, le projet HOSPITEXTE se proposait alors d'instrumenter un prototype fonctionnel.

2.1 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

²XML est, comme SGML, un métalangage permettant de décrire le nom de *balises* autorisées et leur hiérarchie d'apparition, leur grammaire. De ce point de vue, HTML, dont les balises sont prédéterminées est un langage qui peut d'ailleurs être indifféremment décrit en SGML ou XML.

```

<?XML version="1.0" ?>
<?XML:stylesheet type="text/XSL" href="cr-radio.xsl" ?>

<CR-RADIOLOGIE>
  <ENTETE>
    <INFORMATION-SERVICE>
      <HOPITAL>Groupe hospitalier Léonard Devintscie</HOPITAL>
      <SERVICE>Radiologie Centrale</SERVICE>
      <MEDECIN>Dr. Bouaud</MEDECIN>
      <TITRE-EXAMEN>Phlébographie des membres inférieurs</TITRE-EXAMEN>
    </INFORMATION-SERVICE>
    <INFORMATION-DEMANDE>
      <SERVICE>Sce Pr. Charlet</SERVICE>
      <MEDECIN>Dr. Brunie</MEDECIN>
      <DATE>29-10-99</DATE>
    </INFORMATION-DEMANDE>
    <INFORMATION-PATIENT ID="236784020">
      <NOM>Donald</NOM>
      <PRENOM>Duck</PRENOM>
    </INFORMATION-PATIENT>
  </ENTETE>
  <BODY>
    <INDICATION>
      Suspicion de phlébite de jambe gauche
    </INDICATION>
    <TECHNIQUE>
      Ponction bilatérale d'une veine du dos du pied et injection
      de 180cc de produit de contraste
    </TECHNIQUE>
    <RESULTATS>
      1 image lacunaire endoluminale visible au niveau
      des veines péronières gauche. Absence d'opacification
      des veines tibiales antérieures et postérieures gauches.
      Les veines illiaques et la veine cave inférieure sont libres.
    </RESULTATS>
    <CONCLUSION>
      <ETATPATH code="I80">Trombophlébite</ETATPATH> péronière
      et probablement tibiale antérieure et postérieure gauche.
    </CONCLUSION>
  </BODY>
</CR-RADIOLOGIE>

```

FIG. 3.1 – Un compte rendu de radiologie en XML

(Tange, 1995). Par conséquent, si l'information médicale est d'une certaine manière neutre quant à son support informatique ou papier, il apparaît clairement 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 annotatives³, 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.

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).

³Comme on le voit dans le dossier papier, où de nombreux documents sont émaillés de notes et surlignages divers.

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, 2001a). 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'examens, 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 structurés et respectent donc chacun une DTD (cf. § 1.2). En résumé, l'utilisateur (*i*) se voit servir les documents originaux, (*ii*) navigue dans l'hypertexte grâce à des documents de navigation et (*iii*) construit son parcours (*i.e.* écrit sa lecture) grâce à des fonctionnalités d'annotation.

2.2 Des objectifs

Dans la pratique, le projet HOSPITEXTE consiste alors à réaliser un plate-forme expérimentale d'informatisation du dossier patient en suivant 2 objectifs :

Création d'un dossier patient virtuel. Cet objectif est lié au problème de la « reconstruction » d'un unique dossier patient à partir d'informations résidant dans d'autres services ou à l'extérieur d'un hôpital.

Élaboration d'une station de travail professionnelle. La surcharge informationnelle et la désorientation qui résulte de la mise sous forme informatique du dossier patient impose une réflexion et la mise en place d'outils permettant une lecture professionnelle du dossier. Par ailleurs, la nécessité de travailler l'information et d'élaborer de nouveaux documents implique de donner des possibilités de créations de synthèses et d'annotations sous différentes formes. Ces deux contraintes doivent être appréhendées simultanément et sont réunies dans le concept de station professionnelle de lecture/écriture mise en œuvre dans une approche d'Ingénierie des connaissances.

La solution au premier problème passe par une organisation de type Intranet où chaque unité médicale est serveur et client du système d'information; la contractualisation de ces échanges d'information est un aspect essentiel du projet que seule l'ingénierie documentaire permet d'appréhender au niveau de l'information médicale. Ce problème, qui se complique de questions de sécurité est développé chapitre 5, paragraphe 5.1.

2.3 Un prototype

Un prototype du système HOSPITEXTE, implémentant ces principes, a été développé au SIM (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'examens, etc) sous format électronique (Word en général) à un dossier patient informatisé *navigable* sur le Web.

La figure 3.2 donne l'architecture du prototype. Les documents sont produits par les outils usuels de l'hôpital, essentiellement le traitement de texte Word, et des automates. Ces documents ont des formats « propriétaires » qu'il faut traduire en SGML. Cette étape est la « traduction ascendante », marquant le fait que l'on passe d'un marquage physique de formatage à un marquage logique de structure. Les documents SGML sont alors gérés dans un système de gestion de base de données (SGBD). Les documents de synthèse générés *a posteriori* sont aussi inclus dans le SGBD. Enfin, les documents SGML sont tous traduits en HTML pour être lus sur un *navigateur*. Ce prototype est opérationnel⁴ et aboutit à des dossiers patients générés par une chaîne de « dématérialisation » entièrement automatique⁵.

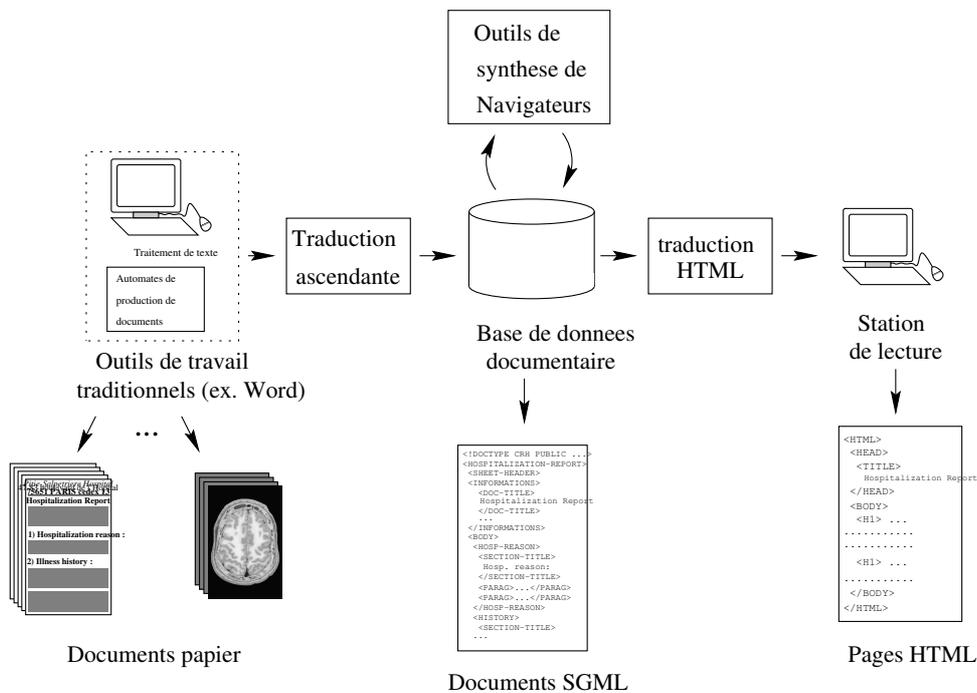


FIG. 3.2 – Architecture du système HOSPITEXTE.

Un des apports principaux – et médical – du prototype a été la définition d'une typologie documentaire de tous les documents constituant un dossier⁶. Cela s'est traduit par la définition d'une cinquantaine de DTD décrivant la structuration de la plupart des documents constituant les dossiers. Ces types sont regroupés dans une DTD générale pour le dossier patient, MEDDOC. Les dossiers médicaux de 10 patients ont ainsi été traités et ils regroupent plus de 2000 documents. La figure 3.3 montre un compte rendu

⁴Les fonctionnalités d'annotation n'ont pas été mises en œuvre dans ce prototype (cf. § 2.4).

⁵Un extrait d'un dossier patient anonymisé est disponible à <http://www.biomath.jussieu.fr/hospibin/patients.tcl>.

⁶Plusieurs dossiers en fait, en pneumologie.

d'hospitalisation avec la table des matières associée – *i.e.* un document de navigation (*cf.* § 2.1).

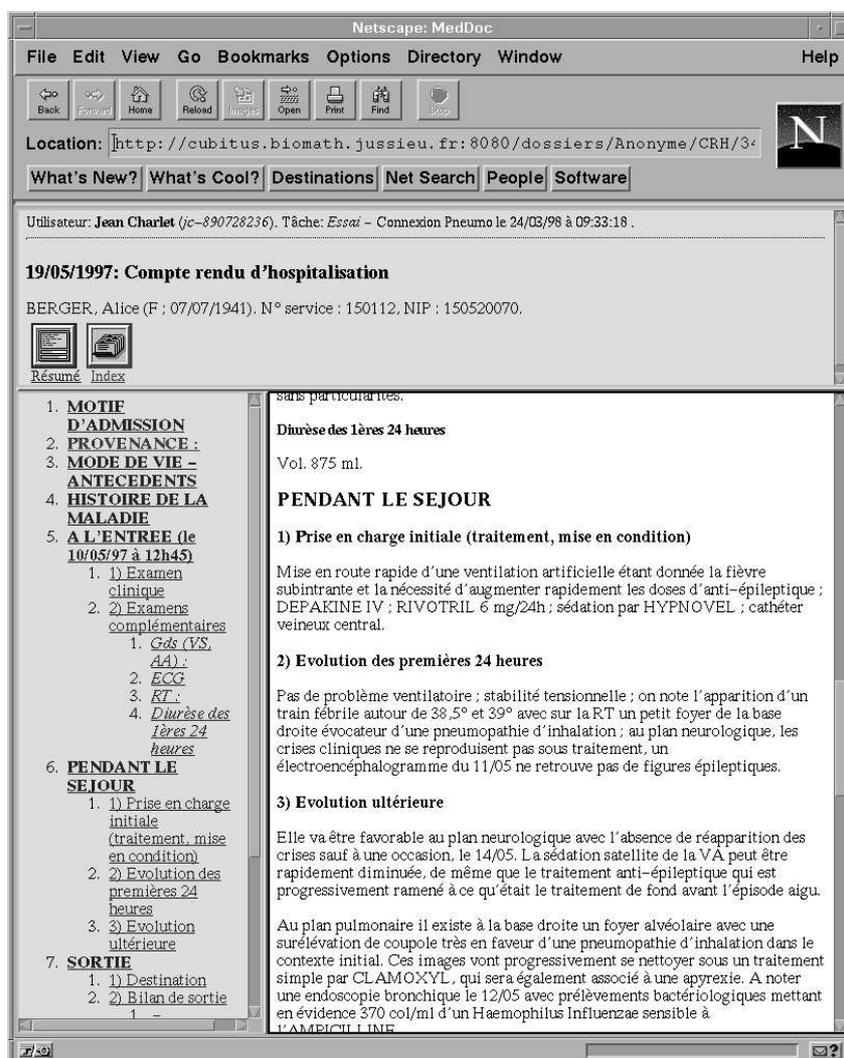


FIG. 3.3 – Visualisation d'un compte rendu d'hospitalisation.

2.4 Perspectives

Les perspectives de développement d'un tel prototype peuvent être envisagées dans un contexte industriel. Pour des raisons trop longues à développer ici, HOSPITEXTE n'est pas en cours de validation à l'AP-HP. En revanche, il sert de base technologique à un projet auquel nous participons – HTSC, `<http://www.laria.u-picardie.fr/Equipes/ic/htsc/>` – de développement d'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, d'autres perspectives sont à envisager par rapport aux buts affichés au départ et plus généralement vis à vis des autres champs d'investigation de l'informatique médicale (Charlet *et al.*, 1998).

2.4.1 Sémantique hypertextuelle

Les fonctionnalités d’annotation qui n’ont pas encore été mises en œuvre dans HOSPITEXTE doivent être réfléchies dans le contexte global des documents de synthèse (cf. § 2.1). En effet, 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 le sens des documents au regard de leur position dans le réseau hypertextuel. Les balises SGML 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 dans le cadre d’une *sémantique hypertextuelle annotationnelle*.

Enfin, le raisonnement à partir de cas, peut permettre de proposer des « bons » parcours à un praticien inexpérimenté en enregistrant les parcours de praticiens expérimentés sur des dossiers de patients et en fonction des buts de la navigation (Kassar & Charlet, 1997)

2.4.2 Balisage XML & TALM (Section réalisée avec le concours de P. Zweigenbaum)

Le traitement automatique du langage médical (TALM), par ailleurs beaucoup étudié vis-à-vis de l’encodage des diagnostics – *e.g.* (Sager *et al.*, 1995; Gundersen *et al.*, 1996; Zweigenbaum *et al.*, 1995b) – est le compagnon naturel de l’approche documentaire : en effet, le TALM intervient, dans ce contexte, pour identifier des portions de texte ayant des significations spécifiques, ici médicales, comme des *diagnostics* ou des *actes*, etc. En accord avec les principes posés ici, le processus d’identification permet de *baliser* le texte et d’annoter le document avec de nouvelles informations, à l’exclusion de tout remplacement par des données telles que des codes (cf. exemples dans (Sager *et al.*, 1996; Zweigenbaum *et al.*, 1998)). L’intégrité du document est donc conservée avec sa propriété de document / texte de *référence*⁷.

Des architectures de traitement du langage (naturel) où les modules agissent sur du langage encapsulé dans des structures SGML ont été définies dans des projets comme MULTTEXT <<http://www.lpl.univ-aix.fr/projects/multtext/>>. Un processus typique enchaîne alors les étapes suivantes : (a) sélectionner la portion de texte à traiter – *e.g.* la section *antécédent* d’un compte rendu d’hospitalisation (CRH) –, (b) exécuter le programme qui délimite des expressions précises – *e.g.* *diagnostiques* –, (c) produire une nouvelle expression pour chacune – *e.g.* un code CIM-10⁸, (d) *baliser* chacune des expressions repérées avec l’élément « diagnostic » en incluant le code produit dans un *attribut* de cet élément. Une telle approche est additive et monotone : elle permet p. ex. d’associer des expressions à des codes ou d’identifier et catégoriser des informations sur la posologie de médicaments (Evans *et al.*, 1996). Par ailleurs, elle permet de respecter le caractère contextuel de l’information médicale tel qu’il est débattu ici et d’effectuer

⁷Référence indispensable dans un cadre de responsabilité médico-légale.

⁸Classification internationale des maladies, version 10.

des repérages dans des portions spécifiques de textes : un diagnostic dans la partie « conclusion » d'un CRH n'aura pas la même signification que s'il est repéré dans la partie « antécédents ».

Enfin, le TALM permet d'envisager de nouveaux « navigateurs » – dont il faudra valider la légitimité – sur des expressions repérées *a priori* : on peut par exemple imaginer naviguer de proche en proche sur toutes les occurrences d'expressions diagnostiques dont le code CIM porte spécifiquement sur un membre inférieur. On crée alors des liens sur la base de parties « cachées » du texte – *i.e.* des *attributs d'éléments*. Ces possibilités ouvrent la voie à de nouveaux usages du dossier médical électronique dans le cadre de mécanismes de navigation dynamiques (Zweigenbaum *et al.*, 1998).

2.4.3 Réflexions et approches complémentaires

Cette approche n'est pas unique en son genre, Frénot *et al.* (Frénot & Laforest, 1999; Laforest *et al.*, 2002) proposent une architecture de dossier médical qui articule un niveau d'acquisition de l'information à l'aide de documents orientés paragraphes et un niveau agrégé de stockage de l'information dans une base de données relationnelles. Cette approche dont la clé est l'extraction des informations contenues dans les documents pour alimenter la base de donnée doit permettre de combiner souplesse d'acquisition et efficacité de stockage.

Le HL7 (*cf.* chap. 5, § 3.2) intègre des réflexions de ce type dans un contexte de communication et d'échange de données médicales.

3 Une aide à la décision « textuelle »

(Les descriptions de cette section sont tirées de (Séroussi *et al.*, 2000))

L'objectif du système ONCODOC est de proposer un ensemble de recommandations thérapeutiques adaptées à un patient donné, une fois le diagnostic posé. C'est un système d'aide à la décision qui se situe dans ce qu'on appelle en médecine le cadre des guides de bonne pratique (GBP). S'opposant aux approches classiques totalement automatisées, ONCODOC (Bouaud *et al.*, 1998) se présente comme un système d'aide à la décision, à mi-chemin entre la formalisation et le texte. ONCODOC apparaît ainsi comme une base de connaissances formalisée mais dans une logique documentaire, c'est-à-dire opérationnalisée par la lecture du médecin qui garde ainsi le contrôle d'une interprétation en contexte des éléments de connaissance qui lui sont fournis. À l'issue d'une navigation hypertextuelle centrée-patient, les recommandations thérapeutiques apparaissent sous la forme d'une liste ordonnée par efficacité de plans de soins complets, chacun d'eux étant composés de modules élémentaires de soins (chirurgie, chimiothérapie, radiothérapie, hormonothérapie) appelés *unités thérapeutiques*.

Les divers composants du modèle décisionnel proposé – les arbres de décision, les critères et leurs valeurs, les unités thérapeutiques – sont implémentés en SGML, à travers une DTD. Une application particulière, comme celle sur le traitement du cancer du sein, s'encode en créant une « instance de document », contenant toutes les connaissances et informations nécessaires, conforme au modèle constitué par la DTD. La représentation SGML, une fois constituée, est le référentiel de l'application. Un ensemble de programmes a été développé pour effectuer les transformations sur le GBP SGML. Tout ou partie de son contenu peut ensuite être extrait sous différentes formes et sur différents supports, « électroniques » pour une diffusion télématique, « papier » pour une lecture plus classique, le format et le support étant adaptés à une tâche donnée. Les productions principales sont les cahiers de validation, équivalents papier de la ressource initiale, et l'hypertexte qui permet la consultation du modèle décisionnel depuis un navigateur. La figure 3.4 propose une représentation de l'architecture générale du système.

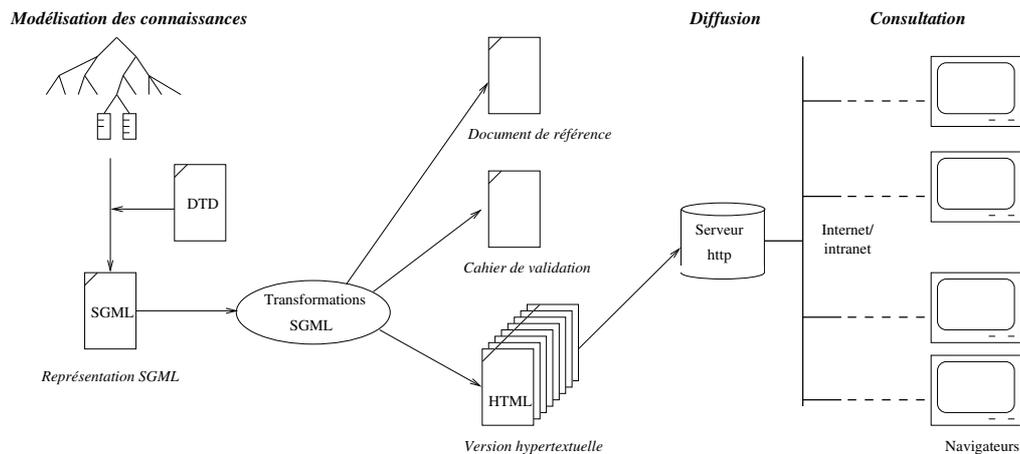


FIG. 3.4 – L'architecture générale du système ONCODOC (figure reprise de (Séroussi *et al.*, 2000)).

Dans une philosophie similaire à celle de HOSPITEXTE, ONCODOC veut éviter l'écueil de la formalisation des connaissances médicales qui ne peut anticiper la variabilité de la médecine que les médecins constatent en traitant certains patients. Le caractère textuel des descriptions et des termes médicaux utilisés – *i.e.* leur ambiguïté – n'est – ici comme pour HOSPITEXTE – pas un problème pour le médecin. Au contraire, cela lui laisse la capacité et le choix d'interpréter les descriptions textuelles en situation : pour un patient donné, le contexte clinique et l'expérience du médecin permettent en effet de désambiguïser ces termes et de leur donner la « bonne » interprétation, c'est-à-dire, celle qui permet l'opérationnalisation des connaissances.

À l'opposé d'une approche classique de décomposition du processus d'affectation thérapeutique où le patient réel est catégorisé par un correspondant formel (*cf.* figure 3.5 (a)), la navigation hypertextuelle informelle au sein de l'arbre de décision permet au médecin de capitaliser et interpréter les caractéristiques cliniques et physiologiques du patient réel et de construire contextuellement un meilleur patient formel pour lequel seront proposées des solutions thérapeutiques optimales (*cf.* figure 3.5 (b)).

Appliqué au traitement du cancer du sein, le système a été évalué en vraie grandeur (Séroussi *et al.*, 1999) par les oncologues du service d'oncologie médicale Pitié-Salpêtrière qui ont respecté à près de 80% les recommandations du système. Ces résultats significativement supérieurs aux meilleurs chiffres publiés démontrent une adéquation de l'approche proposée aux besoins des médecins pour un accès facile et rapide aux informations centrées-patient, tant sur les principes fondamentaux qui s'appuient sur une participation active des utilisateurs dans la catégorisation contextuelle – essentielle – de leur patient, que sur les choix d'implémentation adoptés avec l'élaboration de bases de connaissances sous la forme d'un arbre de décision aux branches exclusives et exhaustives et dont la séquence des nœuds respecte la logique du raisonnement médical.

4 CISMef

RDF (*the Resource Description Framework*) normalise l'écriture de métadonnées associées à un document aussi bien que la référence aux balises utilisées pour les déclarer. Si ces métadonnées ne sont pas exactement une nouveauté et sont déjà connues de ceux qui écrivent du HTML, elles sont ici décrites dans un langage (*a*) normalisé et

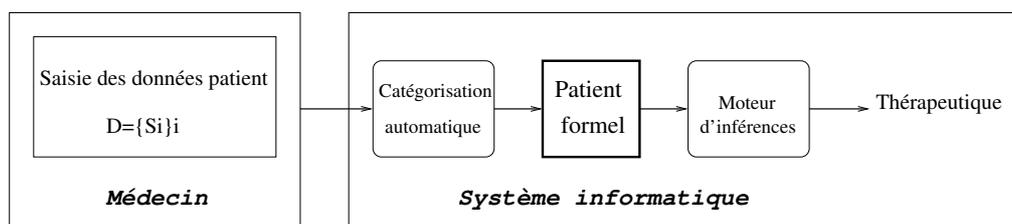


FIG. 3.5 – (a) Affectation thérapeutique : un découpage de l'approche totalement implémentée.

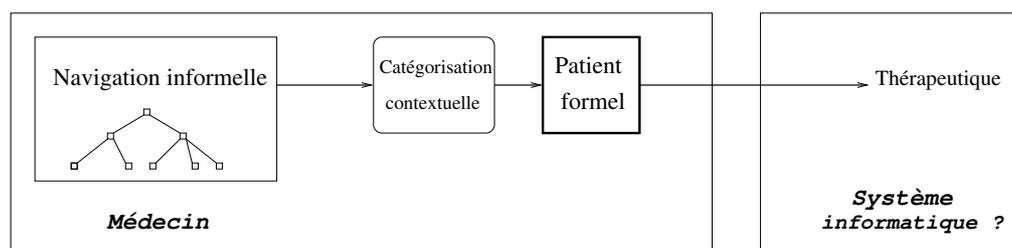


FIG. 3.5 – (b) Affectation thérapeutique : un découpage de l'approche ONCODOC (figure reprise de (Séroussi *et al.*, 2000)).

(b) dont la puissance d'expression répond à la plupart des besoins. Il est par ailleurs intéressant de noter qu'elles permettent d'attacher, à des documents dont la structure est aussi libre que nécessaire, des renseignements précis, par exemple bibliographiques (auteur, date, etc.), dont le format est aussi contraint que nécessaire. C'est ce type de fonctionnalités qui a fait dire à certains que XML retrouvait les bienfaits des bases de données (Cotte, 1999)⁹.

Le projet CISMeF est un premier exemple des résultats, en médecine, de la normalisation des métadonnées. Cet exemple est, à dessein, choisi dans le domaine de la mise à disposition des fonds documentaires car il représente pour nous les premiers pas vers la gestion des connaissances. Ainsi, le site CISMeF du CHU de Rouen, reconnu en France comme étant le site de référence en informatique médicale, « catalogue » et indexe l'ensemble des sites médicaux francophones de qualité. En dehors du fait que cela ne se fait pas sans méthode et sans une certaine force de travail, il est intéressant d'explorer d'un peu plus près les modes d'indexation des sites (Darmoni *et al.*, 2000; Darmoni & Thirion, 1999) : ainsi, quand une page Web est cataloguée, elle est indexée pour pouvoir être retrouvée et reproposée aux intéressés. RDF est alors utilisée pour décrire des informations de type bibliographique au sujet de la page (langue utilisée dans le document, année, auteurs, sujet (mots-clés), etc.) : c'est un ensemble complet de métadonnées développé par une commission internationale et multidisciplinaire et qui décrit des ressources en lignes à l'aide d'une quinzaine d'éléments (Bourda & Hélier, 2000), le *Dublin Core*¹⁰.

On a reproduit figure 3.6 ci-dessous une partie des balises décrivant l'index théma-

⁹Ce n'est ni une surprise ni un problème; c'est même tout l'intérêt de XML de s'adapter à des besoins de structurations divers. Nous ne discuterons pas ici des activités autour des bases de données mais nous pouvons noter que des recherches récentes essaient de mettre au point des modèles de bases de données semi-structurées permettant de prendre en compte des documents XML de la façon la plus efficace possible. À ce sujet, on pourra lire (Le Maître *et al.*, 2000).

¹⁰Un mécanisme implémenté dans RDF, les « espaces nominaux » ou *Namespaces*, permet par ailleurs de décrire la source des balises, c'est-à-dire une page Web où sont répertoriées des balises pour un usage précis. Dans le cadre du projet CISMeF, c'est le *Dublin Core* qui a été retenu (DC, 1999) et les balises utilisées en portent la « marque », DC.

tique du catalogue CISMeF. On voit que les balises du *Dublin Core* (DC.) y sont complétées par des balises correspondant à diverses classifications que l'équipe CISMeF avait besoin d'utiliser pour décrire les sites au regard d'un certain nombre de critères d'accessibilité, de coût d'accès, d'utilisateurs cibles, etc.

```
[...]
<meta name="DC.Language" content="fr">
<meta name="DC.Title" content="CISMeF">
<meta name="DC.Title.Subtitle"
  content="Catalogue et Index des Sites Médicaux Francophones :
  classement thématique">
<meta name="DC.Type" content="texte.guide ressources">
<meta name="DC.Subject.Keywords"
  content="(SCHEME=MeSH) France ; Internet ; médecine ; santé ;
  médecine ; health">
<meta name="DC.Creator" content="équipe CISMeF">
<meta name="CISMeF.Cible"
  content="professionnel de santé ; étudiant ; patient/grand public ">
<meta name="DC.Publisher" content="Rouen University Hospital">
<meta name="DC.Format" content="text/HTML">
<meta name="DC.Identifiant"
  content="http://www.chu-rouen.fr/ssf/santspe.html">
<meta name="DC.Description"
  content="sélection, description et indexation des ressources
  francophones de qualité dans la santé ; accès thématique">
<meta name="DC.Rights"
  content="© CHU de Rouen . Toute utilisation partielle ou totale
  de ce document doit mentionner la source">
<meta name="DC.Date.Creation" content="1995-02-02">
<meta name="DC.Date.Lastmodified" content="quotidien ; daily basis">
<meta name="CISMeF.Acces" content="libre">
<meta name="CISMeF.Cout" content="gratuit">
[...]
```

FIG. 3.6 – Partie des métadonnées de la page <http://www.chu-rouen.fr/ssf/santspe.html>, index thématique des sites en santé du catalogue CISMeF.

Par ailleurs, les mots-clés sont choisis dans le thésaurus MeSH (*Medical Subject Heading*) qui est originellement disponible en anglais et a été traduit en français par l'INSERM. Ce thésaurus offre une organisation hiérarchique et associative et permet par exemple de répondre à une requête sur les « virus » en proposant aussi des documents sur les antiviraux ou les « vaccins antiviraux » (cf. chap. 4, § 3.3).

Cette problématique des thésaurus est évidemment fondamentale pour toute activité où l'on cherche à partager de l'information. On a là les deux clés de l'interopérabilité : respecter les mêmes syntaxes pour pouvoir communiquer – norme XML – et donner le même sens aux mots – *via* les thésaurus – pour échanger des informations qui aient un contenu compréhensible par tous et de façon identique (cf. chap. 5, § 6.1). La question de l'accessibilité des différents utilisateurs de CISMeF aux ressources indexées sera spécifiquement discutée au niveau des problématiques d'indexation (cf. chap. 4, § 6.4).

5 Au sujet du Web sémantique

En guise de conclusion de ce chapitre, nous allons nous intéresser au Web sémantique. Le Web sémantique est un sujet à la mode mais ce n'est pas que cela et, dans le contexte médical, réfléchir aux conditions de coopération d'êtres humains avec des agents logiciels ou d'agents logiciels entre eux – le « but » du Web sémantique défini par T. Berners-Lee – permet d'avoir une vue globale et prospective des problèmes qui se posent et des directions de recherche pour les résoudre de façon, justement, globale¹¹. En reprenant une distinction soulignée dans (Laublet *et al.*, 2002), les problématiques des applications textuelles décrites dans ce chapitre, ressortissent à la question des métadonnées au sein de Web sémantique¹² :

Les métadonnées subjectives. Ce sont, par exemple, les annotations nécessaires à la mise en œuvre d'une sémantique hypertextuelle (cf. § 2.4.1) dans le projet HOSPITEXTE. En effet, pour proposer des services d'agréments de documents ou parties de documents dans le but d'en reconstruire une synthèse – e.g. élaborer un CRH à partir du dossier médical –, il est nécessaire de typer ces (parties de) documents. Ce typage définit plusieurs catégories de documents en fonction des desideratas de l'auteur : il définit des lectures personnelles et temporaires du praticien et est la condition de satisfaction de la fonctionnalité. Ce typage est une métadonnée sur la partie de documents visée.

les métadonnées objectives. Ce sont par exemple les métadonnées du projet CISMED qui reprend le *Dublin Core* et l'augmente d'informations considérées comme nécessaires à la bonne description des sites médicaux comme la cible visée par les documents ou le type d'accès proposé (libre, gratuit, . . .). Ce sont encore des informations simples mais elle fournissent de nombreux renseignements et préfigurent des systèmes beaucoup plus riches.

Enfin, nous pensons que la problématique et le succès du Web sémantique ont de fortes convergences avec les approches textuelles défendues ici : en effet, le succès du Web vient du fait que les utilisateurs ont eu à leur disposition un énorme fonds documentaires d'une information contextuelle, utilisable : des documents. Il y a de nombreux défauts liés en particulier à la crédibilité de ce qui est écrit ou au mode de recherche des moteurs mais, pour peu qu'on sache s'en servir de façon *ad hoc*, on a des recherches couronnées de succès. Dans notre cas, l'approche documentaire déclinée sur l'accès aux documents, l'aide à la décision ou l'indexation fournit des documents contextuellement compréhensibles. Comme pour le Web sémantique, la question de leur récupération, donc de leur indexation dans une problématique de recherche d'information est une des clés de leur accès et un certain nombre de travaux s'y attèlent comme nous verrons plus loin dans ce mémoire.

Après ces considérations spécifiquement liées aux textes et à leurs annotations, nous poursuivons les interrogations liées au Web sémantique appliqué à la médecine au chapitre 4, paragraphe 7.

Crédits

Les travaux développés ici doivent beaucoup à B. Bachimont qui a copiloté le projet HOSPITEXTE avec moi pendant qu'il était encore à l'AP-HP et V. Brunie et S. El Kassar, nos stagiaires respectifs de thèses, pour les travaux liés à HOSPITEXTE. Sans oublier B. Séroussi et J. Bouaud qui ont piloté le développement de la maquette de DOME. Par ailleurs, dans ce chapitre, je décris des travaux que je n'ai pas effectués et ai mis leurs auteurs à contribution : je remercie donc J. Bouaud, B. Séroussi, S. J. Darmoni et P. Zweigenbaum pour leur coopération.

¹¹Nous ne développons pas ici ce qu'est le Web sémantique. Une importante littérature est accessible à ce sujet sur le Web, à commencer par le serveur du Web sémantique <<http://www.w3.org>>. On pourra aussi voir le site de l'action spécifique du CNRS à ce sujet <<http://www.lalic.paris4.sorbonne.fr/stic/>> ou lire un article de synthèse récent (Laublet *et al.*, 2002).

¹²Les métadonnées et les annotations ne correspondent pas exactement à la même chose dans le contexte du Web sémantique. On peut poser rapidement que les annotations sont des informations rajoutées à un document au cours du processus d'annotation/écriture et que les métadonnées sont des informations rattachées de façon externes à un document et respectant un schéma prédéfini. Pour ce qui nous intéresse ici, la nature des informations, nous parlerons indifféremment d'annotations ou de métadonnées.

De nécessaires ontologies ?

Dans ce chapitre nous abordons la notion d'ontologie en prenant une position affirmée sur ce que nous pensons que c'est et que ça n'est pas. Cette position nous semble nécessaire à la compréhension de ce que l'on peut faire ou non avec les ontologies comme avec de nombreux « produits » terminologiques tels que les thésaurus ou les bases de données lexicales. Nous prenons appui sur les travaux développés au sein de la communauté Ingénierie des connaissances, s'appuyant eux-mêmes sur des travaux plus anciens comme la terminologie ou la recherche d'information, puis nous développons la conception des ontologies en nous fondant sur des travaux et exemples récents, en médecine comme dans d'autres domaines. Nous défendons, à l'occasion de la description de travaux développés en collaboration, l'utilisation d'outils du traitement du langage naturel comme mode d'accès privilégié aux concepts d'une ontologie. Des ontologies en médecine, développées durant ces travaux servent d'exemples à notre propos. Nous terminons ce chapitre sur le problème de l'indexation en lien avec les ontologies.

1 Introduction et définitions

1.1 Les ontologies en IC

Comme nous l'avons vu au chapitre 1, les ontologies sont apparues en Ingénierie des connaissances et plus largement en Intelligence artificielle¹, 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.*

¹Pour ne pas dévier de notre propos, nous ne développons pas l'historique de la classification en Intelligence artificielle mais un certain nombre de travaux s'étaient déjà posés des questions au sujet de la classification des objets du monde. Parmi ceux-ci, on peut citer les travaux autour des réseaux sémantiques, comme Brachman (1983; 1991) s'interrogeant sur la nature du lien *is-a* ou Woods (1991) s'attachant à des questions de représentation et de niveau de description des connaissances.

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 – que ce soit celui de l'Ingénierie des connaissances, de l'Intelligence artificielle ou des sciences de gestion et de leurs systèmes d'information ou plus récemment le contexte du Web sémantique – 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 contexte², 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*. On peut alors proposer la définition de Gruber (1993) :

Ontologie (déf. 2) : *Une ontologie est une spécification explicite d'une conceptualisation.*

Cette définition donne une vue figée d'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. Pour avancer dans ce processus, 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'élaboration³, à savoir l'ontologie comme un *système conceptuel informel*, l'ontologie comme *la représentation d'un système conceptuel via une théorie logique et son vocabulaire*. On peut alors proposer la dernière définition (pour ce paragraphe) de ce qu'est une ontologie, en reprenant les spécifications de M. Uschold *et al.* (1996) qui ont appliqué les résultats de ces réflexions dans le contexte applicatif d'une ontologie d'entreprise, *the enterprise ontology* :

Ontologie (déf. 3) : *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 vocabulaire⁴ de termes et une spécification de leur signification.

[...]

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

Les trois définitions proposées fournissent des spécifications des ontologies cohérentes entre elles. Elles précisent, en particulier la troisième, ce qu'est une ontologie par rapport à une application informatique et nous permettent 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 ;

²La construction d'un artefact n'est évidemment pas la seule raison d'une réflexion conceptuelle comme nous le verrons au § 1.2 mais l'implique indubitablement.

³Les développements faits ici sont une libre interprétation des attendus des travaux de N. Guarino et P. Giaretta (1995), en particulier par rapport aux méthodologies abordés au § 1.2. 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.

⁴« 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. § 2).

- 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 partiellement⁵.

Arrivés à ce stade, nous avons une définition qui rend compte d'un certain processus de construction mais qui ne donne aucune méthodologie réelle pour construire une ontologie. C'est ce à quoi nous allons nous attacher par la suite en revenant d'abord vers des conceptualisations plus anciennes, antérieures à l'existence d'artefacts informatiques, l'Ontologie vue par ARISTOTE et les taxinomies en science naturelle.

1.2 Repérer, organiser les objets du monde

1.2.1 L'Ontologie d'ARISTOTE

Repérer et organiser les objets du monde est une activité humaine qui a quelques siècles (!). Par rapport aux ontologies, deux types de travaux vont nous intéresser :

ARISTOTE a défini l'Ontologie comme la science de l'Être⁶. Définition qu'on retrouve dans le Petit Robert avec « **Ontologie** : la partie de la métaphysique qui s'intéresse à l'Être en tant qu'Être ». Mais l'Ontologie est habituellement davantage comprise comme une science des étants que comme une science de l'Être en tant qu'Être, c'est-à-dire qu'elle s'intéresse davantage à ce qui existe (les étants ou existants) qu'aux principes de ce qui existe (l'Être). Cette science, l'Ontologie, produit des ensembles, les ontologies.

Pour ce qui nous préoccupe, ARISTOTE s'intéresse à la nature des choses et aborde la question de l'essence et du genre spécifique.

Il (Aristote) définit l'essence des êtres qui n'existe à part des êtres individuels sensibles. Cette essence doit être discernée au sein des substances sensibles et dégagée au moyen d'une opération intellectuelle et dégagée au moyen d'une opération intellectuelle appelée abstraction.

[...].

Le travail d'abstraction discerne ce qu'il y a d'intelligible dans la substance et qui est dû à sa forme. Cet intelligible s'exprime au moyen de concepts, qui ont chacun une universalité (le concept de cheval vaut pour tous les chevaux). Ce discernement abstraitif s'effectue au départ d'une observation de chaque substance et de la comparaison des substances entre elles. La comparaison implique un classement au moyen des concepts. Ceux-ci ont des extensions diverses et l'un peut recouvrir l'autre. Ainsi les concepts, comme les classes, ont entre eux des relations hiérarchiques de genre et d'espèce. La définition d'une essence désigne toujours une espèce qu'elle situe à l'intérieur d'un genre selon la formule : « espèce = genre prochain + différence » (Canivet, 1999).

Ainsi, l'homme appartient au genre animal et se différencie spécifiquement par sa raison. *L'homme est un animal raisonnable.* ARISTOTE propose ici la première méthode d'élaboration d'une ontologie, dans son cas, l'Ontologie : les concepts se définissent par genre et différence au sein d'une relation de subsomption. Les méthodes proposées section 2 mais, tout de suite, les taxinomies en sciences naturelles s'en sont inspirées.

1.2.2 Sciences naturelles et taxinomies

La science a toujours eu pour premier but de repérer et classifier les objets du monde pour les comprendre, comprendre leur fonctionnement et leur genèse⁷. La recherche s'est systématisée en sciences naturelles, d'abord en botanique et ensuite pour

⁵ La théorie logique représentant formellement l'ontologie est l'*engagement ontologique* de cette dernière (Guarino & 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.

⁶ Si la pratique ontologique date de ARISTOTE, le mot lui-même date du 17^e siècle, avec des dates précises discordantes selon les sources.

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.

Pour être profitable à ses utilisateurs, biologistes et non biologistes, la classification des êtres vivants doit être prédictive : on doit pouvoir y situer les espèces nouvelles pour la science, et pouvoir déduire les propriétés biologiques des espèces, qu'elles soient inconnues ou déjà nommées, d'après leur position dans la classification. La classification doit de plus être aussi indépendante de l'observateur que possible. La classification constitue donc un système de stockage et de restitution de l'information biologique. Le principe d'organisation qui correspond le mieux à ces contraintes est celui des relations de parenté : celles-ci reflètent l'histoire de la vie, qui est unique et donc indépendante de l'observateur ; elles traduisent bien les propriétés biologiques, qui sont d'autant plus partagées par deux organismes vivants que leur ancêtre commun est proche ; enfin elles présentent l'avantage de situer les classifications dans la théorie unificatrice du vivant, qui est la théorie de l'évolution. Les classifications modernes sont donc phylogénétiques, et agrègent les organismes, les espèces, les genres et les familles en fonction de la proximité de leur ancêtre commun. Intégrer de façon aussi pertinente que possible dans les classifications biologiques l'histoire des êtres vivants, leurs propriétés biologiques et les mécanismes évolutifs de leur diversification constitue l'enjeu scientifique majeur de la systématique. Des classifications obtenues dépendent d'innombrables décisions scientifiques, politiques et économiques : en organisant notre vision du monde vivant, les classifications biologiques déterminent en grande partie la façon dont nous interagissons avec les autres espèces de notre planète dans presque tous les domaines de notre activité [Réseau national de biosystématique].

⁷Ces réflexions sont le fruit de différentes lectures, à commencer par « L'Histoire naturelle » de Buffon (Pillot & Cuvier, 1832), les entrées sur les *mammifères* et la *systématique* de l'Encyclopædia Universalis, les URL du réseau national de systématique <<http://rnb.snv.jussieu.fr>> et des pages de l'exposition virtuelle Bioscope à la Faculté Universitaire notre-dame de la Paix (Namur) <<http://www.fundp.ac.be/bioscope>>. Elles n'engagent que leur auteur, en particulier au niveau des savants cités comme de l'historique où nous avons pris le parti de développer certaines idées propres à replacer les ontologies dans un continuum scientifique qui a débuté bien avant l'invention des ordinateurs et de l'Intelligence artificielle.

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 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 même, « avoir un tégument à plumes » est l'essence même des oiseaux alors que « voler » est contingent. Le fait que les oiseaux volent est alors seulement une connaissance du sens commun et inférer le vol à partir de l'existence d'un oiseau est un raisonnement prototypique. Cet exemple (simplifié au regard du nombre d'attributs définitoires des espèces) nous montre encore que décider du définitoire et du contingent n'est pas toujours évident dans ce domaine pourtant ancien, comme dans d'autres et qu'il n'est pas étonnant qu'il soit difficile de repérer et classer les objets d'un domaine : les critères de classification dépendent des buts poursuivis et n'ont rien d'immuables.

1.2.3 Enseignements pour les ontologies

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 – *cf.* § 1.3) 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.* § 2 et 5).
- l'exemple des oiseaux qui volent (en général) nous amène à nous interroger sur la nature de l'ontologie construite : il ne s'agit pas de confondre des connaissances scientifiques (pour les oiseaux) ou techniques (pour des artefacts) avec des connaissances de sens commun. À ce sujet, il n'y a pas de connaissances scientifiques qui seraient plus nobles que des connaissances de sens commun. Il faut simplement prendre garde au fait que ces dernières sont, par leur nature, pas toujours cohérente et pas faciles à conceptualiser, en tous cas encore moins que les connaissances scientifiques.
- Les noms des espèces, les labels des concepts de l'ontologie pour nous, sont très important car ils conditionnent la compréhension qu'on en a. Par ailleurs, la place des espèces au sein d'un arbre construit par différences spécifiques est importante à saisir. D'où, pour VON LINNÉ, une nomenclature binomiale et, nous le verrons plus loin pour les constructeurs d'ontologie, la nécessité de se doter d'interfaces et de ressources textuelles permettant de comprendre la structure hiérarchique de la classification.

- De façon plus prospective, la taxinomie phylogénétique nous interroge sur l'évolution des concepts au sein des ontologies (cf. chap. 6 § 1.3).

1.3 Que représente-t-on dans une ontologie ?

À partir des définitions et caractéristiques proposées jusque-là pour les ontologies, trois grands types de caractéristiques nous permettent de préciser ce qu'on peut représenter avec une ontologie :

Les propriétés. Une ontologie est non seulement le repérage et la classification des concepts mais c'est aussi des caractéristiques qui leur sont attachées et qu'on appelle ici des propriétés. Ces propriétés pouvant être évaluées. Pour reprendre les taxinomies en sciences naturelles, les vertébrés ont justement un tégument (la peau) comportant des poils – *p. ex.* pour les mammifères – ou des plumes – *p. ex.* pour les oiseaux. Dans le cas d'une application médicale, MENELAS, un patient a un âge qui a une certaine valeur ou est soigné par tel médecin. Ainsi, les propriétés vont impliquer potentiellement un (relation unaire), deux (relation binaire) ou, plus rarement, n (relation n-aire) concepts.

Le type d'ontologie. Les méthodes en Ingénierie des connaissances ont répertorié plusieurs types d'ontologie liés à l'ensemble des objets conceptualisés et manipulés au sein d'un SBC. Nous allons en citer quelques-unes : on a (1) l'ontologie du domaine (cf. § 1.4, premier point), (2) l'ontologie générique ou qui se veut comme telle et qui repère et organise les concepts les plus abstraits du domaine ou autre (cf. § 5), (3) l'ontologie d'une méthode de résolution de problème où le rôle joué par chaque concept dans le raisonnement est rendu explicite (*p. ex. signe* ou *syndrome* dans le cadre du raisonnement médical), (4) l'ontologie d'application qui se veut une double spécialisation : d'une ontologie du domaine et d'une ontologie de méthode, enfin (5) l'ontologie de représentation qui repère et organise les primitives de la théorie logique permettant de représenter l'ontologie (*p. ex. la frame ontology* d'ONTOLINGUA– Gruber, 1993).

Les liens reliant les concepts. La relation de subsomption *is-a* qui définit un lien de généralisation – *i.e. hyperonymie* – est la plus utilisée dans les ontologies et ce depuis ARISTOTE. Mais ce n'est pas la seule possible et, surtout, pas la plus utile dans certains cas. On peut avoir besoin de relations de partie-tout ou *méronymie*. Ce type de conceptualisation est, par exemple, indispensable en anatomie médicale où il est nécessaire de décrire des organes ou des systèmes et ce qui les compose (cf. § 3.4.3).

1.4 Quelle utilité ?

Pour terminer cette entrée en matière au sujet des ontologies, il est nécessaire de préciser leur utilité au sein des SBC et dans un cadre plus large. Ce que nous allons faire en reprenant des travaux en Ingénierie des connaissances, *p. ex.* (Kassel, 2001; Uschold & Gruninger, 1996), en développant aussi de nouveaux usages qui apparaissent, liées par exemple au Web sémantique.

Les connaissances du domaine d'un SBC. La première et originelle utilité d'une ontologie était liée à une volonté de réutilisation (cf. chap. 4, § 4.2). Cette propriété étant attendue mais ni démontrée, ni définitoire, elle est restée comme pétition de principe alors que la nature de l'ontologie se précisait au fur et à mesure que des travaux utilisaient le concept. Ainsi, réutilisable ou pas, l'ontologie devait servir de représentation des connaissances du domaine pour un SBC. Plus précisément, on peut dire maintenant qu'elle sert de squelette à la représentation des connaissances du domaine dans la mesure où elle décrit les objets, leurs propriétés et la

façon dont il peuvent se combiner pour constituer des connaissances du domaine complètes.

La communication. Les ontologies peuvent intervenir dans la communication entre humains. Dans ce cas, elles servent par exemple, à créer au sein d'un groupe ou d'une entreprise un « vocabulaire » standardisé. Pour de tels besoins, on est plutôt dans le cadre d'une ontologie informelle. On peut même se poser la question de l'utilité d'une ontologie dans cette situation et préférer un produit de type thésaurus. Mais nous verrons, dans la suite de ce chapitre, que des critères précis de construction d'ontologies sont nécessaires à la maintenance de thésaurus, sans parler des cas où une première approche informelle au sein d'une entreprise a vocation à être réinvestie dans des questions de référentiels de systèmes d'informations.

Dans le cas de la communication entre êtres humains et ordinateurs, l'ontologie est formelle et sert en général une tâche précise dans le SBC ou le système d'information (cf. point suivant). Cette situation pose la question de l'accès aux concepts de l'ontologie pas toujours évident pour un non expert. La solution peut passer par l'association d'un thésaurus métier à l'ontologie (cf. § 6).

L'aide à la spécification de systèmes. Comme pour les SBC mais au niveau du système d'information, l'ontologie fournit une liste classifiée des objets que doit manipuler le système : c'est le référentiel du système d'information. Une volonté de réutilisabilité, présente déjà dans les SBC, sous-tend l'utilisation des ontologies dans le cadre des systèmes d'information (cf. note 7, p. 8).

L'interopérabilité. L'interopérabilité est une spécialisation de la communication, dans ce cas vue entre deux ordinateurs. L'ontologie répertorie alors les concepts que des applications peuvent s'échanger même si elles sont distantes et développées sur des bases différentes. Cette interopérabilité est *l'interopérabilité sémantique* qui s'appuie d'abord sur une *interopérabilité syntaxique* (cf. chap. 5, particulièrement le § 6.1).

L'indexation et la recherche d'information. Plus récemment, les travaux autour du Web sémantique ont réactivé la problématique et l'utilisation des ontologies : en sus d'un rôle de médiateur, les ontologies y sont utilisées pour l'indexation, fournissant les index conceptuels décrivant les ressources sur le Web. Ce type d'usage, ressortissant comme certains points précédents à la communication entre être humain et machines, pose la question de l'accès et la compréhension de l'ontologie.

2 Quelle méthodologie de construction ?

2.1 Au commencement était la langue

Les méthodologies de construction d'ontologies ne sont pas légion. Nous entendons par là, la donnée argumentée de procédures de travail, d'étapes, qui décrivent le pourquoi et le comment de la conceptualisation puis de l'artefact construit. De nombreuses méthodes de construction d'ontologies sont orientées sur des problèmes de cycle de vie de l'ontologie vue comme un logiciel. Elles sont par ailleurs complétées par de « bons » principes (cf. § 2.3) mais ne proposent pas de réelle méthodologie. À l'inverse, la méthodologie proposée par B. Bachimont est linguistiquement et épistémologiquement fondée (Bachimont, 2000a) et c'est elle que nous allons décrire après l'avoir plusieurs fois expérimentée. Cette méthodologie qui a été élaborée à la suite du projet MENELAS (cf. § 2.4) et de la construction de son ontologie, s'est enrichie de considérations linguistiques développées dans (Bachimont, 1996). Enfin, elle est complétée ici de considérations sur les corpus textuels développées au sein du groupe TIA (cf. note 12, p. 13) qui a constitué et constitue encore un groupe de réflexions fructueux sur, entre autres, les

corpus, les produits terminologiques et leur construction (Bourigault & Slodzian, 2000; Aussenac-Gilles *et al.*, 2002).

D'autres méthodes fondées sur des principes proches ont été élaborées au sein du groupe TIA. Ne voulant pas développer une comparaison de ces méthodes, nous renvoyons le lecteur à, par exemple, (Aussenac-Gilles *et al.*, 2000).

La première question qui se pose pour développer une méthodologie de construction d'ontologies, est le matériau de départ : nous avons développé au chap. 2, § 1 que l'Ingénierie des connaissances avait souvent recours aux textes comme matériel de base pour élaborer ses artefacts. Ensuite, il y a le matériau d'arrivée, ici une ontologie formelle qui doit servir dans un SBC⁸. La question est alors de caractériser le passage d'une connaissance exprimée sous forme linguistique à une connaissance formalisée. B. Bachimont (2001b) note alors que *les langages formels mobilisent des symboles primitifs dont le choix est fixé par le langage lui-même, et d'autres qui doivent être spécifiés par l'utilisateur du langage. On appelle les premiers les primitives logiques, les seconds les primitives non logiques. Ce sont en particulier les symboles de prédicats et de fonctions du langage formel. L'objectif des ontologies est donc de spécifier le choix des primitives non logiques ainsi que leur signification.* On peut alors poser cette quatrième définition d'une ontologie :

Ontologie (déf. 4) : *Définir une ontologie pour la représentation des connaissances, c'est définir, pour un domaine et un problème donnés, la signature fonctionnelle et relationnelle d'un langage formel de représentation et la sémantique associée* (Bachimont, 2000a).

Cette définition ne ressortit pas, contrairement à ce que suggère une lecture rapide, à la seule problématique formelle. Au contraire, la question de la *sémantique associée* au langage formel pointe toute la difficulté qu'il y a à construire une représentation des connaissances formelle d'un monde qui ne l'est pas. Autrement dit, la donnée d'un langage formel de représentation des connaissances laisse ouverte la question des symboles fonctionnels et relationnels nécessaires et de la sémantique à y associer. Il faut construire, lister et organiser un ensemble de primitives non formelles, avec leur signification. Ces primitives – les concepts de l'ontologie – seront manipulées dans le cadre d'une théorie logique pour construire l'ensemble des significations possibles en son sein. L'art et la manière sont principalement proposés par B. Bachimont dans la méthodologie qui suit.

2.2 Une méthodologie constructiv(ist)e

2.2.1 1^{re} étape : la primauté du corpus et son analyse

Nous sommes, par hypothèse, dans des domaines où les données, connaissance, s'expriment en langue – la médecine en étant un des meilleurs exemples (*cf.* chap. 3, § 1.1). Il est alors aisé⁹ de trouver des textes explicitant les connaissances du domaine, dans des buts de pratique ou de transmission des connaissances. On peut ainsi 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

⁸Nous nous plaçons ici dans ce contexte où nous cherchons à construire une ontologie formelle. Nous avons vu au § 1.4 que ce n'est pas toujours le cas, au moins dans un état intermédiaire.

⁹Dans le sens où ces textes existent bien et en quantité. Nous n'abordons pas la question, parfois complexe, de la disponibilité réelle de ces textes sous forme numérique.

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 réseaux¹⁰.

2.2.2 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.

Les résultats de l'extraction terminologique permettent de poser des relations entre des unités linguistiques fonctionnant comme des termes selon la morphologie et la syntaxe. Ces relations sont par conséquent intralinguistiques. Pour entreprendre une modélisation sémantique, il paraît donc idoine de retenir une sémantique reposant sur ce type de relation. La sémantique différentielle est une telle sémantique : elle permet d'explicitier le sens (en contexte) d'une unité linguistique ou sa signification (hors contexte, quand le linguiste veut la considérer pour elle-même) par d'autres termes de la langue, sans faire appel à des notions ou entités extralinguistiques. Dans cette perspective, les unités linguistiques ont du sens dans la mesure où leur usage les distingue les unes des autres, si bien que le sens d'une unité se constitue des identités et des différences qu'elle entretient avec les autres unités de la langue. Nous adoptons cette sémantique et proposons de caractériser chaque unité par les unités avec lesquelles elle est identique sous certains aspects et celles avec lesquelles elle est différente sous d'autres (Bachimont, 2001b).

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 (cf. § 1.2.1), un arbre. Sans reprendre les arguments qui ont été longuement développés dans (Bachimont, 1996, 2000a), cette structure arborescente est une affirmation forte de cette méthodologie et les arguments n'ont rien à voir avec une implémentation dans un langage informatique ou un autre qui ne permettrait pas d'exprimer des héritages multiples comme dans les treillis : on est simplement en train de préparer l'élaboration des primitives non formelles d'une théorie formelle et on ne peut se permettre que leur signification soit « instable », héritée soit d'un père soit d'un autre dans le treillis. Enfin, les principes de construction de cet arbre découlent directement des fondements épistémologiques de cette méthodologie : sans les détailler, ils reprennent les principes aristotéliens de genre proche et différence spécifique (cf. § 1.2.1), additionnés de principes liés au paradigme différentiel choisi (Bachimont, 2000a).

Construction et lecture de l'arbre ontologique sont duales et l'arbre de libellés linguistiques ainsi construit pourra être lu et interprété de façon unique.

Les principes différentiels constituent donc une grille de lecture et sont des prescriptions interprétatives qu'il faut suivre pour savoir comment interpréter le libellé. C'est donc le respect de ces principes qui permet de considérer ce libellé non pas comme une unité linguistique dont le sens varie selon le contexte de son utilisation, mais comme une primitive au sens invariable. [...] On obtient un réseau dans lequel la position d'un

¹⁰Cette étape sera approfondie au § 4.2.

nœud conditionne sa signification. La signification définie par la position dans l'arbre est invariable selon les contextes. [...] En respectant les principes différentiels, en s'engageant à suivre la sémantique qu'ils prescrivent, les nœuds de l'arbre ontologique correspondent à des concepts pouvant être utilisés comme des primitives de modélisation et de formalisation. Nous venons donc de définir l'engagement sémantique à la base de l'ontologie : ensemble des prescriptions interprétatives qu'il faut respecter pour que le libellé fonctionne comme une primitive (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).

2.2.3 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 formelle doit alors être abordée.

Premièrement, les concepts formels vérifient les relations d'identité unissant les concepts sémantiques. En effet, les concepts sémantiques s'interdéfinissent par identités et différences. L'identité correspond au fait qu'une notion est comprise dans une autre : la notion de bistouri comprend celle d'instrument d'incision. Par conséquent, tout objet qui est un bistouri est un instrument d'incision. [...] Deuxièmement, la différence entre concepts sémantiques, où deux notions s'excluent mutuellement, ne se répercute pas directement sur les concepts formels. Prenons un exemple avec les notions d'acteur et d'être humain. Ces notions s'excluent : la notion d'acteur correspond à un rôle, celle d'être humain à une entité biologique. Or, une entité biologique n'est pas un rôle [...]. Mais, dès lors que l'on adopte une sémantique formelle, on ne considère plus des notions, mais des extensions d'objet. Par conséquent, un acteur, c'est l'ensemble des objets qui sont des acteurs, c'est-à-dire qui jouent un rôle. Les êtres humains sont également un ensemble d'objets. [...] à l'évidence, un objet qui est un homme peut être un objet qui est un acteur, même si la notion d'acteur n'est pas celle d'homme. Cela implique que l'on retrouve dans les relations unissant les concepts formels les relations d'héritage des concepts sémantiques, mais pas les exclusions. La structure des concepts formels n'est plus obligatoirement un arbre, mais, plus généralement, une structure de treillis. Cela se comprend d'ailleurs facilement pour la raison suivante : si la sémantique des concepts est référentielle, les relations entre les concepts sont des relations entre ensembles. La structure des concepts formels doit correspondre à la structure algébrique des ensembles, c'est-à-dire un treillis (Bachimont, 2000a).

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 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.

Pour finir, il est important de noter que nous sommes bien dans le contexte de l'*engagement ontologique* de N. Guarino : la formalisation proposée est une spécification formelle donc extensionnelle de l'ontologie ainsi définie et le sens des concepts est dans les objets définis en extension. Comme rappelé précédemment (cf. note 5, p. 45 et définition 3 de l'ontologie), les contraintes de cette formalisation font qu'elle ne peut rendre

compte exactement du sens visé au niveau sémantique, celui de l'ontologie régionale, ce qui justifie qu'elle ne puisse en rendre compte que « partiellement ».

2.2.4 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*.

Les langages de représentation les plus connus sont, comme noté auparavant, les graphes conceptuels et les logiques de description (cf. chap. 1, § 4.3), permettant, l'un comme l'autre, d'effectuer un certain nombre d'opérations sur des ontologies : inférences propres aux structures de graphes comme la jointure ou la projection pour les graphes conceptuels, classifications dans des structures arborescentes pour les logiques de description. Ceci étant, même s'ils ne percent pas réellement, d'autres langages permettent d'intéressants débats sur les niveaux de représentation et tentent de satisfaire, en navigant entre les deux extrêmes, les contraintes d'expression d'une ontologie régionale, formelle et computationnelle et les nécessaires inférences à faire dessus (cf. chap. 1, § 3.5).

2.2.5 Épilogue : les relations

Arrivés à la fin de la méthodologie, il reste à définir plus précisément les relations dans l'ontologie.

[...] elles ne se définissent pas de la même manière que les concepts, car, unissant des concepts, elles se caractérisent à partir d'eux. Si l'on ne retient que des relations binaires, les relations se définissent de la manière suivante : (1) Une relation se définit par les concepts qu'elle relie : par exemple, être animé et action ; ces concepts constituent la signature sémantique de la relation. (2) Une relation se définit en outre par un contenu sémantique intrinsèque articulant les deux concepts : par exemple, le fait que l'être animé est l'agent de l'action. La sémantique intrinsèque de la relation est spécifiée vis-à-vis des autres relations possédant la même signature sémantique selon les principes différentiels vus plus haut. Par exemple, la relation patient entre être animé et action se définit par identité et différence avec la relation agent. L'identité, c'est le fait d'avoir la même signature, la différence, c'est le fait de subir l'action plutôt que de l'exercer. L'identité n'est pas forcément réduite au fait d'avoir la même signature : par exemple, les relations agent volontaire et agent involontaire possèdent comme identité, outre la même signature, le fait d'avoir un père commun, la relation agent. Autrement dit, chaque signature sémantique est potentiellement la racine d'un arbre différentiel de relations possédant la même signature et spécifiées selon les principes différentiels. Les signatures sémantiques constituent également un arbre : on a donc un arbre de relations venant compléter l'arbre des concepts (Bachimont, 2000a).

Avec les relations, on complète les briques conceptuelles – non formelles – nous permettant de construire toutes les représentations formelles valides du domaine¹¹. Il faut remarquer que les concepts et relations de l'ontologies sont duals l'un par rapport à l'autre. Un concept primitif pourrait être un concept défini, une relation pourrait se retrouver implicitement définie au sein d'un concept primitif. Ce sont les choix de la

¹¹Ces relations sont appelées *rôles* en logique de description. Ce nom traduit bien la nature de la propriété attachée au concept par une relation : elle est non définitoire, voire temporaire comme par exemple, le fait d'être un étudiant pour un être humain (cf. § 3.1.2).

deuxième étape qui auront permis de décider de ce qui est essentiel – et donc primitif – ou non (*cf.* § 2.2.2).

L'ensemble des étapes et la nature des objets élaborés est résumé dans la figure 4.1.

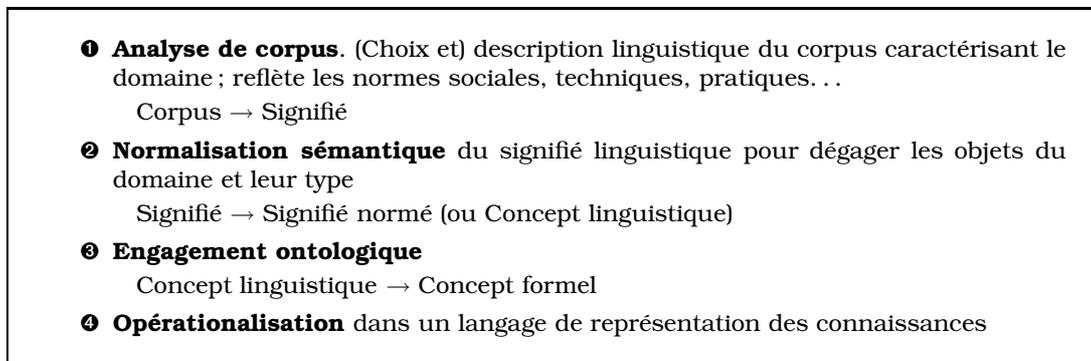


FIG. 4.1 – Les étapes de la méthodologie.

2.3 Quelques bons principes

Passés les questions de méthodologies fondamentales, un certain nombre de travaux proposent des principes de construction d'ontologies. Nous allons nous attarder sur deux d'entre eux, paradigmatiques, les travaux de Th. R. Gruber (1993) et ceux de M. Fernández *et al.* (1999).

Th. R. Gruber propose ainsi un certain nombre de principes à respecter pour construire une ontologie :

Clarté. Les ambiguïtés doivent être réduites, quand une définition peut être axiomatisée, elle doit l'être. Dans tous les cas, des définitions en langage naturel doivent être fournies.

Cohérence. Une ontologie doit être cohérente. Les axiomes doivent être consistants. La cohérence des définitions en langage naturelle doit être vérifiée autant que faire se peut.

Extensibilité. L'ontologie doit être construite de telle manière que l'on puisse l'étendre facilement, sans remettre en cause ce qui a déjà été fait.

Biais d'encodage minimal. L'ontologie doit être conceptualisée indépendamment de tout langage d'implémentation. Le but étant de permettre le partage des connaissances (de l'ontologie) entre différentes applications utilisant des langages de représentation différents.

Engagement ontologique minimal. Une ontologie doit faire un minimum d'hypothèses sur le monde : elle doit contenir un vocabulaire partagé mais ne doit pas être une base de connaissances comportant des connaissances supplémentaires sur le monde à modéliser.

D'autres principes du même type sont proposés par d'autres auteurs (*cf.* article de A. Gómez-Pérez (2000)). Il est facile de voir que la méthodologie que nous avons décrite (*cf. supra*) fournit des moyens d'appliquer ces principes, au niveau sémantique ou au niveau formel selon les cas. Le dernier principe est moins clair qu'il n'y paraît : nous avons vu (*cf.* 2.2.3), en accord avec N. Guarino, qu'un engagement ontologique est le passage obligé vers l'ontologie formelle. La question du contenu de l'ontologie par rapport à une base de connaissances est un problème qui n'est pas uniquement lié à la formalisation mais s'y surajoute (*cf.* § 3.2).

M. Fernández *et al.* proposent, dans la méthodologie METHONTOLOGY, de construire une ontologie en respectant des activités de gestion de projet (planification, assurance qualité), de développement (spécification, conceptualisation, formalisation, implémentation, maintenance) et des activités de support (intégration, évaluation, documentation). On retrouve là des problématiques de génie logiciel et de gestion de projet informatique qu'on a tout intérêt à voir s'appliquer à la construction de grandes ontologies, si on a une méthodologie réelle de construction, évidemment. Cette méthodologie rejoint pour partie celle décrite par F. Gandon (2002) et développée au sein de l'équipe ACACIA de l'INRIA.

Enfin, nous renvoyons à l'article de A. Gómez-Pérez (2000), pour une description plus complète d'ontologies et de méthodologies centrées sur des problématiques de cycle de vie des ontologies.

2.4 Une ontologie en médecine : MENELAS

MENELAS est un projet européen piloté de 1992 à 1995 par le DIAM/SIM/DSI/AP-HP. Le but du projet MENELAS était la conception et l'implémentation d'un système pilote capable d'accéder à des rapports médicaux rédigés en langage naturel dans 3 langues, l'anglais, le français et le néerlandais : ce système devait pouvoir analyser le contenu de rapports médicaux (comptes rendus d'hospitalisation ou CRH) et l'archiver dans une base de données sous la forme d'un ensemble de structures conceptuelles (graphes conceptuels de Sowa (Sowa, 1984)). Ces structures, qui constituent la représentation de chaque CRH, devaient pouvoir ensuite être consultées pour accéder à des informations spécifiques contenues dans le CRH. Une partie des informations était encodée à l'aide de nomenclatures internationales, ce qui permettait leur échange à partir de CRH écrits en différentes langues¹² (Zweigenbaum *et al.*, 1995a).

MENELAS repose sur l'hypothèse que la compréhension d'un CRH consiste à construire une représentation conceptuelle de la situation du monde décrit dans le texte. Cette hypothèse peut être justifiée par le fait que nous nous intéressons à des rapports techniques qui décrivent ce qui est arrivé au patient durant son hospitalisation. Le sous-système d'analyse du langage naturel inclus dans MENELAS utilise un analyseur morpho-syntaxique, un analyseur sémantique et un analyseur « pragmatique ». L'analyseur sémantique produit une représentation du sens sous forme de graphes conceptuels. Cette représentation correspond au sens littéral des phrases ; elle est construite à partir d'une phrase en associant des concepts à des mots grâce à un lexique sémantique : on passe « du mot au concept ». La compréhension d'un texte repose sur l'utilisation de connaissances médicales et de connaissances de sens commun qui permettent d'inférer de nombreuses informations implicites. Ces informations correspondent à des inférences effectuées naturellement par un spécialiste du domaine lorsqu'il lit un CRH. L'analyseur pragmatique a pour tâche d'obtenir un niveau de compréhension plus profond en construisant un modèle de la situation décrite : il va « du concept au concept » (Zweigenbaum *et al.*, 1995a). La figure 4.2 montre un schéma de MENELAS avec l'analyseur pragmatique et ses bases de connaissances.

La question de la construction de l'ontologie de MENELAS a été abordée de façon approfondie dans (Bouaud *et al.*, 1994) et (Charlet *et al.*, 1996a). Elle a amené la mise au point de la méthodologie (*cf.* § 2.2), principalement par B. Bachimont, et l'implémentation d'un système opérationnel, principalement par J. Bouaud (1992), se servant de cette ontologie. En dehors du fait que le pilote a bien fonctionné et que le projet a obtenu

¹²Le projet a été confronté aux problèmes habituels de la compréhension de textes en langage naturel. Il s'agit bien sûr des problèmes inhérents au langage, comme la paraphrase, l'ambiguïté, et de façon plus générale la description et la mise en œuvre de connaissances syntaxiques et sémantiques adéquates et d'une couverture suffisante. Il s'agit aussi des problèmes généraux de représentation du sens des énoncés et des connaissances à fournir au système. On rejoint alors des problématiques classiques en Intelligence artificielle : acquisition, représentation, mise en œuvre, validation de connaissances complexes.

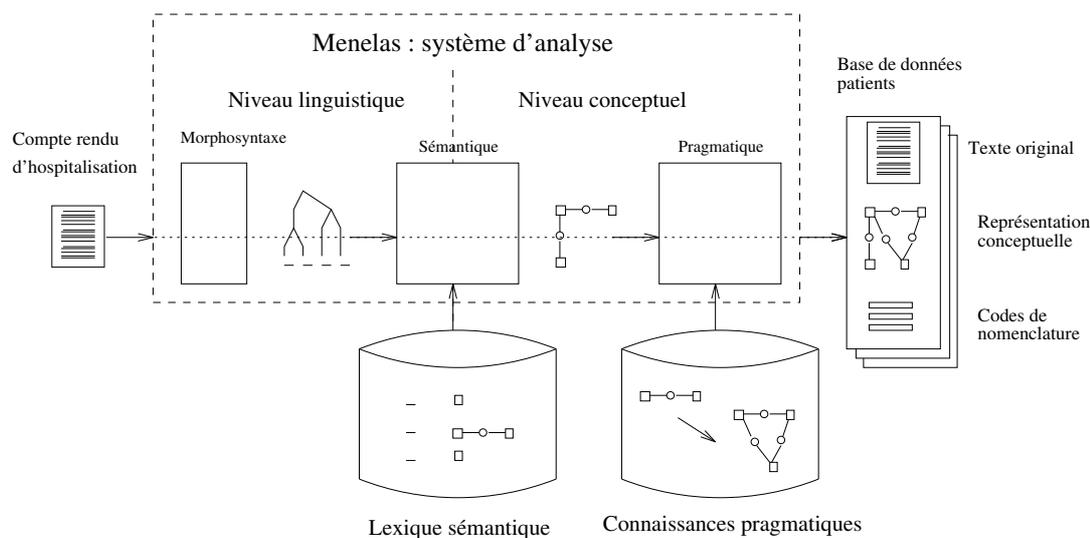


FIG. 4.2 – Analyse d'un compte rendu d'hospitalisation.

des résultats intéressants, on peut résumer rapidement les principaux enseignements de ce projet, du point de vue des ontologies, à travers plusieurs points :

- la représentation canonique construite dépend des buts visés ;
- l'ontologie construite dépend de la tâche assignée au système ;
- la réutilisabilité est problématique (cf. § 5) ;
- il est nécessaire de faire appel à des connaissances de sens communs, ou encyclopédiques, à rajouter à l'ontologie pour pouvoir effectuer les inférences nécessaires (cf. fig. 4.3 où l'on voit la représentation de ce qu'est une admission à l'hôpital pour le système).

Finalement, l'ontologie construite comporte plus de 1800 types et 300 relations. Elle est visible à l'URL <http://www.biomath.jussieu.fr>.

3 Quelques questions autour des ontologies

3.1 Deux exemples pas trop « jouets »

3.1.1 Une question de représentation

Ce premier exemple, tiré du projet MENELAS, mais il aurait pu l'être de beaucoup de projets en médecine, montre que des choix de représentation différents ont pour conséquences des possibilités d'inférences différentes. Cet exemple décrit trois représentations du genre sexuel d'un être humain (cf. fig. 4.4) :

1. Dans la première, nous avons deux types de concept, *patient* et *patiente* sur lesquels le système ne peut rien « dire » sauf qu'ils sont différents ;
2. Dans la seconde, nous avons un concept primitif (*patient*), deux attributs différents, *male* et *female* qui différencient ainsi les deux concepts mais sans que le système ait des raisons formelles de « dire » en quoi ils sont différents.
3. Dans la troisième et dernière représentation, nous – i.e. le système – avons, comme pour la deuxième, la possibilité de différencier les deux concepts définis mais, en plus, la différence est explicitement liée au genre sexuel puisque c'est un concept qui a une valeur explicite, *male* ou *female* selon le cas.

Cet exemple suggère deux types de commentaires :

```

Model admission(_x) is [admission: _x]-
  (pat)-->[human_being:_pat]
  (agt)-->[human_being:_doc]-
    (state_of)<--[knowledge_state:_ks]
    (state_of)<--[want_state:_ws] %
  (motive)-->[knowledge_state:_ks]

-->(content)-->[out_hospital]--(state_of)-->[human_being:_pat]

(motivated_by)-->[state_of_physical_object:_path]--(state_of)-->[human_being:_pat]
(result)-->[in_hospital:_ih]

(performative_goal)-->[intentional_change]--(purported_obj)-->[human_being:_pat]
(descriptive_goal)-->[state_of_physical_object:_ps]-
(intentional)

(state_of)-->[human_being:_pat]--(defines_area)-->[spatial_object:_sp3]
--(loc)-->[spatial_object:_sp3]<--[defines_area]--[ward:_w]
(consists_loc)-->[spatial_object:_sp3]%
(reason)-->[want_state:_ws]

-->(content)-->[in_hospital:_ih]--(state_of)-->[human_being:_p]
(descriptive_result)-->[state_of_physical_object:_t]<--[real]

(attr)-->[admission_performance_attr]--(val_qual)-->[admission_performance_val]%

```

FIG. 4.3 – Connaissances de sens commun sur l'admission.

1. Nous, lecteurs humains de ces représentations, lisons le côté formel et sémantique des représentations, sans être toujours attentif, parce que nous faisons inconsciemment les inférences, à ce que le SBC pourra réellement inférer ou pas.
2. Il n'y a pas une représentation meilleure qu'une autre : il y a des représentations plus précises que d'autres, nécessaires ou pas selon les cas. Mais il est sûr qu'il n'est pas viable, en termes calculatoires comme de compréhension humaine, de représenter tout le corps humain en essayant d'aller au plus profond des processus biochimiques et structures anatomiques.

3.1.2 Une question d'engagement sémantique ou ontologique

Cet exemple correspond à une question qui s'était posée aux concepteurs de l'ontologie de MENELAS, à savoir si une personne était un objet physique ou un être vivant. La réponse avait été la modélisation indiquée en seconde partie de la figure 4.5, à savoir qu'un être humain était un être vivant, qu'il avait un corps qui était, lui, un objet physique¹³.

En dehors du fait que ce choix s'est révélé correct pour le projet, il est intéressant de constater qu'on retrouve là la même question que N. Guarino a abordé (1999) et résolu en proposant, au niveau formel, une méthode à base de métapropriétés (*identité, rigidité, anti-rigidité, dépendance*) qui contraint l'organisation d'une ontologie formelle, en particulier la relation de subsomption, en fonction des propriétés des nœuds qui la composent.

Ainsi, en s'intéressant aux universaux de modélisation – qu'il appelle des propriétés –, Guarino affirme qu'un *type* est une propriété à la fois porteuse d'un critère d'identité – C'est-à-dire qu'on est capable de donner un critère qui définit une instance par rapport à une autre – et *rigide* – *i.e.* nécessaire pour toutes ses instances. Avec les mêmes métapropriétés, un *rôle matériel* est porteur d'un critère d'identité et *anti-rigide* – *i.e.*

¹³Nous ne démarrons pas là une discussion sur la possibilité d'un être vivant sans corps physique. À ce jour et pour la médecine, cette hypothèse est sans objet.

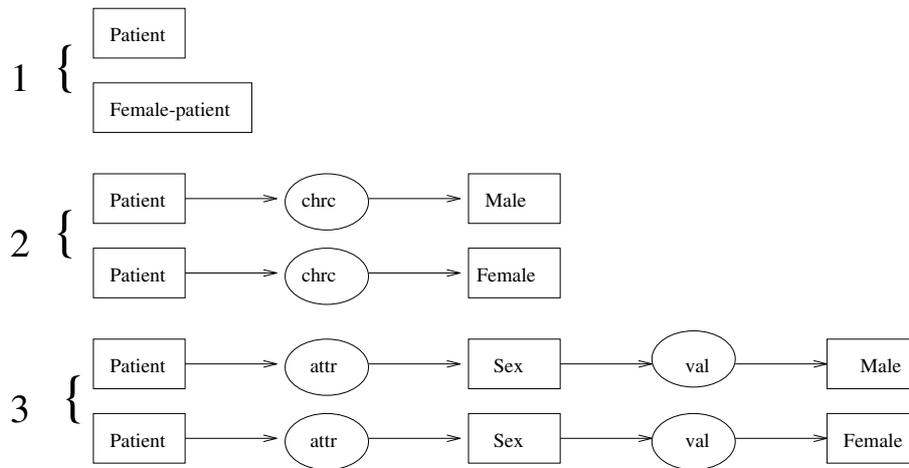


FIG. 4.4 – Trois représentations du genre sexuel en utilisant la représentation graphique des graphes conceptuels où les rectangles sont des concepts et les ovoïdes, des relations.

optionnel pour toutes ses instances. L'auteur donne d'abord l'exemple d'une *Personne* qui est porteuse d'un critère d'identité – *e.g.* l'empreinte digitale si c'est le critère choisi dans notre conceptualisation – et est *rigide* (on ne peut cesser d'être une personne) : *Personne* est un type. L'auteur donne ensuite l'exemple de *Étudiant* qui est porteur d'un critère d'identité et *anti-rigide* (on peut cesser d'être étudiant à tout moment) : *Étudiant* est un *rôle matériel*.

La question du *type* est alors intéressante pour les ontologies car si on suppose que chaque *type* a un critère d'identité différent, alors un *sous-type* spécialise un *type* en rajoutant un critère d'identité et la structure des *types* est obligatoirement un arbre. Ce sont, dit autrement, exactement les mêmes arguments que B. Bachimont (*cf.* § 2.2.2). Cette affirmation est faite à un niveau sémantique. Sa conséquence au niveau formel, en manipulant les métapropriétés indiquées, est la possibilité d'outils formels pour « nettoyer » une ontologie formelle, à commencer par lui constituer une « colonne vertébrale » de types arborescente¹⁴. Sa conséquence, pour nous, est de confirmer qu'une ontologie comporte d'abord un arbre de concepts primitifs.

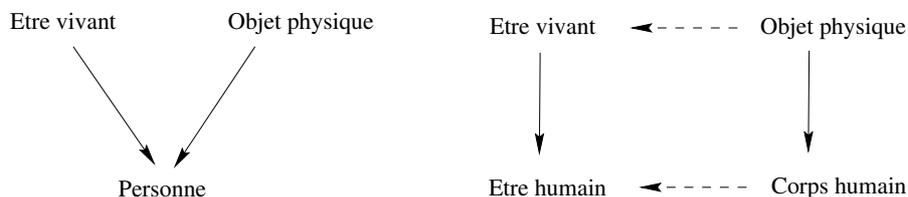


FIG. 4.5 – Le corps de la personne.

¹⁴La spécification des métapropriétés est détaillée dans (Welty & Guarino, 2001). Les liens entre engagement sémantique et engagement ontologique et leurs conséquences méthodologiques sont développés par B. Bachimont dans (Bachimont, 2001b). Enfin, nous n'entrons pas plus avant dans le débat entre B. Bachimont défenseur d'ontologies régionales dépendant du domaine et N. Guarino qui recherche une ontologie universelle même si le lecteur peut imaginer notre penchant, non universaliste.

3.2 Ontologies versus base de connaissances

Avec le but originel de la réutilisation, une affirmation voulait qu'une ontologie serait la conceptualisation du monde là où une base de connaissances contiendrait une ontologie et un certain nombre de connaissances du domaine plus complexes et fondées sur cette ontologie. En abandonnant cette question de la réutilisation, battue en brèche par de nombreux travaux, on peut dire qu'une ontologie recense ce qui existe et le définit par ses propriétés essentielles. Elle ne rapporte pas tout ce qui arrive. La base de connaissances, elle, utilise les descripteurs ainsi fournis par l'ontologie, pour énoncer tout ce qu'il faut savoir sur le domaine. Autrement dit, l'ontologie contient l'essentiel et le définitoire, la base de connaissances le propre et le contingent.

Dans l'exemple de MENELAS, l'ontologie est l'arbre des types primitifs et l'arbre des relations, elles aussi primitives. Les connaissances du domaines construites avec ces briques, sont des types définis, combinaisons de types et de relations primitifs, et des modèles encyclopédiques (cf. fig. 4.3) nécessaires au raisonnement. Il faut noter qu'il y a dans l'ontologie, les propriétés – en utilisant le vocabulaire de N. Guarino (cf. § 3.1.2) – comme rôle. En effet, si rôle ne peut être le type ontologique d'une personne (cf. § 3.1.2 et les caractéristiques de *Personne versus Rôle*), c'est un type pour lui même et, comme tous les objets primitifs du monde modélisé, il a sa place dans l'ontologie.

3.3 Ontologies versus thésaurus

Une modélisation de type ontologique existe depuis longtemps dans le domaine de la recherche d'information au sein des bibliothèques et dans le domaine de la terminologie. Nous allons essayer de caractériser un certain nombre de produits terminologiques au regard de leurs caractéristiques conceptuelles dans le but de savoir si ils peuvent être ou servir d'ontologies et à quelles conditions. Pour cela, nous allons reprendre ici, rapidement, trois définitions par rapport à des produits terminologiques existant parallèlement aux ontologies, les *thésaurus*, *classifications* et *terminologies*¹⁵. **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 pour l'analyse de contenu, le classement et donc l'indexation de documents d'information (dans de nombreux cas, les thésaurus proposent aussi une définition des termes utilisés). **Une classification** est l'action de distribuer par classes par catégories (rien n'est dit sur le type d'objets classifiés). **Une terminologie** est un ensemble des termes particuliers à une science, à un art, à un domaine. Les termes y sont également définis par un texte en langue naturelle et caractérisés par différentes propriétés linguistiques ou grammaticales suivant l'usage prévu de cette terminologie. Avec leur mise sur support informatique, les terminologies ont beaucoup évolué et sont parfois enrichies de relations entre termes, formant ainsi un réseau terminologique.

À partir de là, on va pouvoir s'intéresser successivement à WORDNET, au MESH, à la SNOMED et à UMLS et voir quelle est leur nature exacte, ce qu'on peut en faire et pourquoi.

WORDNET est une base de données lexicales. Les termes y sont organisés sous formes d'ensembles de synonymes, les *synsets*. Chaque *synset* étant alors un concept lexicalisé (en reprenant les termes de M. Slodzian (1999)). Ces concepts lexicalisés sont reliés par des relations conceptuelles (*is-a*, *has-a*). Les concepteurs de WORDNET affirment ainsi construire une ontologie linguistique. Si nous pensons qu'une ontologie a pour but de conceptualiser un domaine et nous le maintenons, alors le processus qui mène à cette ontologie doit clairement en prendre la direction. Ce n'est pas le cas

¹⁵Ces définitions seront complétées par d'autres quand nous aborderons la question de l'indexation, § 6.

de WORDNET qui jongle allègrement entre relations conceptuelles et termes¹⁶. A. Gangemi *et al.* (2001) proposent un travail intéressant en posant l'hypothèse que WORDNET est une ontologie et qu'il faut la « nettoyer ». Ils utilisent pour cela les métapropriétés formelles qu'ils ont définies (cf. § 3.1.2). La profondeur du « nettoyage » laisse penser que l'objet de départ n'était peut-être pas une ontologie telle que nous – mais Gangemi *et al.* aussi – la définissons, à savoir une conceptualisation organisée en vue d'aboutir à un objet formel. Ainsi, le mélange de relations *is-a* et *instance-of* au sein d'une même structure hyperonymique ou la confusion entre rôles et types dans les concepts sont, parmi d'autres, des sources d'erreur qui remettent en cause la validité de WORDNET en tant qu'ontologie (Gangemi *et al.*, 2001). Précisons bien la motivation de ces remarques : WORDNET est un énorme dictionnaire hypermédia de l'anglais-américain (plus de 100 000 *synsets*) et sa richesse et sa facilité d'accès en font un intéressant outil pour la recherche d'information ou d'autres tâches comme le traitement du langage naturel mais ce n'est pas une ontologie, cela n'en prend pas le chemin et essayer de l'utiliser tel quel ou avec un minimum de modification dans un système formel est voué à l'échec.

Le MeSH est un thésaurus médical. C'est le thésaurus d'indexation de la base bibliographique MEDLINE¹⁷. Il est traduit en français par l'INSERM et sert aussi de thésaurus au site CISMeF (cf. chap. 3, § 4). Le MeSH offre une organisation hiérarchique et associative et comprend jusqu'à neuf niveaux de profondeur. Les principaux composants du MeSH sont les *Headings* (MH pour *Main Headings* par la suite), les *Subheading* et les *Supplementary Concept Records*. Les MH respectent un certain nombre de propriétés (Nelson *et al.*, 2001) : d'abord, ils couvrent tout le champs de la médecine qu'on veut bien leur faire couvrir et ne se recoupent pas les uns les autres. Ils forment une partition du domaine. Le seul recouvrement accepté est celui de généralisation (*broader-than*) et de spécialisation (*Narrower-than*) mis en œuvre dans leur structure hiérarchique. Ces arbres proposent des hiérarchies selon plusieurs points de vue et partagent les même MH. Il est alors évident que ceux-ci ne peuvent être des concepts : ils représentent un ou plusieurs concepts et constituent des classes de descripteurs (appelés par la suite, simplement, descripteurs). On voit, figure 4.6, le descripteur « nez » impliqué dans trois hiérarchies de l'anatomie, une liée aux régions du corps [AO1], une autre au système respiratoire [A04] et une dernière aux organes sensitifs [A09]. Chaque descripteur a pour label un terme préférentiel pris parmi les termes préférentiels de chacun de ses concepts (dans notre exemple, le nez dans les trois hiérarchies différentes). Les relations de subsomption des hiérarchies sont principalement des hyperonymies et des partinomies mais, dans le domaine de la recherche d'informations, qui est celui du MeSH, on trouve des relations liées au sujet d'intérêt comme dans l'exemple de la hiérarchie des accidents qui subsume « prévention des accidents ». Alors qu'un certain nombre de propriétés du MeSH lui donnent un statut proche de celui des ontologies (objets conceptuels, partitions, arbres), les auteurs lui ont conservé une orientation recherche d'information, avec des liens de sujets d'intérêts ou des conceptualisations reflétant une vue de la littérature des usagers et pas une conceptualisation destinée à permettre des inférences médicales. Ce qui est affirmé dans le paragraphe suivant :

Many individuals have to tried to use MeSH as a concept representation language with only modest succes. That the relationships in the MeSH tree structure were designed with a different view, and with a different (an not formal) meaning of "broader-than", has frustated their efforts. The MeSH hierarchical structure was designed to reflect a view of the literature for a user.[. . .] The trees thus indicate what appears to be a useful set of relationships, based on the perceveid needs of searchers (Nelson et al., 2001).

On a encore une preuve, si besoin était, qu'une organisation conceptuelle dépend

¹⁶M. Slodzian (1999; 2000) développe très longuement des arguments épistémologiques et linguistiques à ce sujet. Nous renvoyons le lecteur intéressé à ces articles.

¹⁷Le MeSH est accessible à <http://www.nlm.nih.gov/mesh/meshome.html>

des buts pour lesquels elle a été faite et que la recherche d'informations ne nécessite pas les mêmes connaissances - précisément, la même formalisation de la médecine - que, par exemple, un SBC de reconnaissance du langage naturel.

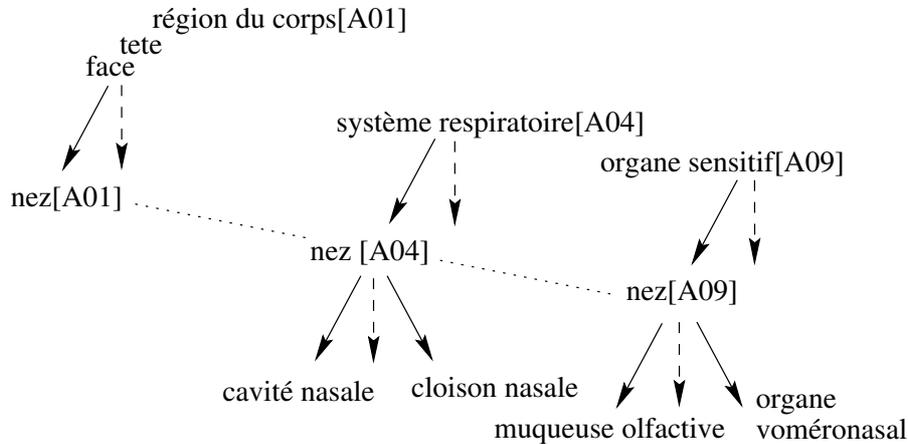


FIG. 4.6 – Le descripteur « nez » dans le MESH.

La SNOMED (pour *Systematized Nomenclature of Medicine*) est une classification multi-axiale (*topographie, morphologie, fonction, organismes vivants, médicaments/produits chimiques, agents physiques, etc.*) et multi-domaines (College of American Pathologists, 1993). À l'intérieur de chaque axe, les éléments sont organisés suivant une structure hiérarchique. La classification d'un terme repose sur une décomposition de celui-ci en combinaison de termes appartenant à différents axes. La SNOMED est une des classifications médicales les plus complètes mais un même concept peut y être décrit de différentes façons et rien n'empêche de créer par combinaison des concepts inconsistants (le Bozec, 2001). En raison de ce type de défauts, la SNOMED a évolué en SNOMED-RT (pour *Root Procedure*) puis en SNOMED-CT (pour *Clinical terms*), fusion de la SNOMED-RT et des *Clinical Terms* de la NHS (services de santé britanniques). Dans cette nouvelle configuration, la SNOMED-RT respecte un certain nombre de principes : structure hiérarchique de concepts, définitions de types ou de rôles pour des concepts, consistance, exploitation dans le cadre d'une logique de description (Dolin *et al.*, 2001) qui permettent de penser que c'est bien une ontologie formelle. La principale difficulté qui semble apparaître dans la transformation de la SNOMED en une ontologie est le choix qui a été fait de conserver les termes de la classification comme concepts et labels des concepts de l'ontologie (Spackman *et al.*, 2002) : la volonté de transformer la SNOMED en une réelle ontologie formelle se heurte à la « nécessité » que s'imposent les auteurs, de conserver la classification - presque - telle quelle avec les souplesses et les imprécisions d'un paradigme de construction pour partie linguistique. Notre point de vue est que, en l'état, cette transformation n'est pas totalement assumée, en particulier au niveau de la normalisation.

UMLS a été mis en place dans le but d'améliorer l'accès à l'information médicale à partir de sources diverses : bases de données bibliographiques, bases de données d'enregistrements cliniques et bases de connaissances médicales (Lindberg & Humphreys, 1990). Un des moyens de UMLS est alors de définir un vocabulaire médical de base, un « Metathesaurus » qui reprend et dédoublonne les termes de l'ensemble des ressources terminologiques qu'il inclut (MESH, SNOMED, etc.)¹⁸. Ce métathésaurus propose une description hiérarchique des connaissances médicales utilisées dans divers documents et SBC. De plus, un réseau sémantique de 134 types sémantiques environ permet de

¹⁸UMLS est accessible à <<http://www.nlm.nih.gov/research/umls/>>.

typer tous les termes du métathésaurus. L'intérêt d'UMLS réside dans sa grande couverture du domaine médical (plus de 870 000 « concepts » dans le métathésaurus à ce jour) et dans sa disponibilité. Le métathésaurus n'est pas une ontologie : il n'a pas été fait dans ce but et une tentative de réutilisation comme une ontologie (l'utilisation du métathésaurus pour construire l'ontologie de MENELAS) s'est soldée par un échec (Charlet *et al.*, 1996a). Le réseau sémantique a une structure beaucoup plus proche d'une ontologie mais ce n'est pas non plus son but : il recense 134 types plus ou moins généraux en médecine.

À la fin de ce panorama, on peut constater que peu parmi les produits terminologiques étudiés, en particulier en médecine, sont des ontologies. Ce n'est pas étonnant, les besoins de la médecine ayant d'abord été focalisés sur des problèmes de terminologie médicale, les produits construits avaient les caractéristiques correspondantes. Les ontologies n'étant pas sans rapport avec les terminologies, on peut trouver dans ces thésaurus, en particulier UMLS pour notre expérience mais pourquoi pas le Mesh, des ressources pour amorcer une ontologie. Mais il faut prendre garde au fait que ce ne sont que des « ressources pour » et pas des « embryons de ». Maintenant, par rapport à l'interopérabilité, la médecine a les mêmes besoins que les autres domaines, voire plus car ils sont jusqu'ici très mal satisfaits. Les ontologies sont une des clés de cette interopérabilité, encore faut-il qu'on les définisse, les construise et les utilise correctement. Après avoir développé jusque-là la question de leur élaboration, nous allons maintenant nous intéresser de plus près à leur utilité à l'éclairage des expériences décrites dans cette section.

3.4 Quelle utilité en médecine ?

3.4.1 Prologue

En prologue à la réponse posée ici, il nous semble important de rappeler que la médecine, comme d'autres, fait bien partie des domaines où une modélisation ontologique est nécessaire à partir d'un matériau non formel, en l'occurrence linguistique. Comme montré préalablement (*cf.* chap. 3, § 1.1), la médecine est une pratique quotidienne où les informations se consignent dans des « dossiers », regroupements cohérents de documents essentiellement textuels et n'est une science que pour ses principes biochimiques, physiques et ses protocoles d'études épidémiologiques. Cela n'a rien de péjoratif mais nous explique pourquoi le matériau n'est pas formel et nous annonce des difficultés pour le formaliser. Enfin, et c'est le début de la réponse à la question de l'utilité, si l'on veut utiliser l'informatique pour améliorer un certain nombre de processus de partage de connaissances, de décision, la formalisation est nécessaire.

3.4.2 Addenda à l'utilité

L'utilité des ontologies en médecine découle évidemment de l'utilité des ontologies telle que nous l'avons décrite au paragraphe 1.4 mais elle se spécialise en fonction du contexte propre à la médecine. Évacuons rapidement les points, évidents, que sont les connaissances du domaine d'un SBC ou la spécification d'un système d'information : les problèmes y sont identiques dans un domaine ou un autre, même s'ils ne sont pas aisés à résoudre, et la médecine ne fait pas exception à la règle. MENELAS (*cf.* § 2.4) est un exemple paradigmatique de cet usage. Pour les systèmes d'information, l'usage est identique en médecine par rapport à un cas général mais il faut noter que les systèmes d'information prenant en charge de l'information médicale – avec la quantité d'objets que devrait avoir un référentiel dans ce domaine – ne sont pas légion et doivent plutôt être considérés encore comme prospectifs (*cf.* chap. 6, § 4.2.1).

En revanche, Pour la question de la communication (qui englobe l'interopérabilité ici), le contexte médical est très spécifique : la médecine a, depuis longtemps, cherché à

normer la façon dont on parle de la pratique médicale, dont on l'écrit. Elle a de plus mis en œuvre depuis longtemps une politique d'indexation et de recherche d'information (RI). Elle a pour cela mis en place des thésaurus, souvent très grands, comme nous venons de le voir (cf. § 3.3). C'est vis-à-vis de l'existant et donc de ces thésaurus, que l'utilité des ontologies va apparaître en médecine, que ce soit pour « remplacer » les thésaurus ou être utilisées dans la même application (Rector, 1998; Dolin *et al.*, 2001) :

Communication entre êtres humains. Ici, les thésaurus sont utilisés pour que les êtres humains, en l'occurrence les praticiens, « parlent » la même langue pour décrire un patients, ses maladies, etc. Les ontologies doivent permettre de gérer les concepts utilisés, que ce soit au sein d'une écriture libre ou dans des interfaces intuitives.

Médiation (communication entre agents logiciel). Il est admis par les praticiens de l'informatique médicale qu'il est nécessaire d'avoir différents types de terminologies, que ce soit pour l'analyse de données cliniques, pour des travaux de traitement du langage naturel, pour des systèmes d'aide à la décision. Ici, les ontologies doivent permettre de créer un pont applicatif que les thésaurus et terminologies, non conceptuels, ne peuvent assurer.

Indexation et Recherche d'informations. Des thésaurus comme UMLS ou le MESH ont d'abord été développés pour l'indexation et la RI. Leur mise sous forme numérique, dans le même élan que la mise sur support numérique de plus en plus de ressources documentaires ou d'informations médicales (articles, etc.), a fait naître de nouveaux besoins en RI avec des requêtes plus complexes, demandant des inférences, elles aussi, plus complexes. Le remplacement de ces thésaurus par des ontologies, seules à même de servir de support à ces inférences, est envisagé en médecine, même si nous n'en sommes qu'au début des travaux de recherche dans ce domaine.

Élaboration et maintenance des terminologies. Les thésaurus médicaux sont tellement gros et complexes (cf. *supra*), 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, en particulier autour du projet GALEN : ce sont des *serveurs de terminologie*¹⁹.

3.4.3 Un serveur de terminologie, GALEN

Un serveur de terminologie est une plate-forme logicielle qui fournit, dans un domaine d'expertise donné, un certain nombre de services, a minima conceptuels, ensuite linguistiques. Ces services sont à destination d'un utilisateur ou d'un programme (requêtes normalisées). La délégation de ces services au serveur permet une prise en charge centralisée des problèmes terminologiques pour une communauté de travail ou un ensemble d'applications. Ces services s'appuient d'abord sur une ontologie du domaine et sont de différents types :

- Navigation dans l'ontologie,
- Édition/modification de l'ontologie (selon autorisation),
- Spécification et construction d'une expression conceptuelle complexe d'éléments de l'ontologie en fonction des signatures des relations (les contraintes de combinaison),
- Vérifications portant sur une expression conceptuelle (validité, canonicité, mise sous forme canonique),
- Distance sémantique entre concepts.

¹⁹Le projet MENELAS, décrivant les diagnostics de la CIM-10 en se servant de son ontologie, poursuivait, à une échelle plus restreinte, exactement les mêmes buts.

Ces services nécessitent la mise en œuvre d'outils de classification et d'inférences sur les ontologies. Les logiques de descriptions sont des candidats privilégiés dans ce contexte. Les services linguistiques sont rendus par la mise en relation des concepts (ontologies) avec des terminologies ou des thésaurus du domaine et permettent, dans un premier temps, la maintenance, la comparaison ou la fusion de thésaurus.

Le projet GALEN, développé à l'université de Manchester, vise à mettre en place un serveur de terminologie en médecine. Développé depuis une dizaine d'années (au sein de projets européens successifs), il est centré sur un *common reference model*, une ontologie de la médecine telle que nous l'entendons ici. Cette ontologie respecte une structure arborescente au niveau de ses types primitifs et est le cœur du système et des services qu'il propose (Rector, 1998). Nous n'allons pas décrire le système par le menu²⁰ et allons plutôt nous intéresser à son *common reference model* et à sa mise en œuvre pour préciser quelques points :

- L'ontologie de la médecine ne couvre que les domaines dans lesquels le projet s'est développé, où des opportunités se sont créées. Ainsi, le département de santé publique et d'informatique médicale de Saint-Étienne participe au développement de la classification commune des actes médicaux (CCAM) et justifie l'utilisation d'un tel outil pour le développement cohérent d'une terminologie médicale (Rodrigues *et al.*, 1998, 1999).
- Les promoteurs du projet sont confrontés au problème de l'évolution rapide de la médecine et donc des ontologies attenantes, au point qu'il est difficile d'avoir une vue complète et cohérente d'une ontologie d'un domaine précis²¹.
- Confronté à la difficulté de compréhension des ontologies, le projet GALEN y a répondu, même si ça n'était pas au début le but d'un tel module, par un générateur de langage naturel permettant de valider les représentations proposées avec les praticiens. Cela rejoint les problématiques d'indexation (*cf.* § 6).
- Enfin, les promoteurs du projet, s'ils proposent de construire – et construisent – des ontologies, ne proposent pas réellement de méthode argumentée et constructive. C'est ce que nous avons fait et allons compléter maintenant, en enrichissant la méthodologie décrite au paragraphe 2.2 de considérations sur les corpus et les outils d'analyse de corpus.

4 Acquérir des ontologies à partir de corpus

4.1 Introduction épistémologique

Après avoir proposé que le corpus soit le point de départ d'une méthodologie de construction d'une ontologie, il est maintenant nécessaire de développer aussi une méthodologie d'analyse de ce corpus et de son utilisation dans le contexte de cette construction. Cette méthodologie (et d'autres) se fonde sur les travaux du groupe TIA (*cf.* note 12, p. 13) et est en accord avec la première étape de la méthodologie de construction d'une ontologie présentée dans ce chapitre (*cf.* § 2.2.1). Les présupposés terminologiques sont des hypothèses fortes pour la communauté du même nom et nous en reprenons rapidement à notre compte deux points saillants (Bourigault & Slodzian, 2000) :

- L'hypothèse selon laquelle l'expert d'un domaine est dépositaire d'un système conceptuel qu'il suffit de mettre à jour est non productive. La tâche d'analyse terminologique vise alors la construction d'une description des structures lexicales à

²⁰Nous renvoyons le lecteur intéressé à, par exemple, (Rogers *et al.*, 2001).

²¹De ce point de vue là, le projet MENELAS avait contourné le problème avec des travaux étalés sur la stricte durée du projet avec des corpus d'apprentissage et d'expérimentation rapidement fixés. Face à de nouveaux corpus, nous ne manquerions pas d'être confrontés, plus de six ans plus tard, à de nouveaux concepts médicaux ou de nouvelles manières de décrire les pathologies coronariennes.

l'œuvre dans le corpus textuel à partir d'une analyse réglée de ce corpus.

- Le corpus doit être soigneusement choisi en fonction du domaine et de l'application visée (la tâche). La taille des corpus et la masse d'informations contenues impliquent l'utilisation d'outils de terminologie textuelle (concordancier, extracteur de candidats termes, extracteur de relations, etc.). L'utilisation de ces outils doit être spécifiée au sein d'une démarche méthodologique claire précisant à quel stade du processus et selon quelles modalités il convient de les utiliser.

Dans ce contexte, un certain nombre d'outils de traitement automatique des langues (TAL) qui proposent une palette de modes d'analyse de corpus, ont été développés²². Nous allons nous attarder maintenant sur l'un d'entre eux, utilisé dans nos développements d'ontologies médicales.

4.2 L'usage d'un outil de TAL

Cet outil est en fait l'enchaînement des traitements de deux outils, un analyseur syntaxique de corpus spécialisés, SYNTAX, et un module d'analyse distributionnelle, UPERY, tous deux développés par D. Bourigault (Bourigault & Fabre, 2000; Bourigault, 2002)²³.

Le résultat de l'analyse syntaxique est un réseau de mots et de syntagmes : un syntagme verbal (resp. nominal, adjectival) est un groupe de mots dont la tête syntaxique est un verbe (resp. nom, adjectif). Par exemple, *révéler une lésion osseuse* est un syntagme verbal dont la tête syntaxique est le verbe *révéler* et l'expansion le syntagme nominal *lésion osseuse*. Ce dernier syntagme a pour tête syntaxique le nom *lésion* et pour expansion l'adjectif *osseuse*. Dans le réseau construit, dit « réseau terminologique », chaque syntagme est relié d'une part à sa tête (lien T) et d'autre part à son expansion (lien E). Le lien E est étiqueté par le nom de la relation syntaxique de dépendance. Les éléments du réseau (mots et syntagmes) sont appelés « candidats termes ».

Le module d'analyse distributionnelle exploite l'ensemble des données présentes dans le réseau pour effectuer un calcul des proximités distributionnelles entre les mots et syntagmes du réseau. Ce calcul s'effectue sur la base des contextes syntaxiques partagés. Il s'agit d'une mise en œuvre du principe de l'analyse distributionnelle de Harris (1989) que prolongent les travaux de Assadi et Bourigault (2000). Les données de l'analyse sont constituées ainsi : chaque lien dans le réseau de dépendance produit une information élémentaire pour l'analyse distributionnelle ; un lien (Tête, Relation, Expansion) se formalise sous la forme d'un couple (contexte, terme) où le contexte est constitué lui-même par la concaténation de la Tête et de la Relation de dépendance et où le terme est l'expansion. On dispose, pour un contexte donné, de l'ensemble des termes (mots ou syntagmes) qui y apparaissent, et, pour un terme donné, de l'ensemble des contextes dans lesquels il apparaît²⁴.

La méthodologie exacte d'utilisations de tels résultats est encore en devenir et elle s'affine au fur et à mesure des expériences qui sont menées avec ces outils. Nous allons nous intéresser à l'une d'elle qui a permis de mettre au point une méthodologie et de construire effectivement une ontologie de la réanimation chirurgicale.

4.3 Une ontologie en réanimation chirurgicale

Le contexte de ce travail est la réanimation chirurgicale qui est un domaine médical spécialisé dans la prise en charge des complications postopératoires et dans la tra-

²²Nous renvoyons le lecteur à (Aussenac-Gilles *et al.*, 2002), pour une description et un typage de ces outils.

²³Plus précisément, SYNTAX nécessite d'avoir utilisé en amont un étiqueteur qui étiquette les mots d'une phrase avec leur rôle grammatical (nom, adjectif, verbe, etc.)

²⁴La description faite ici est tirée de (Le Moigno *et al.*, 2002) et a été écrite par D. Bourigault. On trouve aussi dans (Bouaud *et al.*, 2000) d'intéressants travaux sur l'usage de l'analyse distributionnelle pour la construction d'ontologies.

matologie. Dans cette spécialité, les pathologies et les actes à prendre en compte sont divers (en fonction de la chirurgie réalisée, du type de complication, etc.). Pour leur activité de codage obligatoire²⁵, les praticiens s'aident d'un thésaurus de spécialité qui a été élaboré de façon à ce que les séjours de réanimation soient le mieux possible valorisés. Mais il est reconnu que l'ambiguïté du thésaurus est une source d'erreurs et de variabilité et ce pour de nombreuses raisons : omissions de la part des codeurs qui ne sont pas d'accord avec les items proposés, manque de données dans le compte rendu d'hospitalisation (CRH), durée du séjour. Une ontologie permettant d'organiser les concepts du thésaurus s'impose alors comme nous l'avons déjà évoqué précédemment (*cf. supra*).

Pour ce qui est de la méthode de travail, elle a consisté en l'exploitation des sorties de UPERY par l'expert non informaticien et non linguiste. Dans un premier temps, il a analysé les couples de termes (nominatifs, adjectivaux) et les couples de contextes (verbaux, nominatifs) qui présentaient des proximités suffisamment élevées pour dégager de grandes classes de concepts. Dans un second temps, la classification des concepts est raffinée par un retour à l'analyse du réseau terminologique en Tête et Expansion. Nous renvoyons à (Le Moigno *et al.*, 2002) pour une description *in extenso* de la méthode proposée.

L'ontologie construite est constituée d'une hiérarchie *is-a* de concepts et d'une hiérarchie *is-a* de relations. La hiérarchie *is-a* de classes est initiée par 2 branches principales : *objet-reel* et *objet-abstrait*. La branche *objet-reel* comprend essentiellement les éléments anatomiques et organiques, les instruments, les produits thérapeutiques et enfin les objets vivants animés. La branche *objet-abstrait* est décomposée en deux branches principales : *objet-description* (*etat*, *attribut*, *objet-fonction*) et *objet-action*. Dans la classe *etat*, on retrouve tous les états pathologiques (maladie, syndrome, lésion et symptômes (cliniques et paracliniques)). La classe *objet-fonction* décrit les fonctions physiques physiologiques et les dysfonctions. La classe *objet-action* comprend tous les actes et les processus pathologiques (on peut assimiler ces derniers à des mécanismes physiopathologiques). La hiérarchie *is-a* des relations comprend 3 branches principales : *objet*, *relation-etat* et *relation-action*. Par choix, peu d'adjectifs sont représentés dans cette ontologie. Les adjectifs sont remplacés par l'adjonction d'une relation au nom se rapportant à cet adjectif (ex cardiaque est remplacé par « localisé au-niv-de cœur »). La majorité des concepts sont construits sur ce principe. L'ontologie est visible à l'URL <<http://www.biomath.jussieu.fr>>.

La validation de l'ontologie se fait ici en testant sa couverture par rapport au thésaurus de la spécialité (découpé en chapitres). Cette couverture est évaluée en vérifiant la possibilité de construire une représentation conceptuelle d'un ensemble de pathologies du thésaurus en utilisant les briques conceptuelles fournies par l'ontologie – en fait, l'ontologie des concepts et l'ontologie des relations. Sur les 658 libellés de pathologies qui composent le thésaurus, 400 (60%) sont couverts par l'ontologie. Dans le domaine choisi de la traumatologie, le niveau de couverture est de 100%. Les concepts de ce chapitre sont simples dans la mesure où la plupart se contentent de relier une lésion ou un processus à une localisation anatomique (fracture du col du fémur). D'autres plus complexes, comme ceux représentés figure 4.7, sont également représentés par l'ontologie. L'ontologie couvre 100% des 100 libellés de pathologie exprimés dans ce chapitre.

Les résultats présentés ici de façon rapide sont approfondis et nuancés dans (Le Moigno *et al.*, 2002). On peut dire que ce travail est une validation des outils de TAL utilisés dans la mesure où l'expert a pu être autonome et travailler dans un temps décent (moins de 200 heures) et que l'ontologie est valable au regard de la couverture du thésaurus, critère qui était le nôtre²⁶. Pour terminer la description de ce travail on peut

²⁵ Le Programme de médicalisation du système d'information (PMSI), vise à introduire des concepts de comptabilité analytique dans la gestion administrative des hôpitaux : les diagnostics et actes effectués dans un établissement de santé sont codés et comptabilisés, rapportés à un patient et aux différents coûts de la structure. Cela permet ainsi de bâtir des indices de coûts relatifs par groupe homogène de malades. Le PMSI utilise un système de codage international, la CIM-10.

Fracture de la base du crâne

```
[LESION : Fracture]
  (LOCALISATION : localisé_au_niv_de) [SITUATION_DANS_OBJET : base]
  (OBJET : du) [OS: crâne]
```

Hémorragie extra-durale traumatique

```
[PROCESSUS_PATHOLOGIQUE_INTERNE : hémorragie]
  (LOCALISATION : à_l'extérieur_de) [ANATOMIE_TISSU_ENVELOPPE : duremère]
  (ETIOLOGIE : est_causé_par) [PROCESSUS_PATHOLOGIQUE_EXTERNE : traumatisme]
```

Lésion traumatique intracrânienne avec coma prolongé

```
[LESION]
  (ETIOLOGIE : est_causé_par) [PROCESSUS_PATHOLOGIQUE_EXTERNE : traumatisme]
  (LOCALISATION : à_l'intérieur_de) [OS : crâne]
  (EXPRESSION : se_manifeste_par) [SIGNE_OBJECTIF : coma]
  (REL_EVOLUTION : évoluée_de_façon) [VALEUR_NIV_EVOLUTION : prolongé]
```

Hémopéritoine

```
[LESION : épanchement hématisque]
  (LOCALISATION : localisé_au_niv_de) [ANATOMIE_TISSU_ENVELOPPE : péritoine]
```

FIG. 4.7 – Quelques exemples de couverture du thésaurus de spécialité

noter que ce n'est pas réellement une validation des autres étapes de la méthodologie dans la mesure où l'expert n'a pas spécifiquement cherché à la suivre (et ce travail de « normalisation » reste à faire ou, *a minima*, à vérifier, en sachant que l'expert l'a quand même abordé de manière assez intuitive). Enfin, on peut voir que l'opérationnalisation d'un tel travail se ferait dans le contexte des serveurs de terminologie (*cf.* § 3.4.3) même s'il n'a pas été développé pour un serveur opérationnel à ce jour.

5 De la réutilisation aux ontologies génériques

5.1 La réutilisabilité des ontologies

Nous avons abordé la question de la réutilisabilité des ontologies à travers les débats qui agitaient la communauté Ingénierie des connaissances (*cf.* chap. 1, § 4.2), et l'avons expérimentée durant le projet MENELAS (*cf.* § 2.4). Une non réutilisabilité affirmée des ontologies en découle, corroborée par la description de la méthode de B. Bachimont et des travaux épistémologiques qui la supportent (Bachimont, 1996). En effet, les ontologies sont des artefacts construits en fonction d'une tâche précise et ne peuvent être réutilisées, en tant qu'objet formel, pour une autre tâche. Des « ontologies » de sens commun comme WORDNET ou CYC (Guha & Lenat, 1990) ont encore moins de chance d'être réutilisables dans la mesure où elles veulent rendre compte de tous les sens des mots et ne normalisent pas, passage pourtant obligé vers le formel.

²⁶Il est important de noter que, dans des développements méthodologiques en plein devenir, chaque expérience est particulière. Celle-ci l'est en raison d'un domaine – et, corollairement, des capacités d'un expert – où l'organisation de thésaurus est une vieille et nécessaire tradition. On lira dans Bourigault *et al.* (2003) une comparaison entre le travail de développement de cette ontologie en réanimation chirurgicale et la réalisation de ressources ontologiques dans d'autres domaines (industriel, droit). On constate que l'apparente facilité avec laquelle les questions de normalisation ont été traitées ici ne sont pas de mise dans un domaine industriel où l'ontologie doit être construite *ex nihilo* – à partir de corpus évidemment – sans qu'une tradition et des classifications disponibles ne viennent aider les ingénieurs cogniticiens. Le processus est alors obligatoirement beaucoup plus long et itératif.

5.2 Des ontologies génériques

Depuis quelques temps et comme conséquence de cette non réutilisabilité, des chercheurs travaillent sur la constitution d'un « haut » d'ontologie, une *ontologie générique*, le nom fournissant le but. C'est le cas du groupe SUO (*Standard Upper Ontology* – <<http://suo.ieee.org>>) qui réfléchit à la constitution d'un haut d'ontologie, *the SUMO (Suggested Upper Merged Ontology)*, qui se voudrait universel pour les grandes catégories d'objets et de pensées. Les discussions intenses au sein de ce groupe de réflexion semblent montrer que le but ne sera pas atteint tout de suite. Mais sera-t-il atteint un jour ? Avec les mêmes arguments épistémologiques que précédemment sur la non réutilisabilité des ontologie, on peut penser qu'on ne sera pas capable de construire ce type d'ontologie universelle.

La question que nous nous posons en médecine est de savoir si, en se restreignant à la seule médecine, nous ne serions pas capable de construire une ontologie générique pour une grande classe d'applications comportant toutes les applications liées à des problématiques de la langue, des thésaurus, de la RI, etc. Cela revient à poser l'hypothèse que les grandes catégories des objets médicaux et du raisonnement médical traversent toute la pratique médicale. C'est ce que nous avons commencé à tester en démarrant un travail qui vise à préciser et valider le haut de l'ontologie de MENELAS pour vérifier ensuite la possibilité qu'elle soit le haut d'une ontologie de la périnatalité en développement au sein d'un réseau de santé (Weis, 2002) et qu'elle soit aussi le haut de l'ontologie de la réanimation développée précédemment.

6 L'indexation

6.1 Introduction

Sans étudier ici la problématique de l'indexation dans son ensemble, ce qui n'est pas notre propos, nous nous intéressons à l'élaboration des index en général, en raison de leur nature par rapport aux ontologies. Nous complétons d'abord les définitions données au § 3.3, avec celles de *l'indexation* et de *l'index*. **L'indexation** est un processus destiné à représenter par les éléments d'un langage souvent naturel des données résultant de l'analyse du contenu d'un document ou d'une question. On désigne également ainsi le résultat de cette opération (voir *Index*). **Un index** est une table alphabétique des mots, des termes correspondant aux sujets traités, des noms cités dans un livre. Dans le domaine technique, l'index d'un document est alors son analyse sommaire présentée sous forme de mots-clés, rubriques, etc.²⁷

Ces définitions étant posées, il faut noter deux choses importantes par rapport à l'indexation : c'est une pratique ancienne qui n'a pas attendu l'Intelligence artificielle ou l'Ingénierie des connaissances pour produire des index. En revanche, les pratiques terminologiques en général, et l'indexation en particulier, sont révolutionnées par le numérique et les outils que les communautés du traitement du langage naturel ou de l'Ingénierie des connaissances mettent à leur disposition. Ainsi et pour commencer, des outils de TAL, les mêmes qui permettent d'amorcer des ontologies (cf. § 4.2) permettent de construire de façon semi-automatique des index (cf. § 6.2).

Traditionnellement, une tâche d'indexation consiste à demander, à l'auteur d'un article, quels sont les mots qu'il veut voir servir d'entrées pour indexer son ouvrage. Un

²⁷Nous reprenons pour cela des définitions tirées du dictionnaire Le Robert, de l'AFNOR (AFNOR, 1987) de J. Chaumier (1988) et des travaux menés au sein des l'Action spécifique CNRS sur les corpus, ASSTIC-COT <<http://www.irit.fr/ASSTICCOT/>>. Par ailleurs, on trouve au Canada, à l'Université du Québec à Montréal (UQÀM), au service des bibliothèques, <<http://www.bibliotheques.uqam.ca/bibliotheques/index.html>> (« services techniques »/« analyse documentaire »), d'intéressants documents sur les politiques d'indexation et la gestion des vocabulaires qui ont servi de bases à notre positionnement par rapport aux tâches d'indexation.

des défauts de cette approche est la difficulté à bien repérer les mots dans le texte pour décider lesquels vont être retenus. En utilisant un traitement de texte, on se « facilite » la vie car on peut rechercher les occurrences de ce mot directement. En revanche, on doit avoir en mémoire les formes fléchies de ce mot car un traitement de texte ne lemmatise pas (encore) les mots. En utilisant un outil comme SYNTEX²⁸, le texte est analysé du point de vue morphologique et syntaxique et, si les phrases sont bien formées, on repère toutes les formes fléchies d'un même terme. De plus, on construit un réseau de candidats termes que l'on peut proposer à l'auteur qui a ainsi à sa disposition des listes de syntagmes nominaux (par exemple) et peut visualiser les liens entre syntagmes comme les phrases dans lesquelles ils sont utilisés. À partir de là, on peut envisager la construction de différentes ressources terminologiques servant d'index selon qu'elles sont plus ou moins conceptuelles et donc normalisent plus ou moins les termes utilisés pour l'index – *i.e.* les *descripteurs*.

6.2 Une question de normalisation

La décision de normaliser plus ou moins va dépendre du but visé et, en particulier, du document que l'on veut indexer. Si l'on cherche, par exemple, à indexer des ouvrages techniques, on va construire un index normalisant fortement les descripteurs en construisant une ontologie régionale (*cf.* § 2.2.2). Cela permettra alors de présenter à l'utilisateur une liste de notions abordées dans le texte ou, entre autres, de résoudre les polysémies en dégroupant les entrées ambiguës – *e.g.* réseau, électrique versus informatique. C'est ce qui a été réalisé à l'EDF dans la cadre de la conception et la réalisation de systèmes hypertextuels de documentation technique (SCDT – Gros & Assadi, 1998). En construisant une ontologie comme index, on est dans le contexte d'un thésaurus fortement normalisé et on retrouve, comme pour le thésaurus MESH (*cf.* § 3.3), le fait que, thésaurus ou ontologie, dès qu'il y a normalisation, l'artefact est à construire pour le but visé et exclusivement celui-ci.

Dans une position moins normalisatrice et pour un ouvrage (Charlet *et al.*, 2000b) sur l'Ingénierie des connaissances que nous avons coordonné en collaboration, nous avons construit un index permettant d'accéder aux 35 articles le constituant. Nous nous étions imposé de travailler en « vraie grandeur », c'est-à-dire de produire un index acceptable dans un délai de quelques semaines, à partir des résultats de l'extraction terminologique effectuée par SYNTEX²⁹ sur le corpus des 35 articles (150 000 mots). L'index final comporte environ 570 entrées et sous-entrées et 2 000 renvois. Ces contraintes matérielles assumées jusqu'à l'achèvement de l'expérience nous ont permis de développer une ébauche de méthodologie et de faire un inventaire des problèmes rencontrés (Bourigault & Charlet, 1999). Citons principalement :

Le choix des descripteurs. En dehors de la difficulté inhérente au choix d'un descripteur, nous laissons passer des candidats descripteurs intéressants et ce pour deux raisons : premièrement, l'analyse, très large, fournit un très grand nombre de candidats termes parmi lesquels des heuristiques de choix fondés sur la fréquence aident à faire le tri. Or, nous savons que ce n'est pas parce qu'un terme est peu fréquent qu'il ne peut être un bon descripteur. Au contraire. Secondement, le corpus est analysé de façon homogène alors que la structure logique des articles permettrait de repérer des zones de texte (résumé, introduction, titre, etc.) sûrement plus « riches » en descripteurs.

La normalisation. Ce n'est pas un problème mais il vaut d'être noté : nous avons fait le choix de ne pas normaliser les descripteurs et de rester le plus neutre possible avec, dans l'index, des termes attestés en corpus pour fournir, dans un premier

²⁸Il y a d'autres outils mais celui-ci nous a servi pour des expérimentations passées et présentes et est paradigmatique de notre approche.

²⁹En fait son prédécesseur, développé aussi par D. Bourigault, LEXTER.

temps, une image relativement fidèle de ce dont « parlent » les textes. Nous préparions un index pour ouvrage de type scientifique, dont les utilisateurs potentiels ont le profil chercheur. Il ne s'agit pas d'imposer une vue sur le domaine qui nous serait trop personnelle.

La validation. Il y a en général peu ou pas de validation des index construits. Dans notre cas, une validation a été effectuée *a posteriori*, l'ouvrage étant déjà sorti. Cette validation s'est effectuée dans le cadre du projet TH(1C)² et, si elle n'a pas eu d'influence sur l'index, elle a permis de pointer un certain nombre de choix et de difficultés de l'approche comme de la validation des index en général (Aussenac-Gilles & Bourigault, 2000).

Inversement, dans une position très normalisatrice, le projet (KA)² visait à réaliser des ontologies de la recherche en acquisition des connaissances (Benjamins V. R., 1998). Il était (et reste) intéressant sur la question du processus d'élaboration des ontologies par un groupe de personnes : se voulant consensuel et un travail de la communauté KA (*Knowledge Acquisition*) sur elle-même, il a fait appel à des communautés de chercheurs pour qu'ils indexent leur recherche et participent ainsi à la construction des ontologies. Nous y avons nous-même participé et avons pu constater la difficulté à réaliser un travail collaboratif dans ce but.

Pour terminer, provisoirement, sur cette question de la normalisation, notons le travail de N. Aussenac-Gilles et A. Condamines (1998) qui s'interrogent sur les apports potentiels des bases de connaissances terminologiques (BCT) à la consultation documentaire. Une BCT est attachée à un corpus donné. Elle est le résultat d'un processus d'analyse linguistique méthodique de ce corpus qui conduit à identifier un réseau de concepts et un ensemble de termes reliés d'une part à ces concepts et d'autre part aux textes, le tout constituant une modélisation (aussi) objective (que possible) du corpus. Dans leur réflexion sur l'utilisation d'une BCT comme point de départ pour construire un index, les auteurs expriment de façon claire les points communs et surtout les différences entre index et thésaurus : un index permet d'accéder au contenu d'un texte, il est construit en référence à ce texte ; un thésaurus est utilisé pour indexer et retrouver des documents dans une collection de documents portant sur un domaine, il est le plus souvent construit par introspection et consultation d'experts de ce domaine (Bourigault & Charlet, 1999).

6.3 Difficultés et perspectives

Les différents travaux sur l'indexation décrits jusqu'ici, nous permettent de pointer un certain nombre de difficultés et de perspectives de solutions qui se développent dans les communautés de recherche (principalement Ingénierie des connaissances, RI et Web sémantique). De notre point de vue, trois sont spécifiquement intéressants :

L'articulation terminologie/ontologie. Dans un article récent, C. Roussey *et al.* (2002) séparent explicitement une terminologie du domaine de sa conceptualisation et définissent le thésaurus sémantique comme *une normalisation des notions du domaine auxquelles sont associées des terminologies*. Le point essentiel pour nous ici est la similitude d'approche – l'identité même – avec les serveurs de terminologie de la médecine dans lesquels le couple thésaurus/ontologie joue le même rôle que le thésaurus sémantique (*cf.* § 3.4.3). C'est, dépassant le champ de la médecine, un type d'approche qui nous semble extrêmement prometteur, séparant explicitement les conceptualisations des façons d'en parler, faisant ainsi entrer – de notre avis par la bonne porte – les ontologies dans la problématique des index.

Rejoignant les attendus du projet MENELAS pour le multilinguisme (*cf.* § 2.4), les auteurs abordent ce problème avec le même type d'outil conceptuel – *i.e.* la séparation du niveau conceptuel et du ou des niveaux terminologiques –, notant qu'à langue différente, conceptualisation du monde différente (Roussey *et al.*, 2002). Il

n'est alors pas inutile de noter que l'ontologie du domaine devient l'objet central de négociation sur lequel il faut s'entendre, y compris de façon transnationale, et qu'elle est l'origine et l'aboutissement de cette même négociation.

L'index et les requêtes. Après avoir construit un index, il faut pouvoir s'en servir pour faire des requêtes, ce qui sous-entend qu'on maîtrise son contenu. Avec la taille que peuvent avoir ces index (cf. les thésaurus médicaux ou autres), cette maîtrise n'est plus du tout évidente. Des approches récentes profitent du réseau terminologique que constituent les index (cf. § 4) pour faire de l'expansion de requêtes en fonction des proximités des termes (Bourigault & Lame, 2002). D'autres approches utilisent un index décrit dans un langage normalisé, les *Topic Maps*, pour proposer des analyses conceptuelles à base de treillis de Galois et des outils de visualisation interactive pour naviguer dans l'index ainsi décrit (Le Grand, 2001). Ces deux approches partagent le fait de ne pas pousser la conceptualisation de l'index jusqu'à une ontologie même si elles l'envisagent toutes deux dans le futur.

Vers des index plus riches. Enfin, dans tous les cas de figure, les index grossissent, proposent ou vont proposer des fonctionnalités plus riches (Aït El Mekki & Nazarenko, 2001). Quittant l'index « papier » très linéaire où les seules indirections que l'on s'autorise sont celles que l'on peut appréhender sur la feuille, on trouve aujourd'hui des index hypertextuels proposant, à partir d'un descripteur, des liens vers des informations diverses telles que les variantes orthographiques, la polysémie, l'hyponymie, la méronymie, la conceptualisation, etc., toutes les informations d'ordre linguistique ou conceptuel qui permettent de mieux appréhender et utiliser les descripteurs.

En collaboration avec T. Aït El Mekki, A. Nazarenko (2001) et D. Bourigault, nous avons initialisé un travail sur un index électronique permettant de naviguer à travers les articles d'un livre succédant à (Charlet *et al.*, 2000b). Le but, prudent et pourtant source de nombreuses difficultés et réflexions, est de développer sur support numérique un index enrichissant quelque peu les fonctionnalités de l'index « papier ». Le futur est évidemment à imaginer dans un large développement des fonctionnalités proposées au paragraphe précédent par nos collègues.

6.4 Indexer la médecine

La médecine rencontre les mêmes problèmes que les autres disciplines et les solutions présentes et futures ne sont pas fondamentalement différentes. Nous allons nous intéresser à deux approches paradigmatiques, que nous avons déjà abordées, les projets CISMef et GALEN :

Indexer à partir d'un thésaurus : CISMef. Comme nous l'avons déjà vu, le thésaurus d'indexation de CISMef est le Mesh (cf. chap. 3, § 4). Ce thésaurus correspond à une organisation conceptuelle du domaine très précise et à visée d'indexation. Mais, développé dans un contexte professionnel, il ne satisfait plus les desideratas des concepteurs de CISMef (Soualmia *et al.*, 2002). Ainsi, il leur a d'abord fallu rajouter des descripteurs des spécialités biologiques ou médicales comme la chirurgie. Ensuite, dans le but de créer un CISMef-patient, il a été nécessaire de mettre en place une démarche spécifique d'indexation dont les points principaux sont : (a) création de descripteurs accessibles au grand public – *e.g. manie* plutôt que *trouble bipolaire* –, (b) organisation de vues simplifiées de l'arborescence du thésaurus Mesh et (c) indexation de ressources spécifiques compréhensibles par le grand public. Sans entrer dans les détails de mise en œuvre, il est intéressant de noter que ce travail s'est effectué en créant des liens sur le même thésaurus d'origine et que le travail d'indexation des ressources n'a pas été perdu.

On a ici un exemple intéressant d'utilisation d'une ressource, le Mesh, qui évolue doucement et qu'il faut augmenter d'un certain nombre de descripteurs sans tou-

cher à la structure centrale qui est sa force – on indexe la médecine avec le Mesh depuis des années – et sa faiblesse.

Indexer à partir d'une ontologie : GALEN. Vu les caractéristiques du serveur de terminologie, GALEN, rapprochées des problématiques d'indexation, il était logique d'essayer de l'utiliser comme *thésaurus sémantique* pour indexer des documents. C'est la recherche qui a été menée par D. Karlsson (2001) dans le domaine des infections urinaires. Pour ce faire, et toujours en raison de la complexité des ontologies, il a été développé une ontologie des infections urinaires et une liste de termes reliés aux concepts de l'ontologie. Les documents sont indexés en fonction des *phénomènes* qu'ils décrivent (un concept de l'ontologie de GALEN) et forment une hiérarchie dirigée par les hiérarchies des *phénomènes*. Une requête se fait à partir des termes. L'identification des documents pertinents se fait alors (1) à partir des concepts reliés aux termes de la requête et qui décrivent un document puis (2) se poursuit sur les concepts subsumés par ceux de la requête (*cf.* (Karlsson, 2001) pour une description plus complète et précise de l'algorithme).

Comme noté précédemment, cette expérience montre la nécessité de passer par une terminologie pour accéder aux concepts des ontologies dans une phase de recherche d'information. Un autre problème, important et noté par les auteurs, est que si l'on recherche un document lié à un organe (respectivement une pathologie), il peut être pertinent de récupérer les documents liés à la fonction de cet organe (respectivement aux signes cliniques de cette pathologie). Ces liens peuvent être des liens d'association qui ne sont pas ceux de l'organisation conceptuelle de l'ontologie (*cf.* les liens du Mesh, § 3.3). On retrouve la question de la motivation de la construction de l'ontologie et la question de savoir si cette construction peut poursuivre plusieurs buts.

7 Des ontologies aux textes : le web sémantique

À la suite du chapitre 3, paragraphe 5, nous nous intéressons aux problèmes abordés par les recherches sur le Web sémantique du point de vue des ontologies et de l'indexation et les questions abordées dans ce chapitre sont importantes à trois niveaux :

Les langages de description et de traitement des ontologies. Le Web sémantique est historiquement une réflexion sur les langages, en particulier ceux qui permettent de représenter les ontologies et de faire des inférences dessus. Les travaux convergent vers le langage OWL (*Ontology Web Language*) qui résulte de la convergence de projets européens et américains. Ce langage est, à ce jour, construit sur RDFS (*RDF Schema*), lui-même construit sur XML. Le système d'inférences utilisant ces ontologies est une logique de description comme FACT (Horrocks, 1998).

La construction et la maintenance d'ontologies. Le Web sémantique recherchant la possibilité que des agents logiciels puissent interagir, les ontologies et les recherches sur icelles sont indispensables (1) comme spécification des connaissances de SBC, les agents logiciels devant d'abord agir avant d'interagir et (2) comme spécification formelle des connaissances qui seront partagées par les agents logiciels. Dans un soucis d'interopérabilité, ces ontologies devront évoluer, fusionner, finalement être gérées comme des objets évolutifs avec leurs différentes versions.

L'appréhension des index et l'indexation. D'abord pensé pour la coopération entre agents logiciels (*cf.* § 5), le Web sémantique est aussi, évidemment, fait pour des utilisateurs humains. Ces mêmes utilisateurs feront probablement que le Web sémantique sera d'abord fournisseur de ressources qui seront opérationnalisés, interprétés, par eux-mêmes, avant une interopérabilité entre agents logiciels plus

longue à mettre en œuvre (Laublet *et al.*, 2002). Dans ce contexte, l'accès aux ressources, en particuliers les index, ontologies ou pas, et leur utilisation doivent d'abord être pensées pour des utilisateurs humains. C'est l'objet des travaux décrits précédemment (*cf. supra*) comme des réflexions sur l'utilisation des *Topic Maps* pour un Web plus « cognitivement sémantique » (Caussanel *et al.*, 2002).

À la suite de ces considérations et de l'interaction agent logiciel-utilisateur, les index devront finalement être utilisés par des agents logiciels et leur caractère formel, celui des ontologies, deviendra primordial en même temps qu'il faudra que les ressources soient indexées avec icelles. La recherche de procédures plus ou moins automatique d'indexation sera une des clés de l'existence du Web sémantique permettant à des agents logiciels d'interopérer à grande échelle.

Après ces considérations spécifiquement liées aux ontologies, nous poursuivrons les interrogations liées au Web sémantique appliqué à la médecine au chapitre 5, paragraphe 6.2³⁰.

8 Conclusion

Au long de ce chapitre, nous avons essayé d'éclaircir la notion d'ontologie en recherchant son origine, puis en mettant en perspective un certain nombre de travaux de la communauté Ingénierie des connaissances. Nous espérons avoir montré que la notion d'ontologie a évolué dans un continuum de réflexions cohérentes pour que puisse se tracer un chemin d'évolution vers la stabilisation du concept. Les exemples, en particulier médicaux, dans des projets auxquels nous avons participé nous ont permis de préciser autant que faire ce peut les tenants et aboutissants d'une démarche ontologique.

Les outils de TAL que nous avons utilisés dans plusieurs projets nous ont convaincu de leur utilité, si ce n'est leur caractère indispensable, pour peu, évidemment, qu'on ait un corpus adéquat.

Enfin et pour terminer, nous avons abordé les autres produits terminologiques que sont les index, thésaurus et terminologies pour voir comment ils étaient utilisés dans des démarches de conceptualisation ou d'indexation/recherche d'information, en particulier dans le contexte du Web sémantique. Cela nous a permis de définir précisément les possibilités de ces différents produits et de montrer que si une ontologie est indispensable dans certaines situations, elle a un « coût » qui peut obliger à envisager d'autres produits – *e.g.* un thésaurus – dans d'autres situations, en particulier celles liées à l'indexation. Ce coût est celui de la formalisation et de la recherche d'un accord. Dans ce contexte, il reste encore de nombreuses recherches et expérimentations à mener pour que les ontologies soient des objets accessibles et que leur ingénierie soit aisément transmissible.

³⁰Le concept de médiation entre terminologies, abordé dans ce chapitre (*cf.* § 3.4) n'est approfondi ici par rapport au Web sémantique que dans la question de l'alignement des ontologies mais les médiateurs englobent des problématiques d'interopérabilité qui seront donc développées dans le chapitre suivant.

Crédits

Les résultats sur les ontologies développés ici sont le fruit de réflexions menées en commun avec de nombreux chercheurs, que ce soit Bruno Bachimont sur l'ontologie de MENELAS et pour ses recherches épistémologiques sur la constitution d'ontologies, Jacques Bouaud et Pierre Zweigenbaum sur l'ontologie de MENELAS.

Plus récemment, le groupe TIA <<http://www.biomath.jussieu.fr/TIA>> a été un lieu de réflexion privilégié sur la constitution d'ontologies et autres produits terminologiques. Plus récemment encore, les chercheurs de l'action spécifique ASSTICCOT <<http://www.irit.fr/ASSTICCOT/>> et, en particulier, les membres du sous-groupe Ingénierie des connaissances, m'ont donné matière à réflexion et synthèse sur les ontologies et leur constitution. Les rapports de synthèse de Gilles Kassel (2001) et Fabien Gandon (2002) m'ont été d'une aide précieuse dans, justement, ce travail de synthèse.

Sur des applications récentes, Sophie Lemoigno a développé l'ontologie de la réanimation chirurgicale. Avec Marie-Christine Jaulent qui a co-encadré le DEA de Sophie avec moi, elles m'ont permis de faire avancer ma réflexion sur les ontologies en médecine. Jean-Christophe Weis que j'ai dirigé en stage de DEA a travaillé sur l'ontologie de la périnatalité et sa tentative de réutilisation du « haut » de l'ontologie de MENELAS a été riche d'enseignements.

Enfin, Didier Bourigault, par son long et précis travail sur le développement d'outils du TAL dévolus à la constitution de produits terminologiques et plus précisément d'ontologies a été un partenaire efficace et réfléchi sur le développement d'ontologies médicales ou d'index en Ingénierie des connaissances . . .

Des systèmes d'information de santé et de l'interopérabilité

Dans ce chapitre, nous voulons aborder le contexte de la mise en œuvre d'une gestion des connaissances médicale. Ce contexte est celui des systèmes d'informations. Systèmes dont le champ d'action grandit, originellement limité à un même établissement de soins – hôpital ou clinique –, concernant maintenant des communautés de soins plus larges que sont les « réseaux de santé » qui s'étendent à la médecine de ville et impliquent l'ensemble des professionnels de santé pour une pathologie ou une situation donnée. Cette situation crée de nouvelles contraintes pour la mise en œuvre de ces larges systèmes d'informations réseaux – des systèmes d'information de santé –, en particulier au niveau d'un de leurs buts premiers qui est d'assurer l'interopérabilité des différents acteurs du réseau. Cette interopérabilité passe et doit passer par des standards qui se construisent lentement et s'imposeront encore plus lentement. Cette lenteur doit être vue comme nécessaire à la reconnaissance du bien fondé de ces standards et à leur acceptation. Cette acceptation étant, comme nous le verrons, très dépendante des contextes organisationnels et techniques des différents systèmes de soins et des systèmes d'information de santé sous-jacents.

1 L'interopérabilité

Dans ce chapitre, nous allons nous intéresser à l'interopérabilité des Systèmes d'Information dans le domaine médical. Nous prendrons le point de vue de l'Ingénierie des connaissances et essaierons de repérer les liens il y a entre les connaissances à représenter et le système d'information. Nous montrerons que des choix de représentation sous-tendent des modes différents d'appropriation des connaissances, des modes différents d'échanges et donc des spécifications différentes de l'interopérabilité.

Pour cela nous allons décrire 3 approches emblématiques de l'interopérabilité en santé : (a) une approche dite textuelle *centralisée* où l'information médicale est conservée dans son format d'élaboration, le texte (cf. section 2), (b) une approche normalisatrice où les données médicale sont échangées via des messages structurés typant fortement ces même données (cf. section 3) et (c) une approche « d'enveloppe » essayant de conserver les liens et la sémantique d'élaboration de l'information médicale (cf. section 4). Dans la section 5 nous essaierons de proposer une grille de lecture de ces

approches et de relier les questions de ce qui ne semble, dans un premier temps, que de la syntaxe à des questions de sémantique et donc d'interopérabilité sémantique (*cf. infra*) puis à des questions d'architecture informatique et des problématiques organisationnelles.

Auparavant, il est nécessaire de définir ce que nous entendons par interopérabilité. On peut trouver une première définition intéressante dans (CSA, 1992) :

Interopérabilité : *Faculté que possèdent des produits et des ensembles de produits informatiques hétérogènes de fonctionner conjointement. L'une des conditions fondamentales permettant la communication entre ordinateurs et logiciels hétérogènes est l'utilisation de langages et de protocoles communs. Possibilité de communication, d'exécution de programmes ou de transfert de données entre unités fonctionnelles différentes de manière que l'utilisateur n'ait que peu ou pas besoin de connaître les caractéristiques propres à chaque unité. Le meilleur exemple d'interopérabilité ce sont les protocoles qui gouvernent le fonctionnement d'Internet et qui permettent à des ordinateurs utilisant des technologies et des systèmes d'exploitation différents d'échanger de l'information.*

La médecine ne diffère pas d'autres secteurs professionnels pour l'échange de messages et cette définition s'y applique parfaitement. Comme pour les autres secteurs, il y a d'abord la syntaxe de l'échange où l'on définit la nature, le type et surtout la structure des messages échangés. Le respect de cette syntaxe entre applications permet une *interopérabilité syntaxique*. Si l'on s'intéresse au contenu des messages, à la signification de ce qui est réellement échangé (quoi? Quel type d'information? ...) et qu'on essaie de créer les conditions pour que des applications échangent des informations qui ont le même sens aux deux extrémités de la chaîne, alors on crée les conditions d'une *interopérabilité sémantique*¹. (Degoulet & Fieschi, 1997).

Munis de ces définitions, il nous est alors possible d'étudier les trois approches sus-nommées proposées.

2 L'information médicale et son contexte

Nous allons reprendre ici, les principales conclusions de la section du même nom du chapitre 4 (*cf. chap. 3, § 1*) et une description rapide du projet HOSPITEXTE au regard des questions d'interopérabilité discutées ici (*cf. chap. 3, § 2*).

Le caractère contextuel de l'information médicale (*cf. chap. 3, § 1.1*) fait qu'elle ne s'énonce bien qu'à travers le langage naturel écrit dans des documents textuels. Cette forme 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, des notes de suivi ou des résultats des examens complémentaires, sous des formes et avec des mises en page constantes. Ainsi, 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*.

La numérisation – le changement de support – de ces documents doit alors être envisagée à travers des documents hypertextuels, accessibles via une interface Web. C'est ce qu'envisagent un certain nombre d'expérimentations autour du dossier médical informatisé, en particulier le projet HOSPITEXTE.

Dans cette optique, XML est par son caractère de standard et sa capacité à prendre en charge l'enregistrement d'informations semi-structurées, le langage d'implémenta-

¹ Comme nous l'avons vu tout au long du chapitre 2, cette question du sens est reliée à l'interprétation des utilisateurs des applications. Ce point ne sera pas redéveloppé ici. Il sera juste réabordé en fin de chapitre sur les conditions de l'interopérabilité sémantique (*cf. 6.1*).

tion privilégié d'un dossier médical fondé sur les documents. Enfin, le Web, devenu un standard ergonomique de fait est l'interface évidente de tels dossiers hypertextuels même s'il n'est pas sans problème (cf. chap. 3, § 2.4).

3 De l'échange de données

3.1 XML et l'EDI

L'activité de normalisation autour de XML, s'est rapidement focalisée sur le transport de données informatisées. En termes de normes, il existait jusque-là dans ce domaine, la norme EDIFACT qui prenait en charge, comme son nom l'indique, l'échange et le transport de données informatisées, de l'envoi à la réception, en passant par les différentes vérifications d'intégrité des messages.

En ce domaine, XML est utilisé pour sa capacité à encoder des messages structurés et normalisés. On peut résumer en deux points principaux les avantages d'une mise en œuvre des concepts de l'EDI avec XML plutôt qu'EDIFACT :

- La structure du message est déclarée à l'extérieur des programmes qui vont la décoder, dans la DTD, et un programme simple (un *parser*) peut vérifier la conformité de la structure du message à cette DTD – *i.e.* la spécification. Cette déclaration de la spécification à l'extérieur du programme qui s'en sert – *i.e.* la *programmation déclarative* – permet une évolutivité plus grande des structures de messages qui sont ainsi plus facilement modifiables ;
- Il est possible de déclarer une structure de message générique repectant une DTD, puis, sur cette même base, des structures plus spécialisées pour des besoins plus précis. Ce point recouvre en pratique deux propriétés différentes, (1) le fait qu'un document puisse respecter une DTD qui ne spécifie pas toute sa grammaire puis une autre DTD, plus précise, la spécifiant complètement ou (2) la possibilité d'utiliser les *formes architecturales* standardisées par HYTIME (ISO/IEC, 1997) pour spécifier le niveau logique de certaines sections comme on le fait, par exemple, avec *section* ou *subsection* en \LaTeX^2 .

Ceci dit, il ne faut pas croire que XML utilisé ici résout tout : c'est ici un langage qui permet de fournir de façon déclarative des structures de messages. Il ne préjuge en rien du transport de ce message, de la sécurité de ce transport, ou même de l'intégrité du message³. La figure 5.1 présente un extrait de la DTD MEDDOS de transport de tout ou partie d'un dossier médical. On voit rapidement que de telles DTD ne servent qu'à spécifier – ce qui est en soit important – le type et la structure de données échangées. Ici, comme ailleurs, XML est un langage de description mais des choix de représentation – que l'on peut qualifier de « sémantiques » au regard des problèmes de syntaxe XML – doivent être faits : Qu'est-ce qu'un élément de dossier médical ? Quel en est l'auteur ? Quel contexte transporte-t-on ? Quelle compréhension doit en avoir le récepteur ?

3.2 HL7/CDA : a « Clinical Document Architecture »

À l'intersection de réflexions sur la représentation des connaissances médicales (cf. § 1) et sur l'échange de données cliniques (cf. § 3.1), un groupe de réflexion spécifique au sein du HL7 XML *Special Interest Group*⁴, le *Kona Group*, a fait des propositions d'ar-

²Si les formes architecturales HYTIME nous sont très utiles comme nous le verrons plus loin dans le chapitre, il n'est pas évident que la norme perdure au sein des langages XML. Les spécifications des *schémas* XML peuvent jouer le même rôle par rapport à l'usage que nous en faisons mais nous ne développerons pas ce problème ici.

³Nous verrons section 4 que pour ce faire, des protocoles comme SOAP ont été défini. Ils sont eux-aussi, dans la philosophie des langages de la « galaxie » XML, décrits en XML.

⁴Le HL7 ou *Health Level 7* est un organisme qui propose des spécifications de messages pour la communication entre applicatifs dans le domaine de l'informatique médicale. C'est en pratique le seul qui compte au

```

<!ELEMENT MEDDOS (MESSAGE_MEDDOS, EXPEDITEUR, DESTINATAIRE, PATIENT)>
<!ELEMENT MESSAGE_MEDDOS EMPTY>
  <!-- ATTLIST MESSAGE_MEDDOS -->
    local_id                ID          #IMPLIED
    identification_emetteur  CDATA     #FIXED    "05"
    identifiant_recepteur   CDATA     #FIXED    "01"
    date_heure_preparation  CDATA     #REQUIRED
    meddos_count            CDATA     #REQUIRED
    type_du_message         CDATA     #FIXED    "MEDDOS"
    num_version              CDATA     #FIXED    "962"
    num_revision_version    CDATA     #FIXED    "1"
    agence_de_controle      CDATA     #FIXED    "PR"
    nom_du_message          CDATA     #FIXED    "dossier médical"
  >

<!ELEMENT EXPEDITEUR EMPTY>
  <!-- ATTLIST EXPEDITEUR -->
    message_local_id        ID          #IMPLIED
    signid                  CDATA     #REQUIRED
    profid                  CDATA     #REQUIRED
    active_flag             CDATA     #IMPLIED
    meddos_conid           CDATA     #REQUIRED
    orgid                   CDATA     #IMPLIED
    nom                     CDATA     #REQUIRED
    prenom                  CDATA     #IMPLIED
    title                   CDATA     #IMPLIED
    adress1                 CDATA     #IMPLIED
    adress2                 CDATA     #IMPLIED
    adress3                 CDATA     #IMPLIED
    ville                   CDATA     #IMPLIED
    codepost                CDATA     #IMPLIED
    tell                    CDATA     #IMPLIED
    tel2                    CDATA     #IMPLIED
    fax                     CDATA     #IMPLIED
    email                   CDATA     #IMPLIED
  >

```

FIG. 5.1 – Un extrait de la DTD MEDDOS de transport de tout ou partie d’un dossier médical. Le message MEDDOS a été conçu durant le projet européen PROMPT. Originellement implémenté en EDIFACT, il a été ensuite porté en XML en 1998 par ICSF <<http://www.icsf.fr>>.

chitectures de dossier médical utilisant XML et intégrant à la fois la représentation des connaissances et le transport de données. Cela se traduit maintenant par une proposition d’architecture de référence (*the Clinical Document Architecture – CDA – HL7-SD-TC*, 2001) qui tire parti des propriétés de XML pour organiser les informations médicales en trois niveaux de spécificité (cf. fig. 5.2) :

Coded Header, un niveau de documents repérés de façon unique grâce aux informations enregistrées dans le *header* (nom, identifiant, etc.) sans contrainte sur le contenu textuel ;

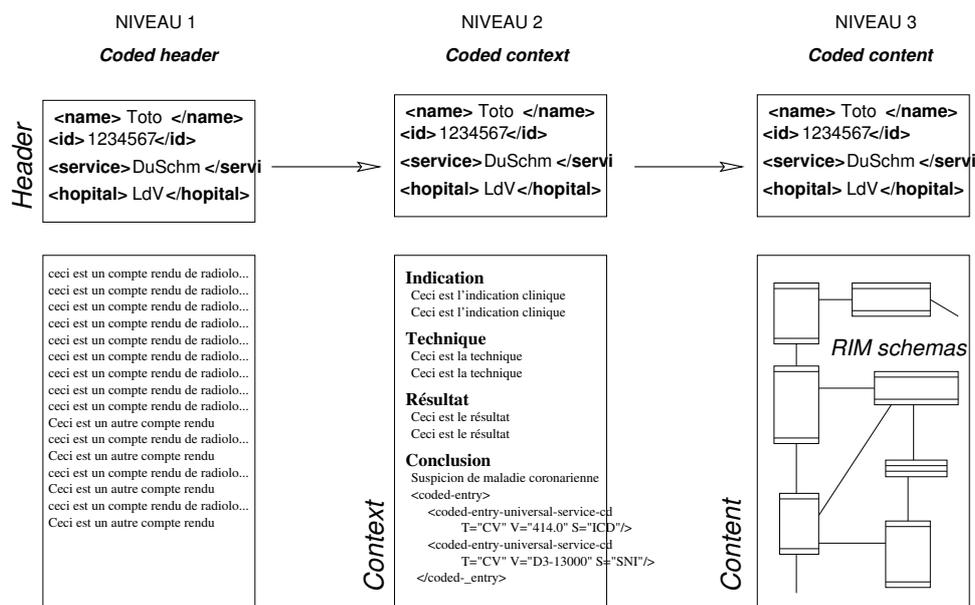
Coded Context, un niveau de spécificité plus fin, respectant les mêmes *headers* que le niveau précédent mais avec un contenu textuel organisé logiquement (*a minima* sections), permet p. ex. une organisation des documents selon les nécessités (raisonnement, administratives, etc.) ;

Coded Content, un niveau de découpage permettant de structurer de façon détaillée (cf. § 5.2) des informations médicales en s’appuyant pour cela sur le *Reference Information Model* (RIM). Le RIM présente une vue consistante des données à échanger et permet de construire des messages que pourront s’échanger différents applicatifs.

Ceci posé, cette architecture veut respecter un certain nombre de principes :

- architecture XML multi-niveaux ;

niveau des États-Unis. Il est un des interlocuteurs privilégié du CEN TC251, l’organisme européen en charge de l’élaboration des normes dans le domaine de l’informatique médicale <<http://www.centc251.org>>.

FIG. 5.2 – Schéma de la *Clinical Document Architecture*

- respect des formes architecturales. Comme pour HOSPITEXTE (Brunie *et al.*, 2000), le recours aux formes architecturales standardisées par HYTIME (ISO/IEC, 1997) semble obligatoire si l'on veut pouvoir organiser correctement les informations médicales liées à un patient ;
- fonctions d'organisation, enregistrement, management, agrégation des documents requises à tous les niveaux ;
- nécessité pour chaque institution de développer des DTD correspondant à ses propres usages (les schémas et/ou DTD d'échanges ne *sont pas* adéquats à l'enregistrement des informations) en respectant les formes architecturales des niveaux concernés. Ce principe est signalé pour bien montrer qu'on conserve, édite, élabore des documents pour un usage précis lié au travail et qu'on en édite d'autres destinés à gérer la communication. Cela n'empêcherait pas d'envoyer un document sous sa forme locale à un interlocuteur mais à condition que ce dernier en comprenne la sémantique et soit donc capable de le lire⁵.

Ces principes sont évidemment garants de l'utilisabilité et de la généricité d'une telle architecture. Le dernier niveau, *Coded Content*, fait directement référence au *Reference Information Model* qui correspond à une normalisation de l'information à communiquer. Ce modèle est, dans sa dernière version en cours de développement (3.0), décrit en UML et est développé en collaboration avec le CEN TC251 en essayant de faire converger le *Reference Information Model* avec la prénorme européenne sur l'architecture des dossiers informatisés communicants (ENV 13606, appelée maintenant EHRCOM).

⁵On peut imaginer que des institutions se mettent d'accord pour développer des DTD de représentation communes mais on sort là de la problématique de l'interopérabilité puisque les interlocuteurs se mettent dans ce cas dans des conditions homogènes. Situation par expérience anecdotique dans un paysage par définition hétérogène.

4 Un standard d'enveloppe

4.1 Contexte de la proposition

EbXML [6], proposé par le consortium d'industriels de l'informatique et des télécommunications OASIS (*Organization for the Advancement of Structured Information Standards*) et l'UN/CEFACT (*United Nations Center For Trade Facilitation and Electronic business*), vise à définir un format universel d'échange de messages dans le cadre du commerce électronique au sens large. Il intègre depuis peu le format d'échanges SOAP – *Simple Object Access Protocol* <<http://www.ebxml.org>>. Le message ebXML a alors la structure suivante, basée sur le concept d'enveloppe (cf. fig. 5.3) : (a) une *entête* renseignant les caractéristiques d'expéditeur, de destinataire, de routage, etc., (b) un *bordereau* fournissant la liste des documents envoyés et (c) les documents eux-mêmes.

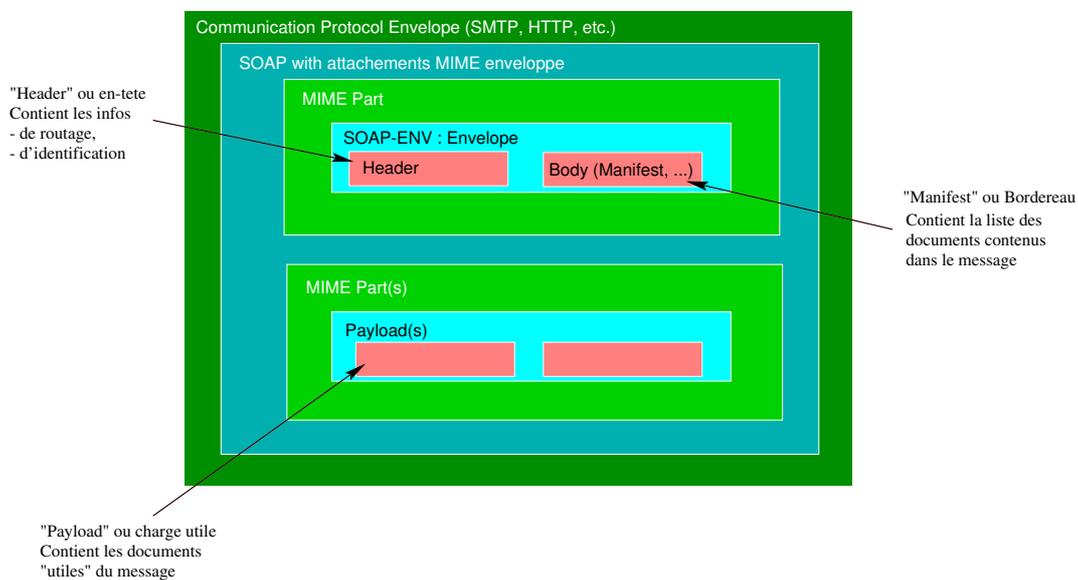


FIG. 5.3 – Structure d'un message ebXML (version 0.98b du 13 mars 2001).

L'objectif du GT11 (groupe de travail 11 au sein d'EDISANTÉ⁶) est d'émettre des propositions techniques sur un format générique d'échange de données médicales, de les expérimenter dans le cadre de démonstrations et de les promouvoir au sein des organismes ad hoc. L'approche proposée par le GT11 est pragmatique, en ne faisant que créer des passerelles entre les standards existants ou émergents, et en appliquant sur un standard générique (ebXML) les éléments spécifiques à la santé issus du CEN ou d'HL7 et DICOM⁷.

Les propositions du GT11 portent sur une structure permettant de transporter des données et des documents hétérogènes mais avec des informations associées, renseignant sur la finalité du message et son contenu, et en permettant la gestion et le traitement – c'est le concept d'enveloppe. Ce pré-standard ne porte donc pas sur d'autres sujets à savoir : (a) une structure de dossier patient, (b) les protocoles de partage de

⁶La mission d'EDISANTÉ est de faire des propositions qui rassemblent les acteurs significatifs du domaine (utilisateurs, industriels...) et de les soumettre aux instances de normalisation. Son domaine d'intervention concerne les formats d'échanges de documents électroniques dans le domaine de la santé <<http://www.edisante.org>>.

⁷DICOM est une norme qui définit une méthode de communication pour les différents équipements d'imagerie médicale numérique(). DICOM s'intéresse depuis déjà de nombreuses années à l'échange de documents médicaux multimédia, au delà de son champ d'intérêt initial qui était celui de l'imagerie biomédicale.

données médicales (l'échange supposant un « transfert » d'une copie de l'information d'un acteur à un autre acteur, le partage supposant un accès de deux acteurs à la même copie de l'information) et (c) le contenu lui-même des échanges de données administratives ou non médicales.

4.2 La notion d'enveloppe d'échange médical

Il a donc été introduit la notion d'enveloppe d'échange médical (cf. fig 5.4). Ce concept rejoint totalement le concept récent d'enveloppe ebXML, défini dans le cadre beaucoup plus général des échanges sur Internet pour le commerce électronique (cf. § 7). Il s'en différencie par deux points principaux :

Le patient comme unique objet de la transaction. Un tel échange ne saurait être anonyme du point de vue du couple émetteur-récepteur. Il concerne donc un émetteur et un destinataire, qui tous deux sont impliqués et responsabilisés dans l'échange. Le seul moyen de permettre à un émetteur de signer un envoi réservé à un récepteur précis, concernant un patient qui a le droit d'exiger d'en connaître le contenu, aboutit à une structure nécessairement unique pour le triplet {émetteur, récepteur, patient}. La seule réserve est la notion de destinataires multiples (y compris les aspects de délégation) mais il convient dans ce cas de savoir gérer les avis de remise de manière différente pour chacun de ceux-ci (traçabilité).

Le caractère multimédia des informations transportées. Une analyse de la situation chez la majorité des acteurs de santé fait apparaître majoritairement l'existence de sources multiples d'informations concernant un même patient. Ces informations médicales ne sont pas nécessairement liées entre elles, notamment sur le plan informatique, et se présentent sous des formes et sur des supports divers (cf. fig. 5.4) :

- base(s) de données patient « communes » ou « de spécialité(s) »,
- document textuel saisi au moyen d'un logiciel de traitement de texte en dehors du logiciel de dossier patient,
- document numérique reçu de l'extérieur mais non intégré dans la base de données,
- document textuel numérisé et non intégré,
- image produite de manière numérique,
- image analogique (film, support papier, vidéo), numérisée,
- commentaire vocal,
- signal physiologique.

Si ces informations ne sont pas toujours gérées de manière centralisée chez l'émetteur, il peut être fondamental de les réunir à l'occasion d'un échange avec un autre acteur de santé, qui, lui, saura éventuellement les intégrer dans sa base de données. Même sans lien informatique structuré, le fait de les envoyer ensemble a un sens sur le plan médical, par rapport au contexte précis de l'échange. Sur le plan de la traçabilité de l'échange, il est donc fondamental pour l'émetteur et le récepteur de pouvoir prouver que ces informations ont été transmises ensemble.

5 Comparaison des approches

Les approches développées précédemment ont des implications différentes en termes d'organisation du transfert de l'information, d'organisation des communautés de travail et finalement du SI sous-tendant ces organisations.

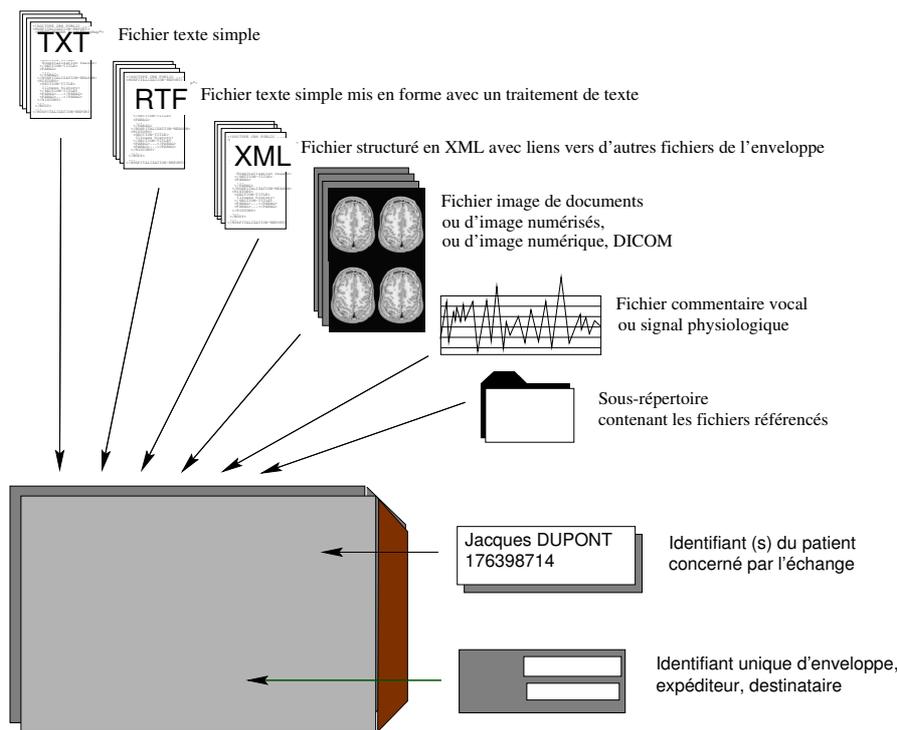


FIG. 5.4 – Principes du contenu de l'enveloppe d'échange

5.1 Une information textuelle disponible sur un serveur Web.

Une approche à la HOSPITEXTE (cf. chap. 3, § 2) est souvent dite *centralisée* dans la mesure où l'information est, *a priori*, disponible en un seul endroit, sans redondance, même si l'ensemble de l'information concernant un patient est répartie dans la totalité des lieux où l'a amené sa *trajectoire de soins*.

Vis-à-vis de la technique, une approche fondée sur l'hypertexte et les langages HTML/XML se prête particulièrement bien à des accès via le Web puisque respectant les bases même de la « philosophie » de ce dernier comme ses normes informatiques et ses protocoles d'échanges.

Si les considérations de sécurité et de confidentialité peuvent être satisfaites dans le domaine médical⁸, elles ne satisfont pourtant pas la demande de sécurité des professionnels comme des patients : le Web développe chez les utilisateurs un sentiment d'insécurité qui semble le rendre inapte à stocker des informations de santé nominatives. Ce sentiment n'est pas une donnée réelle de la sécurité sur le Web que beaucoup de développements tendent à augmenter mais une donnée objective de la perception des patients comme de certains usagers professionnels⁹. Vis-à-vis des patients, les récentes expériences sur des dossiers médicaux partagés semblent converger vers l'idée que si il n'est pas acceptable de « voir » sur le Web des données fortement médicales comme

⁸Les cartes de professionnels de santé (CPS), les cartes SESAM-VITALE, les *tiers de confiance* stockant les numéros d'identifiant des patients concourent à la mise en œuvre de normes de sécurité spécifiques à la médecine, en dehors des processus de cryptologie maintenant adoptés par tous. On lira sur ce sujet, plus compliqué qu'il ne semble au premier abord, des propositions d'architectures et de protocoles de sécurisation adaptés à la médecine (Maret & Ghédira, 2002; Wack *et al.*, 2002)

⁹Ceci dit et quel que soit le mode de communication de l'information, le problème de la responsabilité liée à l'administration de données qui sont partagées entre différentes institutions de soins n'est pas résolu actuellement. On peut imaginer, au sein d'un réseau, que ce sera l'identité juridique représentant le réseau qui sera responsable mais encore faut-il que ce soit une solution acceptable.

des comptes rendus d'hospitalisation, des données moins « sensibles » correspondant à des prescriptions, des habitudes alimentaires, etc., provoquent moins de réticences. Ce serait l'avènement du « carnet de santé » au détriment du dossier médical¹⁰.

En ce qui concerne les communautés professionnelles – *e.g.* médicales – le partage d'une information centralisée à l'extérieur de la communauté ne semble pas se développer même s'il faut être prudent quand on fait des prédictions de ce type. Dans cette situation, la question est en fait de spécifier quelle est l'étendue de cette communauté, d'un point de vue organisationnel. Au sein d'un grand hôpital, cette volonté de partage peut être limitée à un seul service. Les *réseaux de santé*, entités médicales suscitées et créées pour favoriser le développement de communauté de professionnels autour d'un patient ou d'une entité mère/enfant dans le cadre des réseaux de périnatalité, se bâtissent sur le sentiment d'appartenir effectivement à une même communauté de soins. Dans ce cas, le partage d'une information centralisée est envisagé et parfois réalisé. A *contrario*, l'hétérogénéité des équipements et des accès au Web des membres d'une telle communauté est parfois un obstacle technique à cet accès centralisé.

Ainsi, ce que la technique permet théoriquement peut être freiné par la mise en œuvre pratique ou, plus « sûrement », par la volonté des organisations.

5.2 La normalisation de l'information *versus* normalisation des messages.

Le modèle décrit section 3 est très normalisateur. Dans la première approche proposée (*cf.* § 3.1), XML est utilisé dans sa capacité à proposer une syntaxe de structuration de message. C'est le même principe qui est utilisé dans l'approche de normalisation européenne du CEN TC251. La philosophie semble en être : « puisqu'il est difficile, voire impossible, de se mettre d'accord sur une norme de dossier, ne proposons que de "bons" principes d'architecture¹¹ et focalisons-nous sur les normes d'échange ».

Cette position est maintenue encore plus fermement dans l'approche de HL7 (*cf.* § 3.2) dans une volonté d'intégration qui se traduit, dans la *Clinical Document Architecture*, par une complète dichotomie dans les principes comme dans la réalité entre le modèle du dossier médical et les modèles d'échanges¹². Par rapport aux échanges, le CEN TC251 et HL7 tentent de travailler de concert à leur normalisation. Par rapport au dossier médical, la situation est plus complexe qu'il n'y paraît : d'un côté, le niveau 2 de la *Clinical Document Architecture* est très proche des principes de HOSPITEXTE et nous semble respecter une tendance de fond en système d'information hospitalier dans la mesure où l'architecture ne pose aucun problème de mise en œuvre informatique ; d'un autre côté, celui des échanges, « l'alimentation » du niveau 3, le *reference Information Model* pour construire les messages est passée sous silence : silence révélateur des difficultés qu'il y a et qu'il y aura à passer du niveau 2 au niveau 3. Les solutions sont à chercher du côté des approches ontologiques et de travaux en traitement du langage médical (*cf.* § 6.1).

Enfin et puisqu'on essaie ici de normaliser tous les objets ou les champs échangés (*cf.* figure 5.1), la norme telle qu'elle se met au point au CEN TC251 – ENV 13606, vol. 1 à 4 – est copieuse et complexe.

¹⁰Le projet P2VIE – prévention et prise en charge du vieillissement par l'information électronique – du réseau national de technologie en santé (RNTS) teste justement cette approche de carnet de santé <<http://www.p2vie.org>>.

¹¹Ces principes font aussi l'objet de travaux de normalisation au niveau du CEN TC251 mais ne seront pas développés ici.

¹²Comme le traduit bien le principe : « Nécessité pour chaque institution de développer des DTD correspondant à ses propres usages (les schémas et/ou DTD d'échanges ne *sont pas* adéquats à l'enregistrement des informations) ». Seule l'utilisation du langage XML qui montre là toute ses potentialités peut laisser croire au lecteur distrait qu'il y a des liens structurels évidents.

5.3 L'échange d'extraits de dossier multimédia

L'approche fondée sur une enveloppe (cf. § 4) est obligatoirement normalisatrice sur les conditions de l'échange, l'expéditeur, le destinataire, les conditions de routage, etc. Elle suit en cela les recommandations ebXML enrichies de prescriptions spécifiques à la santé.

En revanche, elle est beaucoup moins normalisatrice sur le contenu transporté par l'enveloppe si l'on se place du point de vue de la syntaxe et du type des objets transportés. Si l'on s'intéresse à la sémantique de l'échange, on voit apparaître, comme pour le projet HOSPITEXTE, la sémantique des relations qu'entretiennent les objets, souvent des documents, entre eux : de la même façon que l'informatique permet de créer de nouveaux documents de « navigation » ou de « lecture » (cf. chap. 3, § 2), elle permet de créer et de matérialiser par des liens de nouvelles proximités entre documents. Ainsi, des documents envoyés « ensemble » dans une enveloppe ont un sens qu'ils n'auraient pas séparément.

Si elle est évidemment toujours présente, la normalisation sur les contenus se fait ici *a posteriori*, au regard d'une normalisation *a priori* dont est porteur l'approche précédente. En étudiant les conséquences d'une telle approche sur le possible pilotage de la fabrication des systèmes d'information (cf. chap 6, § 4.3.2), on peut voir l'approche par échange d'extraits de dossiers multimédia comme beaucoup plus « contextuelle » qu'une approche fortement normalisatrice plus « technicienne » (David & Pallez, 2001).

Enfin, il est à noter que l'approche enveloppe n'est pas incompatible avec l'approche précédente dans la mesure où rien n'empêche de respecter, au sein de l'enveloppe, les labels et les structures d'un certain nombre de champs normalisés¹³.

6 Discussion et perspectives

6.1 De l'interopérabilité sémantique

Tout système d'information, *a fortiori* dans le domaine de la santé – *i.e.* système d'information de santé ou SIS –, manipule des objets formels. Ces objets sont répertoriés au sein de référentiels et sont le *vocabulaire* du SI. Dans un domaine comme la santé et si on essaie de manipuler des objets correspondant à des activités de soins – *e.g.* *signes*, *syndromes*, *diagnostics* – on se retrouve rapidement à organiser des *vocabulaires* gigantesques comportant pour certains plusieurs centaines de milliers de mots. Ces vocabulaires sont alors organisés en hiérarchies, spécialité par spécialité, sous forme de thésaurus de spécialité. Mais tous souffrent d'une sémantique peu claire, de définitions non partagées par les auteurs, etc. En bref, de leur incapacité à assurer une réelle interopérabilité sémantique en SIS (cf. chap. 4, § 3.3). Comme nous l'avons vu au chapitre 4, les solutions sont à trouver du côté des ontologies qui permettent, par leur côté formel, d'assurer la cohérence et l'évolutivité des thésaurus (cf. chap. 4, § 1.4) et de maintenir un objet dual formel de ceux-ci. Ce sont les conditions de l'interopérabilité sémantique tant recherchée (cf. § 1).

6.2 Du Web sémantique

Après le chapitre 4, paragraphe 7, nous poursuivons ici notre approche des concepts du Web sémantique appliqué à la médecine. Ce sont évidemment les questions d'interopérabilité qui sont prégnantes ici, augmentées de la question des services Web :

¹³Cette comptabilité potentielle semble bien comprise par le HL7 qui étudie maintenant ebXML et SOAP pour ses échanges de messages.

L'interopérabilité syntaxique. La question ici est l'hétérogénéité de représentation. Même si des progrès sont à attendre des actions de normalisation du Web sémantique autour d'une représentation des ontologies respectant le langage OWL, on ne peut espérer qu'elles soient partageables à très grande échelle, sauf cas spécifiques, et il sera nécessaire de mettre en place des traducteurs entre les différents langages.

L'interopérabilité sémantique et l'approche médiateurs. Comme nous venons de le voir (cf. § 6.1), l'interopérabilité sémantique est de prime abord une question de partage d'ontologies formelles entre applications. En pratique, la question est plus compliquée dans la mesure où il n'est pas toujours possible d'imposer un schéma conceptuel commun à plusieurs systèmes devant partager des connaissances. L'approche médiateur consiste alors à définir une ontologie regroupant l'ensemble des spécifications modélisant le domaine d'application, en décrivant des vues abstraites de chacune des sources d'information dans les termes de l'ontologie¹⁴. Dans ce contexte de systèmes consistant en des sources d'information hétérogènes et utilisant des ontologies locales, il sera nécessaire d'étudier les méthodes permettant d'automatiser au plus près le rapprochement entre ontologies, en termes d'alignement de recouvrement ou d'inclusion, etc. (Laublet *et al.*, 2002).

Les services Web. Ce vocable met l'accent sur les fonctionnalités attendues de tel ou tel logiciel sur le Web. Il correspond alors en la recherche de techniques et standards pour permettre à des agents logiciels d'interagir à haut niveau, en se repérant mutuellement à travers des annuaires (norme UDDI), en se définissant les services proposés (norme WSDL), en définissant précisément les modalités d'échanges d'information (norme SOAP), etc. L'utilisation de la norme SOAP (cf. § 4) dans l'échange d'information médicales est un pas de l'informatique médicale vers les services Web.

Avec les problèmes de normalisation des syntaxes d'échanges, on a là les deux clés de l'interopérabilité : respecter les mêmes syntaxes pour pouvoir communiquer – norme XML, SOAP, ebXML – et donner le même sens aux mots – *via* les ontologies – pour échanger des informations qui aient un contenu compréhensible par tous (humains et « machine ») et identique (Zweigenbaum & Charlet, 1999)¹⁵.

6.3 Des standards pour l'information médicale

Nous avons décrit ici un certain nombre d'approches de l'interopérabilité, toutes fondées sur des standards ou de futur standards. Ce panorama n'épuise pas, et de loin, les initiatives prises dans le domaine de l'informatique médicale et plus précisément vis-à-vis des SIS. Cette « abondance » est une difficulté et c'est la raison pour laquelle il est important de replacer les approches et les standards impliqués par rapport aux conditions organisationnelles de leur validité. Deux points sont alors cruciaux dans ce contexte (EdiSanté, 2001) :

Les standards, garants de l'ouverture et du bon fonctionnement du SIS. L'ouverture, s'agissant d'échanges entre professionnels, est une exigence absolue. S'agissant de messages médicaux, il faut des systèmes interopérables et fortement sécurisés. Mais, comme on l'a vu, l'ouverture exige plus : il faut que la syntaxe des messages soit respectée entre 2 partenaires mais que la *sémantique* le soit aussi. Ainsi, un patient doit avoir un même numéro d'identification quels que soient les

¹⁴Cette description est une des approches médiateurs possible. On lira (Laublet *et al.*, 2002) pour approfondir la question.

¹⁵Cette affirmation se place évidemment dans le contexte de l'interopérabilité entre machines. Le fait que les ontologies soient aussi compréhensibles par les êtres humains a été longuement discuté au chapitre 4.

partenaires s'échangeant des informations à son sujet¹⁶. Plus, dès qu'il y a codage d'informations, il y a nécessité d'accord *a minima*, ne serait-ce que sur des entêtes ou enveloppes, sinon le « bon vieux » papier reprendra le dessus. Enfin, la classification des patients et leur représentation partagée exige un accord total sur le système de représentation conceptuelle choisi (cf. § 6.1). C'est au respect de ses conditions que nous aurons des SIS efficaces.

Les standards : les respecter pour ne plus en parler. Dans un point de vue plus contextuel (cf. § 5.3) et par rapport aux documents XML, il est important d'insister sur le fait qu'un tel standard permet de se rendre indépendant des outils l'exploitant. En d'autres termes, le fait qu'il s'agisse de norme permet de constituer le dossier en ressource documentaire générique : tout outil qui se conforme à la norme XML permet d'exploiter ces ressources. L'intérêt de s'appuyer ainsi sur des normes est de pouvoir consacrer l'effort méthodologique sur les difficultés essentielles, à savoir le contenu et la forme de l'information médicale. Les hôpitaux, les cliniques ou les réseaux de soins n'ont pas vocation à développer des outils informatiques, mais bien plutôt à réfléchir sur l'information qui structure leur activité – dans ce cas et pour les documents, réfléchir aux DTD à développer. De plus, la pérennité du patrimoine informationnel ne peut être assurée que si le mode de représentation – et d'archivage – est décrit en termes de standards ; cette pérennité est donc incompatible avec une représentation des informations et connaissances médicales qui dépendrait d'un outil ou d'un logiciel¹⁷.

7 Conclusion

L'interopérabilité en santé suscite des démarches normalisatrices qui peuvent sembler incohérentes. Nous espérons avoir montré ici que, d'une part, ces propositions ne sont pas incompatibles et que, d'autre part, elles portent en elles des philosophies d'échange et de partage d'informations différentes. Nous pouvons en conclure que c'est l'appropriation des organisations sous-jacentes par les acteurs professionnels – singuliers, les praticiens, ou collectifs, les réseaux – ou utilisateurs finaux – les patients – qui feront que le modèle d'échange d'une norme s'imposera naturellement. Il est d'ailleurs très probable que – à situations organisationnelles différentes, SIS différents – aucune norme ne s'impose franchement par rapport à une autre mais que les modèles d'échange dont elles sont porteuses se déploient de façon privilégiée en lien avec telle ou telle organisation.

Par rapport à cette vue un peu idyllique, il est à espérer que, au contraire d'autres domaines, des situations de monopole ne se fassent pas jour dans la mise en place des SIS, au moment où l'exigence d'ouverture de la concurrence impose ses visées normalisatrices.

¹⁶Le problème de l'identifiant unique du patient, s'il est résolu au sein d'un hôpital et appréhendable au sein d'un réseau de santé, est encore un sujet de discussions au niveau interhospitalier ou national pour des problèmes de sécurité et d'éthique.

¹⁷On peut aussi noter que si les balises XML ont un sens pour un système formel qui va les utiliser – *i.e.* un programme – elles ont aussi potentiellement un sens pour un lecteur humain. Ce qui a fait dire à ses concepteurs que XML est « machine readable and human readable » (cf. fig. 3.1).

Crédits

Les travaux développés ici doivent beaucoup à des discussions au sein du groupe de travail GT11 de EDISANTÉ et du projet RNTS « EDI, données cliniques ». Ces discussions ont été rendues spécialement riches par Norbert Paquel, délégué général de EDISANTÉ et Emmanuel Cordonnier et Bernard Gibaud sont à l'origine des travaux sur le concept d'enveloppe.

Nos travaux sur le réseau de santé de périnatalité autour de l'hôpital de Poissy nous ont aussi permis de confronter l'existence de normes à un cas réel. Les discussions que j'ai eu, dans ce contexte, avec Michel Daigne et Jean-Christophe Weis ont été elles aussi riches d'enseignements.

La gestion des connaissances médicales

Ce chapitre doit d'abord nous permettre de revenir aux origines de la gestion des connaissances en tant que problématique. Ces origines sont à retrouver au sein des entreprises et, pour le côté académique, chez les chercheurs en sciences de gestion. Cela nous permettra de rechercher un certain nombre de points communs et de convergences avec l'Ingénierie des connaissances. Ceci fait, nous nous intéresserons aux liens entre la santé et la gestion, plus précisément, les rapports de l'hôpital avec la gestion et nous repositionnerons les enjeux de la gestion des connaissances médicales et leur spécificité au regard des problématiques développées dans les chapitres précédents. Les projets comme les développements normatifs sont resitués dans un contexte d'intégration. Enfin, nous essaierons de tirer les perspectives et développements futurs attendus, en particulier au regard des nouvelles structures de soins qui se font jour comme les réseaux de santé.

1 Les connaissances de la gestion

1.1 Les connaissances organisationnelles

Pour traiter la question de la gestion des connaissances, il est nécessaire de s'intéresser au point de vue de la gestion sur la question et donc d'étudier comment les connaissances sont abordées par celle-ci. Nous prendrons principalement pour support la synthèse de F. Blackler (1995) qui fait une revue exhaustive de la littérature sur la connaissance en gestion et son évolution. Nous ferons un résumé succinct de ce travail dans le but d'exploiter ses définitions de la connaissance et ne reciterons pas tous les travaux qui y sont analysés. Nous renvoyons le lecteur intéressé ou familiarisé directement à l'ouvrage (Blackler, 1995).

F. Blackler adopte une perspective historique pour démontrer ou proposer une évolution du point de vue de la gestion sur la connaissance. Ainsi, le courant de *l'apprentissage organisationnel* qui décrit des phénomènes de transfert ou de production des connaissances dans le cadre de l'action collective propose une vue statique et localisée de la connaissance. C'est à partir de ces différentes localisations ou type de savoir que la connaissance sera produite et échangée. F. Blackler propose alors cinq « lieux » ou types de savoir :

Embrained (formalisé, intellectualisé). C'est une connaissance créée par une activité de conceptualisation et d'abstraction. Elle permet, au contraire de connaissances routinières, de comprendre des causalités complexes et de développer des comportements en réaction à de nouvelles situations.

Embodied (savoir-faire corporel, pratique). C'est la connaissance dans le geste. c'est une connaissance qui dépend d'un contexte précis et n'existe que dans celui-ci. Cela correspond à un type d'organisation où l'expert est crucial : c'est celui qui sait.

Encultured (partagé, négocié). C'est la connaissance dans un processus de socialisation et de négociation. Le langage est un des vecteurs importants dans ce contexte. Cela correspond à un type d'organisation où la communication et la collaboration sont importantes. C'est le lieu du travail coopératif assisté par ordinateur (TCAO ou, dans sa version anglo-saxonne, CSCW).

Embedded (mémorisé dans des routines) C'est la connaissance mémorisée dans des routines, dans des procédures formelles.

Encoded (dans des symboles). C'est la connaissance encodée dans les programmes informatiques. C'est le statut de cette connaissance « encodée » dans un ordinateur qui a justement provoqué des réflexions remettant en cause cette classification.

F. Blackler argumente ensuite que ce découpage n'est plus opérationnel dans le contexte des technologies de l'information et de la communication et que, par exemple, une connaissance vue comme *embodied* dans un geste est mal différenciée d'une connaissance encodée si l'outil est l'ordinateur. Plus, l'ordinateur, le support numérique n'est pas du tout neutre vis-à-vis des connaissances et est un médium de leur organisation. Il agit sur les savoir-faire en requérant de nouvelles compétences cognitives, créant de nouvelles tensions dont l'origine (le changement de support) n'est pas nouvelle. Par ailleurs, l'unité d'analyse ne peut plus être les individus ou les organisations mais des systèmes d'activités socialement distribuées. *La connaissance est multi-facettes, complexe, située, abstraites, implicite et explicite, distribuée et individuelle, physique et mentale, dynamique et statique, verbale et symbolique* (Blackler, 1995). Il semble alors beaucoup plus justifié de s'intéresser aux connaissances – et donc de les spécifier – par rapport aux situations de création de savoir, c'est-à-dire vis-à-vis de l'activité dans laquelle le savoir est mobilisé¹. Les différentes caractéristiques de création de connaissance (*knowing*) et donc d'apprentissage sont alors :

L'apprentissage en tant qu'activité médiée. Les systèmes d'activités, anciennement considérés comme séparés se développent, se complexifient et sont en interrelations les uns avec les autres. Les apprentissages sont alors, entre autres, médiés par des supports, des outils.

L'apprentissage en tant qu'activité située. Cette caractéristique met l'accent sur le fait que les acteurs de l'organisation interprètent le contexte dans lequel ils agissent et que les communautés de pratique jouent un rôle fondamental dans l'acquisition des compétences.

L'apprentissage en tant qu'activité non figée. C'est le fait que les modes de développement des savoirs sont provisoires et en perpétuel développement. Développements que l'on n'est pas toujours capable d'anticiper, comme l'arrivée des NTIC en donne un exemple.

L'apprentissage en tant qu'activité pragmatique. C'est l'idée que l'action collective est pilotée par la conception qu'ont les gens de l'objet de leur activité.

¹Nous avons presque totalement fait disparaître l'argumentation car il n'est pas dans notre propos de balayer toutes les théories et recherches en gestion de ces vingt dernières années. Le lecteur intéressé lira (Blackler, 1995) et (Le Masson, 2000).

L'apprentissage en tant qu'activité contestée. C'est l'idée que derrière le développement technique et les changements économiques il y a des relations de domination/subordination en évolution.

Cette description d'une évolution dans la façon d'appréhender la connaissance dans le domaine de la gestion peut se compléter d'une description du concept de gestion des connaissances dont les types de formulations reposent sur une évolution similaire comme nous allons le voir maintenant.

1.2 Une solution *versus* un symptôme

Dans une revue de la littérature sur la gestion des connaissances pour l'innovation, P. Le Masson (2000) explore la notion de gestion des connaissances qui correspondrait moins à une solution qu'au symptôme d'une crise des savoirs ; des savoirs de production, des savoirs de conception et finalement une crise de la gestion stratégique. Des crises où la gestion des savoirs est au cœur de la question, comme face cachée du paradigme de la décision (Le Masson, 2000).

De façon analogique à l'évolution de la définition des connaissances et des contextes de création des connaissances, il propose trois vues de la gestion des connaissances :

1. **Isoler la connaissance pour la gérer.** Dans cette première acceptation, l'information est un modèle de relations entre les acteurs, les types d'acteur dépendant du savoir qu'il détient. C'est dans ce contexte que la question des connaissances tacites *versus* explicites est rendue célèbre par I. Nonaka et H. Takeuchi (1997) (*cf.* § 3.2).
2. **Gérer les situations d'apprentissage plutôt que la connaissance.** Dans cette deuxième acceptation, ce sont les situations d'apprentissages qui sont considérées comme intéressantes plutôt que les caractéristiques des connaissances. La problématique porte sur les formes de production collective de connaissances. La gestion de ces situations amène à concevoir des outils de gestion des ressources humaines ou de réorganiser les locaux pour favoriser cette production.
3. **Gérer la conception des savoirs.** Dans cette dernière acceptation, on s'intéresse à la gestion des savoirs dans l'innovation, à la coévolution des savoirs et de l'action – *i.e.* la conception. Dans ce contexte, on met en parallèle un modèle de l'organisation ou du produit donc de son évolution et de celle des connaissances qui lui est concomitante.

Il n'y a pas de réelle exclusive entre ces trois approches : il faut plutôt voir chacune comme une évolution développée en raison des incapacités de l'approche précédente à prendre en compte certains phénomènes. Ainsi, la question des connaissances tacites *versus* explicites de I. Nonaka et H. Takeuchi (1997), si elle exemplifie la première approche, est présentée par les auteurs, au sein d'un contexte où les connaissances sont (proposées comme étant) souvent recyclées au sein d'un processus de création faisant la part belle à la communication et au langage. Un tel point de vue trace le chemin d'une connaissance appréhendée de façon dynamique dans des processus d'apprentissage (Blackler, 1995).

La dernière acceptation de la gestion des connaissances est fondée sur des travaux de A. Hatchuel et B. Weill (1999) qui proposent d'étudier le processus de conception selon deux points de vue, celui des connaissances et celui des concepts. Mobilisant G. Simondon sur la genèse des objets techniques (Simondon, 1958), ils proposent de voir la création de connaissances et la création de concepts dans deux espaces duals, évoluant en parallèle avec des transferts de connaissances : durant un processus de conception, les connaissances évoluent, permettent de réfléchir à des concepts – *i.e.* des objets – en évolution, non stabilisés. Les connaissances évoluant dans l'espace des connaissances, les concepts dans l'espace des concepts. Puis le concept – *i.e.* l'idée de

l'objet – se stabilise, les connaissances se stabilisent, et l'objet existe en tant qu'objet statique avec ses contraintes et ses usages.

Un autre point de vue sur les connaissances et leur usage dans l'entreprise est fourni par Ph. Lorino (1995) qui s'intéresse au pilotage de cette dernière. Il constate que dans un monde en perpétuel changement avec des systèmes aux frontières jamais stabilisées, il n'est plus possible de contrôler l'entreprise, au sens propre du terme, à savoir obtenir des grandeurs mesurables ou interprétables, agir dessus, vérifier le résultat de l'action. Ph. Lorino propose alors de constater que l'entreprise est « pilotée » par de multiples acteurs, chacun ayant des connaissances dans et au sujet de l'entreprise et ayant chacun une autonomie cognitive. Cette autonomie est une marge d'interprétation pour l'action².

1.3 De la gestion à l'ingénierie : convergences épistémologiques

Alors qu'il nous semblait légitime de s'intéresser à la gestion en raison de notre intérêt pour le développement de SBC qui devaient être utilisés dans une organisation, on s'aperçoit que la gestion et l'Ingénierie des connaissances entretiennent des rapports étroits et des convergences de vue plus intimes qu'il n'y paraît au premier abord. Ces convergences ne sont pas étonnantes dans la mesure où les deux domaines ont un objet d'étude commun, la connaissance. À ce sujet, on aurait pu penser que l'expérience des systèmes experts développée au sein de l'IA et qui portait spécifiquement sur les connaissances aurait créé des synergies, au moins sur les critiques épistémologiques. Mais en dehors de quelques travaux³, il ne s'est pas vraiment créé de champ d'étude commun assumé. Pourtant, des réflexions épistémologiques partagées sont possibles (cf. chap. 2, § 2.1.1) et une histoire commune pourrait s'écrire, comme nous allons l'étudier maintenant :

La nature des modèles. Les modèles de l'Ingénierie des connaissances sont des modèles d'artefacts. Ce n'était pas évident au premier abord, puisque l'Ingénierie des connaissances héritait des approches cognitivistes (modélisation des raisonnements humains) et voyait l'acquisition des connaissances comme un « goulet d'étranglement » lié à la grande quantité de connaissances qu'il fallait « tirer » de l'expert pour construire le SBC (cf. chap. 1, § 1). La gestion a eu la même tendance en envisageant les organisations elles-mêmes comme un système d'interprétation collectif et en en caractérisant le fonctionnement avec des concepts liés à la cognition : c'est le cas par exemple de C. Argyris et D. Schön (1978) qui ont proposé le concept d'*organisation apprenante*. Celle-ci est idéalement définie comme sachant limiter ses routines défensives et capable de ce fait d'un apprentissage organisationnel en « simple boucle » (modification du comportement en fonction de l'environnement) et en « double boucle » (modification du comportement et de l'organisation elle-même via ses représentations – Charlet *et al.*, 2001). Pour conclure temporairement, on peut alors noter que : (1) l'Ingénierie des connaissances modélise des artefacts, les SBC, pour lesquels une approche cognitive n'a plus lieu d'être et (2) la gestion modélise, à plus grande échelle, une organisation impliquant des artefacts (systèmes d'informations, SBC, . . .) et des êtres humains. Ces derniers, êtres humains et collectifs, ressortissant possiblement à une approche cognitive.

Les situations d'apprentissage. Les caractéristiques des situations d'apprentissages sont pour partie convergentes avec nos théories de l'Ingénierie des connaissances

²Comme nous l'avons déjà noté en citant Ph. Lorino au chapitre 2 (cf. § 2.1.2, n° 1). Sans résumer le livre, nous pouvons encore noter que cette interprétation présente au sein du pilotage de l'entreprise met à mal les représentations universalistes de l'entreprise et qu'il est nécessaire de se forger une théorie de l'action.

³On peut, à l'opposé, citer le travail de H. Hatchuel (1992), qui a mené une étude critique des systèmes experts dans l'organisation en tant qu'outils de gestion.

(cf. chap. 2). Ainsi, le fait que les apprentissages soient *médiés* correspond à des connaissances se développant dans un environnement technique et s'inscrivant sur un support, le fait qu'ils soient *situés* correspond au contexte d'utilisation des connaissances et enfin, la caractéristique *pragmatique* des apprentissages correspond à l'irruption de la tâche comme critère de modélisation des connaissances. En abandonnant la caractéristique *contestée* de la connaissance et les enjeux de pouvoir que nous laissons à la gestion, on peut noter que la question d'un apprentissage *non figé* est un apport de la gestion à l'Ingénierie des connaissances. *Cette caractéristique met en lumière la question de l'évolution des connaissances que l'Ingénierie des connaissances a commencé à réfléchir malgré son penchant pour des connaissances stabilisées propres à être modélisées, poussée par des applications comme le Web sémantique, impliquant une réflexion sur les ontologies et leur évolution.*

Les ontologies. Sans que ce soit un apport déterminant à ce jour, il est intéressant de rapprocher les théories des objets techniques de Simondon, utilisées dans la troisième acception de la gestion des connaissances, *la gestion de la conception des savoirs*, de la question des ontologies et de leur méthodologie de construction. Deux points semblent intéressants à fouiller : (1) la question des usages, partie prenante de la genèse des objets techniques comme de l'Ingénierie des connaissances et (2) le caractère phylogénétique de cette gestation, qui rejoint la question des classifications en sciences naturelles (cf. chap 4, § 1.2.2). Pour ce dernier point, il ne semble pas y avoir d'urgence à organiser nos ontologies en suivant un paradigme évolutionniste mais rien ne dit que cela ne soit pas une façon de faire appropriée dans certains cas et pour le futur, en dehors des sciences naturelles évidemment (cf. chap. 4, § 1.2).

La question de l'interprétation. Il n'est pas besoin de développer beaucoup cette convergence qui nous a déjà servi dans nos réflexions épistémologiques auparavant (cf. chap. 2, § 2.1.2) mais *l'interprétation pour l'action* est un concept que nous devons à la gestion et donc à un de ses auteurs (Lorino, 1995) et *la prise en compte de l'action collective au sein des organisations* nécessite de réfléchir la mise en place d'un SBC par rapport à un collectif d'acteurs là où l'Ingénierie des connaissances réfléchit, généralement pour des raisons de modélisation, à un utilisateur unique.

Arrivés à la fin de cette étude des convergences et divergences, il faut noter qu'il n'est pas aisé d'affirmer que les apports sont de tel ou tel domaine. Il n'empêche que les convergences sont nettes et qu'elles nous confortent dans notre travail de réfléchir l'un et l'autre des domaines en synergie quand il s'agit de penser les SBC dans une organisation. Nous essaierons d'en tirer quelques conclusions pratiques après avoir abordé les relations de la gestion avec l'hôpital.

2 L'hôpital et la gestion

2.1 Les affres du pilotage organisationnel hospitalier

Le pilotage organisationnel nécessite la disposition d'un certain nombre de grandeurs d'analyse, leur interprétation, la volonté (ou possibilité) d'agir en fonction des interprétations et enfin la possibilité de contrôler les conséquences des actions (Lorino, 1995). Dans le cadre de l'hôpital ou, plus largement, du système hospitalier, la question de la mise en œuvre de ce pilotage organisationnel est récurrente. Dans une recherche-intervention, O. Lenay (2000; 2001) s'intéresse à la place d'un outil de gestion⁴(le seul existant dans le domaine hospitalier), le programme de médicalisation du système d'information (PMSI), dans la conception de la politique hospitalière française. Ce travail

est suffisamment intéressant dans notre contexte pour que nous nous y attardions.

Sans revenir sur l'historique et la genèse du PMSI, nous pouvons rappeler, en reprenant notre note 25, p. 66, que le PMSI vise à introduire des concepts de comptabilité analytique dans la gestion administrative des hôpitaux : les diagnostics et actes effectués dans un établissement de santé sont codés et comptabilisés, rapportés à un patient et aux différents coûts de la structure. Cela permet ainsi de bâtir des indices de coûts relatifs par groupe homogène de malades.

Le sujet abordé par la thèse est alors le suivant :

Au cours de sa genèse, la philosophie d'usage de l'outil [PMSI] a évolué, d'une perspective de médicalisation du contrôle de gestion à celle de la mise en place d'une politique de régulation. Dans cette dynamique, il faut voir l'influence de modèles économiques développés autour du concept d'incitation. Mais cette politique de régulation demande à être complétée, en prenant pleinement en compte le statut de l'outil de gestion, non seulement vecteur d'incitation mais aussi outil d'apprentissage. Envisagé sous la forme d'un couplage dynamique entre incitations et apprentissages, l'outil de gestion PMSI semble entrer dans un processus de concrétisation. Mais constitue-t-il un savoir actif, c'est-à-dire un savoir permettant de dépasser les règles d'un jeu politique largement critiqué? Permet-il de définir les agences régionales de l'hospitalisation comme des acteurs vraiment nouveaux? (Lenay, 2001).

Les résultats de cette étude sont multiples. Nous allons en balayer les principaux de notre point de vue⁵ :

- *Le PMSI, outil d'incitation, propose une vision de ce qu'est l'hôpital, de son mode d'action, de sa rationalité, mais aussi du fonctionnement du système hospitalier dans son ensemble. Cette modélisation est critiquable a priori et par conséquent difficile à appliquer telle quelle (Ibid.).*
- *L'auteur défend que malgré son inadéquation et parce qu'il est présent, cet outil peut générer des apprentissages. Ces apprentissages sont d'abord la nécessaire amélioration des descriptions de l'activité que sous-tend l'outil (consultations, activités de recherche, certaines activités directement médicales, soins infirmiers, ...). Ce sont ensuite des apprentissages stratégiques où l'établissement hospitalier modifie sa politique d'accueil et de transfert pour augmenter sa productivité (Op. Cit., p. 65).*
- *L'auteur développe par ailleurs une intervention montrant que des réseaux de soin (au sens propre) se créent entre structures hospitalières et sont cachés, non structurés, ni reconnus ni même connus, mais structurant pour l'activité de soin (Op. Cit., p. 193). Dans le même principe mais par des incitations financières, des « réseaux de santé ville-hôpital » se mettent en place. Cherchant à mieux coordonner les activités de soin autour d'une pathologie, ils sont (ou seront) évalués à cette aune et ressortissent à la *planification*.*
- *Ainsi, il apparaît que l'hôpital, l'établissement de soins au sens large, obéit à deux formes de gestion sanitaire différentes, la *régulation* incarnée par le PMSI et la *planification* incarnée par les réseaux, difficiles à concilier. O. Lenay propose alors de mettre en place une nouvelle forme de *gouvernementalité* qui dépasse les deux précédentes en définissant un nouvel objet de gouvernement, les *trajectoires de patients*. Cette troisième forme qui, à partir de catégories de patients et de coûts*

⁴Outil est utilisé ici dans un sens large propre aux sciences de gestion : est outil de gestion un artefact utilisé par des personnes ou un collectif d'une entreprise dans un tâche de gestion. L'outil de gestion est un modèle de l'organisation ou d'une partie de l'organisation. Il constitue alors une représentation de l'action organisée. Cet outil peut-être un tableau de comptabilité qui sous-tend une vision comptable de l'entreprise ou le PMSI qui sous-tend (plus, implique) une vision régulatrice de l'activité médicale hospitalière. O. Lenay note en plus à ce sujet que les fonctions assignées à un outil ne sont pas stables dans le temps et qu'il faut plutôt envisager sa genèse pour les comprendre (Op. Cit., p. 51-2).

⁵Cette liste n'engage que nous. Elle est le résultat de notre *interprétation* de ce travail en fonction de nos buts. Le travail de O. Lenay est évidemment bien plus riche que cela.

moyens de ces catégories, amorce des processus de décision qui combinent spécification de l'offre, orientation des malades et responsabilisation des organisations concernées, nous paraît se ranger spontanément sous le terme d'organisation. Ce type de démarche où la cohérence médicale des cas pathologiques considérés est relativement forte associe très largement les médecins à l'élaboration des typologies et au fonctionnement du dispositif financier (Op. Cit., p. 226).

2.2 Enseignements : des SBC dans une organisation

Nous voulons ici tirer les enseignements de la gestion des connaissances vue par la gestion, avec trois formes d'acceptation (*cf.* § 1.2), et des difficultés du pilotage organisationnel de l'hôpital (*cf.* § 2.1). Ces enseignements ne sont plus de caractère fondamental comme les comparaisons précédentes (*cf.* § 1.3) mais ont pour but de nous donner des indications et des critères de pertinence pour les propositions de gestion des connaissances que nous donnons par la suite (*cf.* § 3.1).

Problématique sous-jacente à tout ce chapitre, l'intégration d'un SBC dans l'organisation peut maintenant être réfléchi en tirant partie des réflexions développées ici et des apports épistémologiques précédents (*cf.* chap. 2). En particulier, d'un point de vue organisationnel, on peut alors dire comme le souligne Le Moigne (1995) que « la connaissance implique un sujet connaissant et n'a pas de sens ou valeur en dehors de lui. ». Ainsi, on ne peut modéliser la connaissance sans se préoccuper des invariants sociaux-culturels que transporte le sujet connaissant, du contexte d'utilisation de cette connaissance (*cf.* chap. 2, § 2.1.2 n° 2), en particulier des interactions et coopérations d'acteurs dans lesquels elle s'insère et des contextes organisationnels ainsi créés. Ce qu'affirment ainsi les épistémologies constructivistes et que découvrent de leur côté d'autres approches théoriques comme la cognition située : les connaissances sont avant tout connaissances pour l'action (*cf.* chap. 2, § 2.1.2 n° 1), elles sont situées dans un contexte physique, socialisées dans un réseau d'acteurs, interprétées par les capacités perceptives autant que de raisonnement de ceux-ci et prennent place dans les organisations qu'ils ont construites (Charlet *et al.*, 2000a)⁶.

Il est intéressant de confronter les points de vue épistémologique et systémique développés jusqu'ici à un point de vue plus formel en remarquant 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. 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 alors⁷ :

- 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 technique, 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 serait 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 ou ONCODOC, chap. 3) rentrent dans cet autre paradigme.

⁶Cette citation rend mal compte du travail de synthèse effectué en collaboration avec Ch. Reynaud et R. Teulier, cette dernière ayant développé l'argumentation organisationnelle.

Nous verrons plus loin que ces deux tendances peuvent et doivent être développées de concert dans un système de gestion des connaissances.

2.3 Quelle gestion des connaissances médicales ?

Nous avons vu dans l'exposé des connaissances et de leur gestion, vues par les sciences de gestion, des exposés propres aux domaines concurrentiels où la réflexion organisationnelle a toujours été présente au service de la productivité. La crise des savoirs en gestion y a précédé la gestion des connaissances, en a même été la cause.

Dans le monde de la santé, plus précisément des établissements au sens large qui dispensent les soins, nous sommes dans un domaine non concurrentiel où l'innovation en gestion n'a, jusqu'à maintenant, pas sa place. En effet, une innovation en gestion n'est possible et surtout pérenne que si elle se concrétise dans un outil⁸. Or, le seul outil dont dispose l'hôpital à ce sujet, le PMSI, est fortement figé dans ses attendus et fonctionnement en tant qu'outil de régulation économique (*cf. supra*). C'est un des résultats du travail d'O. Lenay, de montrer qu'il y a une innovation possible en gestion hospitalière mais qu'elle ne se concrétisera qu'avec des politiques (ici publiques) d'incitations. L'innovation n'est effectivement possible que reconnue et encadrée par l'administration qui favorise ou non des idées, des types d'organisation, en fonction des incitations ou modes de remboursement pratiqués. Par exemple, les réseaux de santé, apparaissent sans incitation, puis ont une reconnaissance officielle par les incitations financières, le FAQSV (Fonds d'amélioration de la qualité des soins de ville) par exemple.

À un niveau plus micro, le concept de l'apprentissage organisationnel trouve, quant à lui, des résonances en médecine avec les travaux de M. Stefanelli (2002) qui y voit une façon de réfléchir et modéliser les procédures de travail et d'échange d'informations dans les unités de soins. Ce en quoi nous le suivons.

On peut alors proposer l'hypothèse que selon le point de vue où l'on se place, du plus macro au plus micro, les connaissances mises en œuvre apparaissent statiques puis de plus en plus dynamiques. Le processus de gestion des connaissances médicales suit la même caractéristique et les trois vues de la gestion des connaissances précédemment décrites (*cf.* § 1.2) sont adéquates à décrire ces vues de l'organisation à différentes « distances » :

Les connaissances gérées par l'administration. Nous sommes ici dans le point de vues des connaissances de gestion vues par l'administration, à travers l'outil de gestion PMSI. C'est la situation décrite par O. Lenay (*cf.* § 2.1) où le modèle est très statique, où l'innovation n'apparaît et se pérennise que quand les incitations administratives le veulent bien. La dynamique des connaissances n'a pas besoin d'être modélisées, les innovations étant rares. Ceci dit, les réseaux de santé sont une innovation récente et majeure (*cf. supra*) et elle demande à réfléchir et faire évoluer les outils de gestion. C'est dans ce contexte, qu'O. Lenay a proposé les *trajectoires de patients*. Cette question sera développée au paragraphe 5.

Les connaissances gérées par l'organisation médicale. Nous sommes ici au niveau de l'organisation de l'unité de soins, avec ses procédures liées à l'échange d'informations et de connaissances. C'est dans ce contexte que, par ses réflexions et outils, l'Ingénierie des connaissances peut apporter beaucoup : que ce soit dans son expérience de modélisation des objets du domaine – *i.e.* les ontologies – ou celle de la tâche ou encore dans le fait de savoir que recueillir la connaissance est en fait en grande partie la reconstruire ou encore dans ses liens avec le tra-

⁷Cette opposition nous a été suggérée par B. Bachimont que nous citons bien volonté ici. Par ailleurs, cette opposition technique-épistémologique est à rapprocher de l'opposition technique-conceptuelle de A. David (*cf.* § 4.3.2, « Des modèles de pilotage [...] »).

⁸Outil est ici entendu dans le sens des sciences de gestion (*cf.* note 4, p. 94).

vail coopératif. Tout cela nourri de nos réflexions sur les formes des SBC dans les organisations (cf. 2.2).

Les connaissances gérées par l'innovation médicale. Nous sommes ici dans le point de vue de l'innovation médicale en tant qu'elle va mobiliser de nouveaux savoirs, de nouvelles connaissances et donc de nouvelles manières d'agir. Un nouvel examen, un nouveau type de soins en sont des exemples, s'ils modifient les connaissances que la médecine peut avoir sur la personne soignée. Plus, ce nouvel examen, par exemple un nouveau geste, vient des connaissances des praticiens ou des infirmières : il y a bien co-conception des connaissances et du concept, ici le geste. Mais cette vue de la gestation et de la gestion des connaissances reste prospective, avec des outils, pour la supporter en médecine, qui restent encore à inventer.

Les trois points de vue de la gestion des connaissances ayant été replacés dans le contexte médical, nous allons revenir sur le deuxième, celui où les besoins d'échange de connaissances sont patents et où l'Ingénierie des connaissances peut et a commencé à agir.

3 La gestion des connaissances médicales

3.1 Les enjeux de la gestion des connaissances médicales

Nous nous plaçons ici délibérément dans le cadre particulier de l'hôpital puisque c'est notre lieu d'exercice professionnel et donc, ici, d'étude. Nous étendrons ensuite ces réflexions à tout l'établissement de soins, dans sa relation avec la médecine libérale⁹. Le contexte est donc celui des connaissances médicales qu'il nous faut maintenant repérer, éliciter et finalement gérer, en même temps que nous identifions les acteurs à l'origine de ces connaissances et de leurs échanges. Cela nous amènera alors à mobiliser les réflexions et outils développés jusque-là pour proposer un cadre technique global de la gestion des connaissances médicales.

Arrivé à ce stade, il est nécessaire de définir ce qu'est la gestion des connaissances. Nous avons, jusque-là, mobilisé des auteurs qui s'interrogeaient sur la notion de connaissance, parfois même sur ce que veut dire gérer cette connaissance en termes pratiques mais sans jamais proposer de définition fonctionnelle. Nous allons donc proposer deux définitions, la première de J.-L. Ermine (1996) et la seconde rapportée par M. Stefanelli (2002) :

La gestion des connaissances (déf. 1) : *vise à (1) rassembler le savoir et le savoir-faire sur des supports facilement accessibles, (2) faciliter leur transmission en temps réel à l'intérieur de l'établissement et en différé à nos successeurs et (3) garder la trace de certaines activités ou actions sur lesquelles on peut devoir rendre des comptes à l'avenir.*

Chaque direction [...] est chargée de définir, dans son domaine de responsabilité, ce qui doit être écrit et conservé et d'organiser cette conservation.

La gestion des connaissances (déf. 2) : *vise à améliorer les performances de l'organisation en permettant aux individus de capturer, partager et appliquer des connaissances collectives pour faire prendre des décisions optimales... en temps réel.*

Ces définitions proposent deux points de vues, en termes de moyens et en termes de buts, mais elles sont cohérentes et nous pouvons les instancier dans un milieu hospitalier – très contraint – en écrivant :

⁹Ce choix est surtout guidé par notre expertise et donc les sujets de réflexions que nous voulons aborder. Il ne nie absolument pas que, justement via les réseaux, la médecine libérale, avance plus rapidement dans la gestion des connaissances médicales. Elle tire alors la dynamique de la gestion des connaissances médicales hospitalières (cf. § 5).

La gestion des connaissances médicales (déf. de synthèse) : *La médecine est confrontée à des exigences de qualité et d'optimalité de soins qui obligent à gérer les connaissances médicales.*

Les outils proposés (en dehors de la régulation PMSI) sont la « traçabilité » des actes médicaux (exigence des tutelles), la protocolisation des soins (exigences médicales), le respect de Références médicales opposables¹⁰ (RMO – exigence médicale des tutelles).

Les acteurs de l'établissement de soins (médecins, infirmières, paramédicaux, etc.) doivent pouvoir partager et appliquer des connaissances collectives et donc collectivement validées, en particulier les RMO ou les protocoles de soins se développant dans de nombreuses disciplines sous la forme de guides de bonne pratique (GBP), pour la partie de la médecine la plus protocolisée. Ils doivent aussi acquérir des savoirs et des savoir-faire qui leur permettront d'utiliser leur relation clinique avec le patient pour « instancier », de la meilleure façon qu'il soit, des connaissances générales au contexte, toujours particulier, du patient.

Pour cela, il faut se donner les moyens de rassembler le savoir et le savoir-faire médical, les protocoles, les GBP, les références bibliographiques de la médecine factuelle¹¹ sur des supports facilement accessibles, faciliter leur transmission en temps réel à l'intérieur de l'établissement et en différé à toute personne habilitée à le demander (patient, autre médecin, ...).

Enfin, Il est nécessaire (exigence des tutelles) de garder les traces de toutes les données médicales d'un patient durant 50 ou 70 ans dans certains cas.

Cette définition posée, on a repéré de façon globale la plupart des problèmes qui se posent à nous. Avant de proposer comment les résoudre en utilisant les acquis de la réflexion que nous avons mené jusqu'ici, il est nécessaire de regarder les types de connaissances mis en jeu dans ce processus de gestion. Elles peuvent être ainsi partagées en deux grandes catégories liées à la spécificité de la médecine et au secret médical : (a) les informations nominatives et (b) les informations non nominatives. Dans la première on trouve des informations nominatives administratives (le « dossier administratif »), des informations nominatives médicales (le « dossier médical », enregistrement de tous les événements ayant trait au patient et le « dossier infirmier », permettant de piloter les traitements). Dans la seconde catégorie, on trouve les sources et les résultats d'études épidémiologiques, des conclusions bibliographiques sur les pratiques médicales (La médecine factuelle, cf. note 11, p. 99), des connaissances sur les conduites à tenir (les protocoles thérapeutiques, les GBP, les RMO). Enfin, les thésaurus médicaux qui, comme nous l'avons montré (cf. chap. 4, § 3.3), jouent un rôle fondamental en médecine. En dehors du fait que cette dichotomie soulève des problèmes de sécurité de façon différenciées, avec ceux de sécurisation importants pour les données nominatives, beaucoup moins pour les autres, elle trace surtout une frontière entre des connaissances médicales contextuelles (les connaissances nominatives sur les patients) et non contextuelles (les connaissances non nominatives). La façon de les prendre en compte va en dépendre fortement comme nous l'avons vu au début du chapitre 3 et nous auront des approches à tendance épistémologique pour les premières et à ten-

¹⁰Les références médicales opposables ou RMO sont l'un des outils conventionnels de maîtrise médicalisée des dépenses. Elles visent à éviter le superflu, voire le dangereux dans la prise en charge de certaines maladies ou la prescription de certains examens ou médicaments. Parce qu'elles prennent en compte l'ensemble des connaissances médicales et leur évolution, les références médicales sont appelées à être réactualisées régulièrement et complétées. En tout état de cause, les médecins demeurent entièrement libres et responsables de leurs prescriptions. Exemple de référence médicale : *Dans la plupart des cas il est inutile de répéter le dépistage du Cholestérol, s'il s'est révélé normal la première fois, avant 3 ou 5 ans selon le cas et en l'absence de pathologie, de traitement ou d'augmentation de poids* <<http://www.chez.com/mach02/>>.

dance technique pour les secondes (cf. § 2.2)¹².

3.2 Le dossier médical, objet focal de la gestion des connaissances

Nous argumentons ici que le dossier médical hypertextuel à la HOSPITEXTE (cf. chap. 3, § 2) est un bon outil de gestion des connaissances. Nous allons pour cela examiner ses propriétés au regard de la gestion des connaissances et plus précisément de la transmission des savoirs. Nous reprenons pour cela les arguments que nous avons développé dans les chapitres 2 et 3¹³.

Pour commencer, étant par définition le « lieu » de convergence des flux d'informations et de connaissances qui circulent dans l'hôpital au sujet d'un patient, le dossier médical est le point de passage obligé et, de par sa fonction d'archivage, le réceptacle de toutes les informations nominatives au sujet du patient. Par ailleurs, nous avons vu que la connaissance n'est accessible qu'à travers un support matériel, que celui-ci n'était pas neutre quant au contenu de cette connaissance et que le format documentaire électronique a une rationalité propre qui se traduit par la possibilité de « calculer » des tables de matières, index, nouveaux documents, etc., comme nous avons commencé à l'expérimenter dans HOSPITEXTE. Mais un document, *a fortiori* un dossier, prescrit par nature une manière de lire et d'appréhender son contenu. Cette capacité à lire un dossier médical est un savoir faire qui se transmet entre apprenants et apprentis, là où se crée le dossier, à l'hôpital. Ce savoir-lire est aussi un savoir-écrire. Ainsi, le dossier médical est le siège d'un apprentissage, qui plus est collectif dans la mesure où c'est une connaissance que les pairs essaient de transmettre collectivement aux apprenants. Ces apprentissages peuvent être statiques quand il s'agit d'apprendre l'usage – *i.e.* la lecture et l'écriture – de tel ou tel document stabilisé dans le temps ou dynamiques quand il s'agit de prendre en compte dans le dossier, de nouveaux documents ayant traits à, par exemple, de nouveaux traitements ou de nouvelles modalités d'imagerie.

Ainsi, point de passage obligé des connaissances ayant trait au patient et outil des apprentissages, le dossier médical hypertextuel peut être considéré comme l'objet focal de la gestion des connaissances.

¹¹L'Evidence-Based Medicine ou EBM, appelée en français la médecine fondée sur des preuves ou médecine factuelle désignait, au départ, une stratégie d'apprentissage des connaissances cliniques. Elle fait maintenant partie intégrante de la pratique médicale et consiste à baser les décisions cliniques, non seulement sur les connaissances théoriques, le jugement et l'expérience qui sont les principales composantes de la médecine traditionnelle, mais également sur des « preuves » scientifiques, tout en tenant compte des préférences des patients. Par « preuves », on entend les connaissances qui sont déduites de recherches cliniques systématiques, réalisées principalement dans le domaine du pronostic, du diagnostic et du traitement des maladies et qui se basent sur des résultats valides et applicables dans la pratique médicale courante.

[...]

L'EBM est une démarche qui nécessite de formuler clairement la question clinique, de rechercher les articles pertinents à ce sujet, d'évaluer validité et l'intérêt des résultats et, enfin, d'intégrer ces preuves dans la pratique médicale courante afin de répondre à la question posée au départ

En aucun cas, ces « preuves » ne peuvent remplacer le jugement et l'expérience du médecin, ce qui explique que la médecine factuelle complète la pratique médicale traditionnelle mais ne la remplace pas <<http://www.ebm.lib.ulg.ac.be/prostate/ebm.htm>>.

Cette pratique, en raison de résultats intéressants, se développe et a amené la création de formations des étudiants à « la lecture critique » d'articles. Mais il nous faut nous aussi garder notre sens critique et constater que l'EBM n'est pas sans poser des problèmes d'application qui dépassent le cadre de notre travail (Perrier & Similowski, 2002). Enfin, les GBP portant sur des conduites à tenir ou les RMO sont analogues à l'EBM dans la mesure où ils se veulent scientifiquement prouvés et peuvent être vues comme une EBM « compilée » (on trouvera de nombreux liens sur la question à la même adresse).

¹²Ce découpage demandera à être affiné puisque des GBP comme en propose ONCODOC (cf. chap. 3, § 3) ressortissent à une approche épistémologique.

¹³Dans notre argumentaire et pour ne pas alourdir notre propos, nous ne faisons pas, ici, de distinction entre dossier médical et infirmier et ne nous intéressons pas, dans ce contexte, au dossier administratif.

3.3 La transformation et le cycle de vie des connaissances médicales

À partir du dossier médical et de l'ensemble des connaissances que nous avons listées précédemment, on peut proposer de voir le flux des connaissances dans l'hôpital comme une boucle de transformations au sein de l'unité de soin avec des connexions vers l'extérieur comme le codage ou des transferts d'extraits de dossiers médicaux selon les modalités décrites au chapitre 5. On voit ainsi figure 6.1, le dossier patient au départ d'une boucle qui va vers les données cliniques (et cela demande de les extraire via un thésaurus appuyé sur une ontologie – *i.e.* un *thésaurus sémantique* (cf. chap. 4, § 6.3 et 6.4) et du traitement du langage médical) puis vers les résultats des études épidémiologiques (des données structurées). Ces études permettent de générer des RMO ou des GBP (que nous considérons comme du texte semi-structuré comme dans le cadre de ONCODOC) ou de permettre au médecin de pratiquer la médecine factuelle (EBM – cf. note 11, p. 99). Enfin, le dossier médical permet d'effectuer le codage des patients (toujours grâce à un thésaurus sémantique) ou de communiquer vers l'extérieur suivant les 3 modalités développées au chapitre 3 : visualisation Web, envoi d'enveloppe ou structuration de messages.

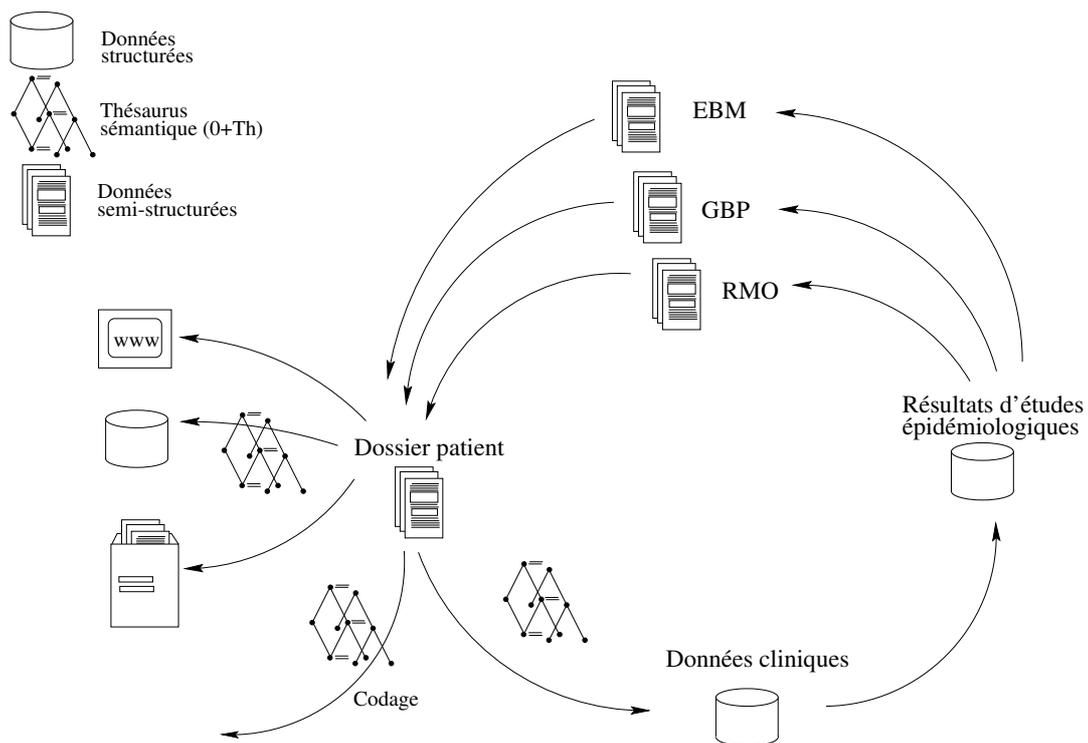


FIG. 6.1 – Les transformations et le cycle de vie des connaissances médicales.

Si les échanges décrits sur la figure 6.1 sont tous possibles si ce n'est réels, leur réalité globale va dépendre d'un certain nombre de développements, plus particulièrement de l'intégration de recherches dans un cadre globale cohérent. C'est ce que nous allons examiner maintenant en abordant les recherches à effectuer selon qu'elles ressortissent principalement à l'Ingénierie des connaissances, les systèmes d'informations ou les sciences de gestion.

4 Des perspectives de recherche

4.1 Perspectives dans le domaine de l'ingénierie des connaissances

4.1.1 Le dossier médical

Comme nous l'avons vu précédemment (*cf.* chap. 3, § 2.4), le dossier médical en tant que tel nécessite la mise en perspectives d'un certain nombre de réflexions. Nous allons les reprendre ici :

Une 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 le sens des documents au regard de leur position dans le réseau hypertextuel. Les balises SGML 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.

Des annotations. 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* et rejoignent la question des annotations pour le Web sémantique.

Le lien avec le TALM Le traitement automatique du langage médical (TALM) est, comme nous l'avons déjà vu, le compagnon naturel de l'approche documentaire : intervenant, dans ce contexte, pour identifier des portions de texte ayant des significations spécifiques, il est le moyen de passer de connaissances contextuelles non formalisées, les textes, à des connaissances formelles, par exemple les ontologies. Nous avons aussi vu qu'il permet d'envisager de nouveaux modes de navigation dans les documents mais de nombreux travaux restent à lancer pour mettre en œuvre ce qui n'est aujourd'hui que des perspectives de recherche.

4.1.2 Les ontologies

Nous avons longuement développé les tenants et les aboutissants des ontologies. Mais pour que celles-ci soient pleinement opérationnelles, un certain nombre de recherches sont encore nécessaires. Citons :

L'intégration globale d'une méthodologie de construction. Nous avons vu une méthodologie de construction des ontologies fondée sur des réflexions épistémologiques. En amont, nous avons défendu, dans une première étape, l'intérêt d'outils de TAL pour analyser les corpus nécessaires à l'élaboration des ontologies. En aval, nous avons abordé la questions des outils formels de spécification des ontologies. Ces étapes sont cohérentes les unes par rapport aux autres et il existe des outils prenant chacune d'elles en charge. Mais il n'existe pas d'outils les prenant en charge de façon globale et cohérente, en particulier entre les analyses de corpus produisant des réseaux terminologiques et les éditeurs d'ontologies supportant les méthodologies de construction. Ils restent encore à développer et font l'objet de projets en soumission à ce jour.

La maintenance et la fusion d'ontologies. Le Web sémantique est demandeur d'interopérabilité mais la médecine aussi. Or, il est bien possible qu'on ne puisse pas se mettre d'accord sur des ontologies couvrant des domaines entiers, que ce soit pour la médecine ou d'autres domaines. Par ailleurs, et particulièrement en médecine, les ontologies représentent des connaissances stabilisées dans des domaines évoluant rapidement. La question de l'évolution, la comparaison ou la fusion d'ontologies doit être posée dans le but de ramener des réponses satisfaisantes.

Thésaurus sémantiques et serveurs de terminologie. Les thésaurus sémantiques et les serveurs de terminologie qui associent thésaurus et ontologie, le second étant le squelette conceptuel du premier, partagent les mêmes buts, à savoir fournir une représentation normalisées des documents textuels impliqués dans le processus. Dans le domaine de la RI, cela est fait dans un but d'indexation. Dans le domaine médical¹⁴, un des usages importants est de fournir une représentation normalisée des données d'un patient à partir d'une description textuelle faisant déjà usage des termes du thésaurus. En collaboration avec des techniques du TALM (cf. 4.1.3), il permet de construire cette représentation pour des tâches de codage ou d'extrait de données cliniques. Si les technologies sont maintenant au point, les outils n'en sont, encore ici, qu'à leurs balbutiements. L'avenir et la diffusion de telles recherche passe probablement par le développement d'une plate-forme « serveur de terminologie » : à destination des professionnels des domaines dans lesquels elle serait développée – *e.g.* médecine ou droit – elle permettrait de leur fournir un certain nombre de services terminologiques et servirait aux chercheurs en retour à tester et valider leur recherche.

Langages de description et raisonneurs. Pour représenter les ontologies et raisonner avec, les principaux langages sont les graphes conceptuels et les logiques de description. Ces dernières semblent prendre le dessus en raison du choix des instances pilotant le développement du Web sémantique, grand pourvoyeur de projets ontologiques et de propositions de normes. Mais les choses ne sont pas si simple ; une ontologie contient une organisation hiérarchique des concepts mais aussi les propriétés associées qui doivent s'hériter ainsi que des connaissances inférentielles. Un ensemble un peu complexe n'est, si on n'y prend garde, pas calculable. Le choix d'un langage de représentation des ontologies puis d'un raisonneur qui agit dessus est la recherche d'un délicat équilibre entre puissance d'expression et calculabilité. Bien que ces choix puissent être de notre ressort, les travaux de normalisation autour des langages du Web sémantique vont probablement s'imposer à nous.

4.1.3 La représentation des connaissances médicales

Nous avons développé dans ce mémoire l'intérêt d'approches textuelles semi-structurées permettant de conserver le contexte d'expression des connaissances. À l'opposé, des objets formels sont nécessaires pour automatiser un certain nombre de tâches. Le noyau de cette formalisation étant les ontologies.

Par rapport à l'accès aux connaissances semi-structurées et le travail avec le dossier hypertextuel, nous avons déjà souligné la nécessité de réflexions autour d'une *sémantique hypertextuelle* et le possible outillage des réponses dans le contexte du Web sémantique avec ce que nous avons appelé des *métadonnées objectives* et des *métadonnées subjectives*. D'autres travaux nous ont aussi amené à réfléchir à l'appréhension des documents textuels en lien avec les images (projet HOSPIMAGE – <<http://www.biomath.jussieu.fr>> / « projets ») mais beaucoup reste à faire dans ce domaine.

¹⁴Qui inclut aussi les problématiques de RI mais ce n'est pas notre propos ici (cf. 4.1.3).

Quant au nécessaire lien entre une représentation semi-structurée et une représentation formelle (ontologie) il est du ressort de travaux en TALM (*cf.* § 4.1.1). Dans ce contexte, il s'agit d'être en mesure de mettre en relation la forme externe des énoncés (texte libre) et une représentation de leur contenu.

4.1.4 À propos du Web sémantique

Le Web sémantique n'est pas une solution à tous les problèmes de la gestion des connaissances médicales mais on a pu constater (*cf.* les chap.-§ 3-5, 4-7 et 5-6.2) que les problématiques abordées jusqu'ici pour les différentes composantes de la gestion des connaissances médicales rencontraient le Web sémantique dans la plupart de ses dimensions. Ainsi, la plupart des thèmes de travail de l'action spécifique « Web sémantique » (*cf.* note 11, p. 41) – les langages, les ontologies, les métadonnées, l'intégration de sources d'information, les services – ont des résonances en médecine.

Ceci étant, s'il n'est pas question de rechercher dans le Web sémantique toutes les solutions à nos problèmes, nous pouvons profiter, en médecine, à la fois des effets sur la normalisation d'un certain nombre de langages et des recherches plus fondamentales qui s'y développeront dans de nombreux champs thématiques (*cf. supra*). Les bénéfices attendus sont symétriques : d'un côté, le développement d'une vue globale de la gestion des connaissances médicales au sein du paradigme du Web sémantique, d'un autre côté, une participation de l'informatique médicale à la réalisation de ce même Web sémantique en médecine (Kamel Boulos *et al.*, 2001).

4.2 Perspectives dans le domaine des systèmes d'information

4.2.1 Les stratégies et perspectives de la normalisation

Avant que les SIS soient une réalité et, qu'en particulier, l'interopérabilité y soit assurée, il reste du chemin à parcourir. Nous avons développé dans ce mémoire des approches d'interopérabilité différentes et concourantes plutôt que concurrentes. Une norme en particulier, « l'enveloppe » avec ses faibles contraintes, nous paraît être un chemin intéressant vers l'interopérabilité. Il n'empêche que ce chemin est encore long d'un point de vue technique comme organisationnel pour aboutir à une interopérabilité de fait et envisager que se mette en place un secteur techniquement et économiquement viable.

Ainsi, si l'usage des technologies va se banaliser, il va falloir que les professionnels de santé aient les moyens techniques de gérer les flux de messages qui vont leur parvenir. Les solutions seront, comme nous avons commencé à le développer, à trouver dans les normes des « services Web » qui, avec leurs annuaires, protocoles d'échanges, répertoire de schéma, permettent d'automatiser autant que faire se peut les flux de données entre les professionnels et les établissements.

D'un point de vue organisationnel, il faudra que le secteur de la santé comprenne que sa survie économique (pour ce qui est des SIS) passe par l'interopérabilité, à l'opposé de toute démarche propriétaire. Les réseaux de santé qui se mettent en place sont un élément moteur de l'organisation des soins et poussent, par l'hétérogénéité des acteurs et comme nous avons pu le constater, les démarches d'interopérabilité (Paquel, 2002).

4.2.2 L'évolution du système d'information de santé

Le système d'information de l'hôpital et, plus largement, d'un réseau de santé (système d'information de santé ou SIS) est en constante évolution. C'est d'autant plus vrai que nous avons pu voir que les organisations de santé qui les sous-tendent sont en

pleine évolution, si ce n'est révolution. Le SIS est alors subordonné aux fréquentes variations de l'environnement et composer en permanence avec ces changements est une problématique récurrente.

L'urbanisation des systèmes d'information cherche à répondre à cette préoccupation en proposant des concepts d'organisation du système. Des travaux récents (Carvalho & Charlet, 2001; Carvalho, 2002) ont permis de définir « MIEU » (Méthodologie Informatique d'Évolution par l'Urbanisme), centrée sur la préoccupation d'évolution du système d'information en proposant une démarche complète de mise en œuvre opérationnelle. Les travaux réalisés l'ont été dans un contexte hospitalier réel, celui de l'Assistance publique-Hôpitaux de Paris.

4.3 Perspectives organisationnelles

4.3.1 Le travail coopératif

Le travail coopératif (TCAO pour Travail coopératif assisté par ordinateur ou CSCW pour *Computer supported Cooperative Work*) est un domaine de recherche à part entière. Les chercheurs dans ce domaine ont souvent opéré dans le milieu hospitalier, champ privilégié pour y étudier les coopération entre acteurs en raison des différents types d'intervenants et des roulements. Des travaux proposent des résultats conceptuels sur « le travail d'articulation » entre les collectifs impliqués (Grosjean, 2000) qui, comme souvent dans ce domaine, permettent d'envisager des changements organisationnels. D'autres, simulant les processus de coopération au sein d'un service hospitalier permettent d'envisager des modifications des processus de coopération et la mise en œuvre d'outils les supportant (Guigue *et al.*, 2000).

Dans un contexte où les technologies de l'information et de la communication se sont fortement développées, ces réflexions évoluent vers la façon de les utiliser pour modifier les organisations dans le but d'améliorer les collaborations.

Ainsi, le rôle des artefacts tend à être étudié dans leur fonction de coordination des collectifs, en particulier – et le CSCW est particulièrement au sein de la gestion dans ce contexte – sur la façon dont les utilisateurs se les approprient (Cicourel, 2000). Assister intelligemment le travail coopératif suppose de construire des modèles conceptuels de l'activité coopérative aussi complexes et aussi variés que ceux des raisonnements individuels. Des bibliothèques de modèles d'activité coopératives dans lesquelles l'Ingénierie des connaissances a un rôle déterminant à jouer restent à élaborer (Charlet *et al.*, 2001). Ces bibliothèques permettraient alors de construire de véritables SBC dédiés au travail coopératif dont il faut étudier des méthodes d'élaboration mais aussi la place et la fonction réelle pour la coopération (Lewkowicz & Zacklad, 2000). Ce type de travail est encore peu développé dans un environnement hospitalier même si les réseaux de santé commencent à susciter de telles recherches (*cf.* § 5).

4.3.2 Les systèmes d'information de santé à l'épreuve de la gestion

Nous voudrions rapidement aborder, du point de vue de la gestion, les conséquences de l'interopérabilité en médecine et de son niveau de développement, en particulier au sein des SIS. Nous reprendrons pour partie et de façon implicite les définitions et hypothèses développés par A. David (2001) et noter d'abord que au vu de la reconfiguration des organisations et du système de soins français, un SIS recouvre ou doit recouvrir un large réseau d'intervenants, des praticiens hospitaliers à la médecine de ville et aux professions paramédicales. Les conditions d'existences d'un SIS étant alors fortement liées à des problématiques d'interopérabilité. Dans ce contexte, nous tenterons de fournir des éléments de réponses à trois questions :

Les SIS sont-ils des éléments de l'organisation ? Comme nous l'avons défendu, des choix informatiques et normatifs sont liés à des types d'organisation. Le problème

dans le domaine de la santé est que l'avènement du Web, de XML, etc. – *i.e.* des NTIC/STIC – a modifié les conditions de gestation des SIS au point que l'informatisation de la médecine est à peine entamée en pratique et qu'il ne se dégage pas encore de mode d'organisation lié à un SIS. Dans ce contexte, en plein devenir, le SIS tarde alors à devenir un élément de l'organisation.

Le marché des SIS pousse-il à la conformation ? Des acteurs dispersés, des usages non encore stabilisés, une offre abondante et disparate, des normes en devenir, des compétences en développement sur le pilotage de projets ressortissant aux STIC. Tout ceci fait que comme précédemment, la réponse à la question est : « pas encore ».

A contrario, nous sommes dans des situations de concurrence ouvertes qui offrent les conditions nécessaires à la réussite de développements normatifs. Dans le futur, ces développements proposeront des types de formalisation normés, reliés à des organisations qui tendront à se stabiliser et à s'imposer en nombre fini à l'organisation réelle (*cf.* problématique de la normalisation, § 4.2.1).

Des modèles de pilotage pour la fabrication des SIS ? A. David (2001) propose de voir le développement d'un SI comme un *processus de définition formelle du SI*, la *formalisation*, et un *processus de réduction de la « distance » qui existe entre le SI et l'organisation qui doit l'utiliser*, la *contextualisation*. Chaque développement d'un SI étant plus ou moins d'un type ou de l'autre, avec des processus où la formalisation précède la contextualisation et inversement.

Ce point de vue se prête parfaitement à la fabrication des SIS : l'élaboration des SIS, en pratique surtout des systèmes d'informations hospitaliers, s'est faite jusqu'à récemment dans des démarches très formelles essayant de contraindre l'organisation à un moule prédéfini. Démarches qui ont souvent échoué dans la mesure où l'adéquation entre moule et organisation était difficile à réaliser. Dans le domaine en plein devenir des SIS, la contextualisation doit évidemment précéder la formalisation, surtout quand les problèmes d'interopérabilité sont en plein devenir au travers de la normalisation ou de la recherche. Plus, les futurs SIS, pour informatisés qu'ils seront, feront sûrement la part belle à la documentation structurée, par définition peu formalisée. Les fabricants de SIS – architectes, urbanistes – devront alors être attentifs à la phase de contextualisation, en particulier à l'intégration d'informations contextuelles au sein des composantes plus formelles du SIS.

5 Les réseaux de santé : de nouveaux modèles de fonctionnement

Les réseaux de santé sont apparus depuis le milieu des années 80. Comme souligné dans le travail de O. Lenay, ils ont débuté sans réelle reconnaissance. Depuis quelques années, ils ont une reconnaissance réelle, un soutien administratif et financier – sous forme d'incitations – réel. Leur définition officielle est : *Le réseau constitue, à un moment donné, sur un territoire donné, la réponse organisée d'un ensemble de professionnels et/ou structures, à un ou des problèmes de santé précis, prenant en compte les besoins des individus et les possibilités de la communauté* <<http://www.cnr.asso.fr/>>.

Combinant les exigences de soins de l'hôpital et des praticiens de ville, additionné de l'éloignement des différents intervenants, les réseaux sont des terrains d'expériences¹⁵ privilégiés qui cumulent de nombreuses difficultés techniques et organisationnelles mais sont porteurs d'espoirs sur la réorganisation des soins. Comme les organisations hospitalières, ils peuvent, dans ce contexte, profiter d'avancées de l'Ingénierie des connaissances pour leur fournir de nouveaux outils de travail ou de gestion. Ceci étant, les réseaux de santé ont des particularités qui donnent une certaine originalité

aux questions posées comme aux réponses à apporter. Nous allons en citer quelques unes, importantes de notre point de vue :

Les trajectoires de patients. La *trajectoire de patient*, nouvel objet de gestion, proposée par O. Lenay comme une forme de gouvernementalité liée à l'*organisation*. Cet objet n'est pas tout à fait nouveau et est déjà étudié par les économistes de la santé et les chercheurs de l'informatique médicale sous le nom de *trajectoire de soins*. Ce qui est important, c'est l'espérance qu'on peut avoir sur le fait que cet outil de gestion sera beaucoup plus médical que le PMSI en raison de l'organisation qu'il représente, beaucoup plus proche de la réalité médicale et donc des médecins. Cela ne veut pas dire une éviction de l'administratif mais bien une synergie entre la gestion de l'hôpital et l'activité médicale pour tendre, en élargissant au réseau de santé, vers une gestion médicale de l'établissement de soins. Il n'empêche qu'il reste encore du travail de modélisation sur l'objet lui-même au niveau, par exemple, de ce qu'est un *épisode de soins* (Naiditch, 2000) et que l'adhésion des professionnels à ce nouvel objet n'est pas assurée.

Coopération et coordination. Les problèmes de coordination au sein des réseaux de santé sont complexes et, dans bien des cas au début, mal résolus. Où est le patient ? Qui a les informations sur sa trajectoire ? Ils se compliquent, dans un paysage qu'il ne faut pas idéaliser, par des résistances liées aux changements organisationnels créés par le réseau (prise de rendez-vous centralisée, agendas partagés, autorisation d'accès aux données médicales, etc. – Grosjean, 2002). Les technologies de l'information et de la communication créent les conditions de la réussite de ces nouveaux modèles de soins mais ne la garantissent pas.

L'interopérabilité et les plates-formes. Nous avons vu que les questions d'interopérabilité sont un sujet de réflexion important en médecine. Avec les réseaux, ces questions sont encore davantage d'actualité. Quelle type d'interopérabilité choisir ? Quelles contraintes pose le type de connexion possible pour certains médecins dans certaines villes ? Comment organiser la sécurité ? Des réponses techniques existent, à ces questions comme à d'autres (cf. chap. 5), mais elles doivent être trouvées de façon cohérentes au sein du réseau. Réponses qui passent parfois par des plates-formes d'intermédiation comme à l'hôpital d'Armentières (Beuscart *et al.*, 1999).

La représentation des connaissances médicales. Nous avons défendu une approche documentaire pour les données nominatives liées au patient et le dossier patient en général, là où il y a des connaissances contextuelles à traiter. Cette proposition tient aussi pour les réseaux avec les mêmes arguments. Ainsi, si l'approche documentaire doit être déployée partout où il y a des textes, il n'est pas sûr que ce soit le cas pour tous les réseaux de santé. Si celui-ci a une vocation de veille sanitaire par exemple sur de la cardiologie ou du diabète, il est probable qu'il y ait peu de production de texte libre et plus de production de données médicales bien spécifiées (en fonction des buts épidémiologiques) : une approche sous forme de données structurées renseignées par des questionnaires préexistera alors à une approche textuelle. Il n'empêche qu'une approche textuelle devra toujours être réfléchie et envisagée, étant la seule à proposer une vue des données médicales comme une ressource générique qui sera possiblement accessible en temps utile et dans des usages que l'on n'aura pas toujours anticipés. Il faut ainsi toujours garder à l'esprit qu'abandonner les textes et les contextes oblitère toute possibilité d'évolution.

Des sujets de modélisation et de réflexion. Nous travaillons sur un réseau de périnatalité autour de l'hôpital de Poissy. Ce type de réseau pose des questions de

¹⁵Ce mot « expérience » ne doit pas oblitérer le fait qu'on est dans le domaine de la personne soignée et que les exigences de réussites sont fortes.

modélisation originales : d'abord, parce que les intervenants institutionnels, médicaux et paramédicaux sont de toutes sortes, médecins, infirmières, assistantes sociales, protection maternelle et infantile, etc. et qu'ils n'ont pas le même point de vue sur le patient (en l'occurrence la patiente). La mise en œuvre d'outils informatiques « intelligents » au sein du réseau passe par la définition d'ontologies dont la construction est compliquée par les différents points de vue (Weis, 2002). Ensuite, nous n'avons pas un patient mais un couple mère/enfant dont la santé doit être appréhendée de façon différenciée mais globale avant l'accouchement et qui, après celui-ci, forme deux personnes distinctes.

Ces particularités ajoutées à l'adhésion des médecins – bien que nous ayons vu qu'elle n'était pas automatique et exempte de résistances – créent des conditions de travail et de réflexions intéressantes. Et si rien ne garantit la réussite des réseaux, nous pouvons penser qu'ils sont l'avenir de l'organisation des soins, pour peu qu'un certain nombre d'outils, conceptuels comme gestionnaires, leurs soient fournis.

6 Conclusion et perspectives

Pour développer notre point de vue sur la gestion des connaissances médicales et les perspectives de sa mise en œuvre, nous avons essayé d'être exhaustif et nous nous sommes d'abord fondé sur notre propre expérience. Cette expérience n'est vraiment productive que si nous prenons du recul pour l'analyser et, dans notre domaine de l'Ingénierie des connaissances, pour développer une réflexion épistémologique sur les connaissances elles-mêmes. C'est de là que vient la structure de ce mémoire où nous avons alterné réflexions épistémologiques sur l'Ingénierie des connaissances (chapitres 1 et 2) la gestion et leurs convergences (chapitre 6) et recherches pratiques sur les dossiers médicaux (chapitre 3), les ontologies et l'indexation (chapitre 4), l'interopérabilité (chapitre 5) et finalement la gestion des connaissances médicales (ce chapitre).

En faisant cela, rajouté au fait que nous avons cité des travaux de personnes avec qui nous avons collaborées ou qui travaillent dans des paradigmes similaires aux nôtres, nous savons avoir balayé une grande partie des questions qui se posent autour de la gestion des connaissances. Malgré cela, nous sommes conscients d'avoir développé un point de vue forcément partiel et d'avoir cité sans les développer des points importants liés à la gestion des connaissances médicales comme la formation, l'aide à la décision et, dans un domaine plus prospectif, les rapports de l'innovation à l'informatique médicale :

La formation. Nous nous sommes intéressés aux apprentissages mais les nouvelles conditions de communication ont permis que se développent un certain nombre d'initiatives sur les apprentissages en présentiel *versus* à distance. Cela pose un certain nombre de problèmes pratiques et épistémologiques qui commencent à être bien réfléchis (Tchounikine, 2002). Dans le contexte de la médecine, les conséquences de ces nouvelles conditions d'enseignement sont encore à tester et un projet d'université médicale virtuelle francophone s'est mis en place et fédère les initiatives de recherche <<http://www.umvf.org>>.

L'aide à la décision. Nous n'avons abordé l'aide à la décision en médecine qu'en faisant part d'une approche textuelle et contextuelle, le système ONCODOC. Or, nous savons qu'il existe d'autres systèmes même si nous pensons qu'ils sont moins suivis et moins prometteurs mais, surtout, nous savons qu'il existe tout un courant de recherche sur l'aide à la décision qui aurait pu nous permettre de resituer plus précisément cette problématique au sein de la gestion des connaissances médicales. Ce travail reste à faire.

Innovation médicale et informatique médicale. Le troisième point de vue que nous avons développé sur la gestion des connaissances médicales faisait l'hypothèse d'une co-conception des connaissances et de concepts comme un nouvel examen

ou une nouvelle modalité d'imagerie. La génomique va très probablement modifier la pratique médicale à plus ou moins long terme. Plus largement, le concept de médecine moléculaire semble nourrir les réflexions de la communauté internationale de l'informatique médicale (Kulikowski, 2002) et il sera intéressant de voir comme la gestion des connaissances médicales et son appréhension de l'innovation peut nous aider devant ce nouveau défi et proposer des outils de gestion et de gestion des connaissances.

Finalement, il reste encore beaucoup à faire pour réaliser la gestion des connaissances médicales à l'hôpital ou dans le cadre, plus large, des réseaux de santé et nous n'avons pas, loin de là, couvert tous les champs de recherche à ce sujet. Ce sera l'occasion de futures collaborations indispensables, interdisciplinaires et enrichissantes.

Crédits

À la fin de ce dernier chapitre, nous avons encore à citer des collaborations sur des sujets récents. Charles Métivier a travaillé en son temps sur le projet HOSPIMAGE. C'est, par ailleurs, avec plaisir que je cite Americo Carvalho qui vient de passer, sous ma direction, sa thèse sur l'urbanisation des systèmes d'information et Michel Daigne avec qui nous avons réfléchi et réfléchissons encore à l'organisation du réseau de Périnatalité de Poissy. Enfin, les propositions autour de la gestion et la gestion des connaissances trouvent leurs sources dans de nombreuses discussions que nous avons eu avec Régine Teulier durant les séminaires « Ingénierie des connaissances et gestion » de ces dernières années.

Arrivé à la fin de ce mémoire, une conclusion technique ne s'impose pas : elle a déjà été faite et mise en perspective au chapitre précédent. En revanche, nous voudrions souligner et défendre l'approche qui a été la nôtre ici, à savoir développer nos réflexions et nos applications en les replaçant dans leur contexte épistémologique et scientifique. Cela nous a amené à écrire un document de synthèse sur les sujets de recherches qui ont été les nôtres dans les dix dernières années.

Les réflexions développées ici concernent d'abord l'Ingénierie des connaissances. Elles se sont enrichies de réflexions d'autres domaines, suscitées par les applications médicales et provoquées par des rencontres, souvent dans des séminaires du GRACQ¹⁶. Les domaines rencontrés ont été, dans l'ordre d'importance pour nous, l'épistémologie, la terminologie, la gestion, les systèmes d'information, la recherche d'information. D'autres rencontres sont encore à faire, en particulier avec les sciences sociales.

La question de la pluridisciplinarité, toujours annoncée, difficilement mise en œuvre a pourtant été d'une grande importance pour nous. Si nous prenons l'exemple de l'Ingénierie des connaissances et de la gestion, nous pouvons essayer de retracer nos motivations et notre démarche.

L'Intelligence artificielle au départ, l'Ingénierie des connaissances ensuite, en tant que spécialisation de la première, ont eu comme champs d'activité, avec les systèmes experts, l'entreprise ou de façon plus générale, des lieux de savoirs collectifs à destination d'une certaine production – *e.g.* l'entreprise déjà nommée mais aussi un hôpital (production de soins). Or, ce terrain était aussi celui de la gestion, ce qui lui donnait toute légitimité à étudier l'arrivée des systèmes experts pendant que les chercheurs en Intelligence artificielle ignoraient, pour la plupart, le champ de la gestion¹⁷. Avec l'évolution de l'Ingénierie des connaissances qui s'est attachée à faire fonctionner des SBC dans les entreprises ou autres puis la pluridisciplinarité organisée dans les séminaires du GRACQ (*cf.* note 16), la coopération Ingénierie des connaissances et gestion devenait évidente dans sa nécessité. À nous, en général et ici en particulier, à faire réellement fonctionner cette pluridisciplinarité en commençant par interroger le point de vue de chacun des domaines sur ce qui nous intéresse, des SBC dans une organisation et, dans le domaine applicatif de la santé, des SBC dans un établissement de soins. C'est ce que nous avons fait en comparant les points de vues de l'Ingénierie des connaissances et des sciences de gestion sur la connaissance et sa gestion. Cela nous a

¹⁶La communauté Ingénierie des connaissances a, dans sa très jeune vie (une dizaine d'année) toujours été très active. Nous avons participé et participons encore à la mise en place de séminaires de réflexions souvent pluridisciplinaires. Cela a pu se faire grâce au soutien du GdR I3 <<http://sis.univ-tln.fr/gdri3/>> et se poursuit aussi dans le cadre des actions spécifiques du CNRS, ciblées sur un sujet mais très profitables aux réflexions pluridisciplinaires elles-aussi.

permis, heureusement, de dégager quelques invariants. Mais, c'est dans les désaccords que l'interdisciplinarité est riche d'enseignements, pour peu qu'on fasse l'effort de les analyser et de rechercher, par exemple, les différentes façons d'appréhender un même objet d'étude. Ainsi, dans notre cas et pour les SBC, on peut citer deux apports intéressants de la gestion à notre réflexion « Ingénierie des connaissances » : (1) les situations d'apprentissage, qui nous interpellent sur l'évolution des connaissances alors que notre penchant est de nous intéresser aux connaissances stabilisées (au moins pendant un temps) pour pouvoir les modéliser ; (2) la prise en compte de l'action collective des organisations qui nécessite de réfléchir la mise en place d'un SBC par rapport à un collectif d'acteur là où l'Ingénierie des connaissances réfléchit, pour des raisons de modélisation, à un utilisateur unique¹⁸ ; enfin (3) les points de vues des sciences de gestion sur les connaissances et leur gestion se sont avérés extrêmement positifs pour notre réflexion de manière générale et pour les propositions qui en découlent en particulier.

Ceci dit et en reprenant un paragraphe du chapitre 6, il faut noter qu'il n'est pas aisé d'affirmer que les apports sont de tel ou tel domaine. À commencer par nos propres lacunes de chercheur maîtrisant (ou ayant au moins une idée argumentées de ce qu'est) l'Ingénierie des connaissances, moins à l'aise avec ses arguments épistémologiques, sûr d'avoir une vue plus étroite de la gestion, celle fournie par quelques rencontres, séminaires et lectures. Par ailleurs et par rapport à la pluridisciplinarité, nous savons qu'elle ne se décrète pas mais nous avons tendance à penser que nous travaillons dans un contexte réellement pluridisciplinaire qu'est l'Ingénierie des connaissances et l'étude des différents séminaires et journées « Ingénierie des connaissances » (appel à communication, sessions, publications) en témoignent.

Du côté des applications, nous nous sommes nourris d'applications médicales qui se renouvellent (projet HTSC de dossier médical hypertextuel à la HOSPITEXTE à l'hôpital d'Amiens et en collaboration avec G. Kassel (*cf.* § 2.4), projet P2VIE de prévention et prise en charge du vieillissement par l'information électronique proposant un portail de services administratifs et médicaux et de continuité de la prise en charge dans un contexte de réseau de soins de gériatrie (*cf.* note 10, p. 83), projet de réseaux de périnatalité (*cf.* chap. 5, « crédits ») et d'applications non médicales, par exemple des travaux sur les index qui ont des résonances avec la médecine.

Ainsi, nos recherches ont évolué, centrées sur l'Ingénierie des connaissances au début et s'ouvrant, dans une approche pluridisciplinaire, à d'autres domaines. Ouverture motivée par la complexité des problèmes abordés en informatique médicale ou dans d'autres champs applicatifs¹⁹.

¹⁷C'est un constat, pas une critique. La gestion ignorait de son côté l'Intelligence Artificielle en dehors de rares travaux (*cf.* note 3, p. 92).

¹⁸C'est seulement une tendance, des acteurs de l'Ingénierie des connaissances réfléchissent aux questions du travail collaboratif entre collectifs.

¹⁹Les actions spécifiques du CNRS que nous co-pilotons – « Web sémantique » et « Ingénierie des connaissances et santé » – sont aussi pour nous une source de rencontres pluridisciplinaires par excellence.

Références

- AFNOR (1987). *Vocabulaire de la documentation*. Paris : AFNOR. 2^e éd. mise à jour.
- ARGYRIS C. & SCHÖN D. (1978). *Organizational Learning*. Addison-Wesley.
- ASSADI H. & BOURIGAULT D. (2000). Analyses syntaxique et statistique pour la construction d'ontologies à partir de textes. In Charlet *et al.* (2000b), chapitre 15, p. 243–55.
- AÏT EL MEKKI T. & NAZARENKO A. (2001). Quel index pour le document électronique ? In *Actes du 4^e colloque international sur le document électronique*, Toulouse.
- AUSSENAC N. (1989). *Conception d'une méthodologie et d'un outil d'acquisition de connaissances expertes*. Thèse de doctorat, IRIT.
- AUSSENAC-GILLES N., BIÉBOW B. & SZULMAN S. (2000). Corpus analysis for conceptual modeling. In N. AUSSENAC-GILLES, B. BIÉBOW & S. SZULMAN, Coordinateurs, *Proc. of the Workshop "Ontologies and Text" associated to the 12th EKAW conference*, p. 13–27, Juan-les-Pins, French Riviera.
- AUSSENAC-GILLES N. & BOURIGAULT D. (2000). The Th(IC)2 initiative : Corpus-based thesaurus construction for indexing www documents. In R. DIENG, Coordinateur, *Proc. of the EKAW conference*, Juan-les-Pins, France.
- AUSSENAC-GILLES N. & CONDAMINES A. (1998). Bases de connaissances terminologiques : enjeux pour la consultation documentaire. In *Actes des premières journées du Chapitre français de l'ISKO*, Lille.
- AUSSENAC-GILLES N., CONDAMINES A. & SZULMAN S. (2002). Prise en compte de l'application dans la constitution de produits terminologiques. In J. LE MAITRE, Coordinateur, *Actes des 2^{ndes} assises nationales du GDR-13*, Nancy : Cépaduès, Toulouse.
- 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*.
- 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.
- BACHIMONT B. (2000b). L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle. In J. PETITOT & R. CASATI, Coordinateurs, *Au nom du sens*. Grasset.
- BACHIMONT B. (2001a). 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. (2001b). Modélisation linguistique et modélisation logique des ontologies : l'apport de l'ontologie formelle. In J. CHARLET, Coordinateur, *Actes des 5^{es} Journées Ingénierie des Connaissances*, p. 349–68, Grenoble, France.
- BENJAMINS V. R., FENSEL D. P. A. (1998). Knowledge management through ontologies. In *Proceedings of the 2nd Conf. On Practical Aspects of Knowledge Management*, Basel, Switzerland.
- BERG M. (1998). Medical work and the computer-based patient record : A sociological perspective. *Methods of Information in Medicine*, **37**(3).
- BEUSCART R., RENARD J.-M., DELARUE D. & SOUF A. (1999). Le réseau ville-hôpital : une nouvelle forme de communication entre professionnels de santé. *Informatique et Santé*, **11**.
- BLACKLER F. (1995). Knowledge, Knowledge Work and Organisations : an Overview and Interpretation. *Organisations Studies*, **16**(6), 1021–46.
- BOOSE J. H. & BRADSHAW J. M. (1987). Expertise transfer and complex problems : using aquinas as a knowledge acquisition workbench for expert systems. *International Journal of Man-Machine Studies*, **26**(1), 3–28.
- BOUAUD J. (1989). *K un Langage pour l'Implémentation d'Outils de Représentation des Connaissances*. Thèse de doctorat, Université Paris 7.
- BOUAUD J., BACHIMONT B., CHARLET J. & ZWEIGENBAUM P. (1994). Acquisition and structuring of an ontology within conceptual graphs. In *Proceedings of ICCS'94 Workshop on Knowledge Acquisition using Conceptual Graph Theory*, p. 1–25, University of Maryland, College Park, MD.
- 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., HABERT B., NAZARENKO A. & ZWEIGENBAUM P. (2000). Regroupements issus de dépendances syntaxiques sur un corpus de spécialité : catégorisation et confrontation à deux conceptualisations du domaine. In Charlet *et al.* (2000b), chapitre 17, p. 275–290.
- BOUAUD J. & SÉROUSSI B. (1996). Navigating through a document-centered computer-based patient record : a mock-up based on WWW technology. *Journal of the American Medical Informatics Association*, **3**(suppl), 488–492.
- 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.
- BOUAUD J. & ZWEIGENBAUM P. (1992). A reconstruction of conceptual graphs on top of a production system. In *Proceedings of the 7th Annual Workshop on Conceptual Graphs*, Las Cruces (NM).
- BOURDA Y. & HÉLIER M. (2000). Métadonnées, RDF et documents pédagogiques. *Cahiers GUTenberg*, (35-36), 37–52.
- 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, Coordinateur, *Ingénierie des langues*, Traité IC2, chapitre 9. Paris : Hermès.
- BOURIGAULT D. (2002). Analyse distributionnelle étendue. In *Actes de la 9^e conférence sur le traitement automatique des langues*, Nancy.
- BOURIGAULT D., AUSSENAC-GILLES N. & CHARLET J. (2003). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas. *Revue d'Intelligence Artificielle*. À paraître.
- BOURIGAULT D. & CHARLET J. (1999). Construction d'un index thématique de l'Ingénierie des Connaissances. In R. TEULIER, Coordinateur, *Actes des 3^{es} Journées Ingénierie des Connaissances*, École Polytechnique, Palaiseau, France.
- BOURIGAULT D. & FABRE C. (2000). Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de Grammaires*, (25), 131–51.

- BOURIGAULT D. & LAME G. (2002). Analyse distributionnelle et structuration de terminologie. application à la construction d'une ontologie documentaire du droit. *Traitement automatique des langues*, **43**(1).
- BOURIGAULT D. & SLODZIAN M. (2000). Pour une terminologie textuelle. *Terminologies Nouvelles*, (19), 29–32.
- BRACHMAN R. (1983). What IS-A Is and Isn't : An analysis of taxonomic links in semantic networks. *IEEE Computer*, **16**(10), 30–6.
- BRACHMAN R. J., MCGUINNESS D. L., PATEL-SCHNEIDER P. F., RESNIK L. A. & BORGIDA A. (1991). Living with Classic : When and how to use a KL-ONE-like language. In J. F. SOWA, Coordinateur, *Principles of Semantic Networks*, chapitre 14, p. 401–456. San Mateo, Ca. : Morgan Kaufmann Publishers.
- 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), chapitre 25, p. 407–22.
- CANIVET M. (1999). La nature humaine. Cours de philosophie, disponible à <http://www.icampus.ucl.ac.be/philolo/>.
- CARVALHO A. (2002). *Évolution du système d'information fondée sur l'urbanisation : application au contexte hospitalier*. Thèse de doctorat, Univ. Paris 6.
- CARVALHO A. & CHARLET J. (2001). Évolution des systèmes d'information : une théorie pour l'urbanisation. *Technique et science informatiques*. Soumis.
- CAUSSANEL J., CAHIER J.-P., ZACKLAD M. & CHARLET J. (2002). Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web sémantique. In B. BACHIMONT, Coordinateur, *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 233–52, Rouen, France.
- CAUVET C., RIEU D., FRONT-CONTE A. & RAMADOUR P. (2001). Réutilisation dans l'ingénierie des systèmes d'information. In C. CAUVET & C. ROSENTHAL-SABROUX, Coordinateurs, *Ingénierie des systèmes d'information*, chapitre 5, p. 115–47. Paris : Hermès.
- CAVAZZA M. & ZWEIGENBAUM P. (1994). A semantic analyzer for natural language understanding in an expert domain. *Applied Artificial Intelligence*, **8**(3), 425–453.
- CHANDRASEKARAN B. (1986). Generic tasks in knowledge-based reasoning : High level building blocks for expert system design. *IEEE Expert*, **1**(3), 23–30.
- CHARLET J. (1989). *LÉZARD : acquisition des connaissances et gestion de l'incertitude dans un système expert de seconde génération*. Thèse de doctorat, Université Paris VI.
- CHARLET J. (1992). ACTE : acquisition des connaissances par interprétation d'un modèle causal. *Revue d'Intelligence Artificielle*, **6**(1–2), 99–129.
- CHARLET J. (1993a). ACTE : a causal model-based knowledge acquisition tool. In David et al. (1993), chapitre 22, p. 495–516.
- CHARLET J. (1993b). Dossier "IA et médecine". *Bulletin de l'AFIA*, **14**, 22–48.
- CHARLET J. (2001). Ingénierie des connaissances : un domaine scientifique, un enseignement ? In J. CHARLET, Coordinateur, *Actes des 5^{es} Journées Ingénierie des Connaissances*, p. 233–52, Grenoble, France.
- 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., BOUAUD J. & ZWEIGENBAUM P. (1996a). Ontologie et réutilisabilité : expérience et discussion. In N. AUSSÉNAC-GILLES, P. LAUBLET & C. REYNAUD, Coordinateurs, *Acquisition et ingénierie des connaissances : tendances actuelles*, chapitre 4, p. 69–87. Cepadue-éditions.
- CHARLET J., BACHIMONT B., BRUNIE V., EL KASSAR S., ZWEIGENBAUM P. & BOISVIEUX J.-F. (1998). Hospitexte : towards a document-based hypertextual electronic medical record. *Journal of the American Medical Informatics Association*, **5**(suppl).
- 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, Coordinateurs, *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., KRIVINE J.-P. & REYNAUD C. (1996b). Causal model-based knowledge acquisition tools : Discussion of experiments. *International Journal of Human-Computer Studies*, **26**(1), 318–336.
- CHARLET J., REYNAUD C. & TEULIER R. (2001). Ingénierie des connaissances pour les systèmes d'information. In C. CAUVET & C. ROSENTHAL-SABROUX, Coordinateurs, *Ingénierie des systèmes d'information*, chapitre 10. Paris : Hermès.
- 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, Coordinateurs (2000b). *Ingénierie des connaissances : évolutions récentes et nouveaux défis*. Paris : Eyrolles.
- CHAUMIER J. (1988). *Travail et méthodes du/de la documentaliste : Connaissance du problème, Applications pratiques*. ESF, 3^e édition. mise à jour et complétée.
- CICOUREL A. V. (2000). What counts as data for modeling medical diagnostic reasoning and bureaucratic information processing in the workplace? *Intellectica*, **30**(1), 115–49.
- COLLEGE OF AMERICAN PATHOLOGISTS (1993). *Systematized Nomenclature of Human and Veterinary Medicine : SNOMED International*. College of American Pathologists.
- CONDAMINES A. & AUSSENAC-GILLES N. (2001). Entre textes et ontologies formelles : les bases de connaissances terminologiques. In M. ZACKLAD & M. GRUNDSTEIN, Coordinateurs, *Systèmes d'information pour la gestion des connaissances dans les organisations : recherches en cours et approches industrielles*. Paris : Hermès. À paraître.
- CORRUBLE V. & CHARLET J. (2002). Dossier « IA & médecine ». *Bulletin de l'AFIA*, (48), –.
- COTTE D. (1999). Le texte numérique et l'intériorisation des dispositifs documentaires. *Document numérique*, **2**(3-4). Numéro spécial « Les bibliothèques numériques ».
- CSA (1992). *Interopérabilité*. CSA, vocabulaire des technologies de l'information. Fichier html mis à disposition par le Centre de recherche en droit public et lu le 20.07.2002 à http://www.autoroute.gouv.qc.ca/loi_en_ligne/glossaire/g083.html%.
- DARMONI S.-J., LEROY J.-P., BAUDIC F., DOUYÈRE M., PIOT J. & THIRION B. (2000). CISMéF : a structured health resource guide. *Methods of Information in Medicine*, **39**(1).
- S. J. DARMONI & B. THIRION, Coordinateurs (1999). *CISMéF : pourquoi, comment*. CHU-Rouen. Page visitée le 6 avril 1999 à <http://www.chu-rouen.fr/ssf/cismef.html>.
- DAVID A. & PALLEZ F. (2001). Les systèmes d'information à l'épreuve de l'organisation. In C. CAUVET & C. ROSENTHAL-SABROUX, Coordinateurs, *Ingénierie des systèmes d'information*, chapitre 2, p. 23–60. Paris : Hermès.
- DAVID J.-M. & KRIVINE J.-P. (1989). Augmenting experience-based diagnosis with causal reasoning. *Applied Artificial Intelligence*, **3**(2-3), 239–248.
- J.-M. DAVID, J.-P. KRIVINE & R. SIMMONS, Coordinateurs (1993). *Second Generation Expert Systems*. Springer-Verlag.
- DC (1999). *The Dublin Core Meta Data Initiative*. Available at <http://purl.org/DC/>.
- DE KLEER J. (1987). Qualitative physics. p. 807–814. John Wiley and Sons, Inc.
- DE KLEER J. & WILLIAMS B. C. (1989). Diagnosis with behavioral modes. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, p. 1324–1330, Detroit, MI.
- P. DEGOULET & M. FIESCHI, Coordinateurs (1997). *L'interopérabilité des systèmes d'information de santé*. Conseil de l'informatique hospitalière de santé.
- DELOUIS I. (1993). *LISA : un langage réflexif pour la modélisation du contrôle dans un système à base de connaissances*. Thèse de doctorat, Université Paris-Sud.
- DOLIN R., SPACKMAN K., ABILLA A., CORREIA C., GOLD-BERG B., KONICEK D., LUKOFF J. & LUNDBERG C. (2001). The SNOMED-RT procedure model. volume 8, p. 139–143.
- EDISANTÉ (2001). *La messagerie et l'échange d'informations médicales*. Rapport interne, Edi-Santé. Available at http://www.edisante.org/documentation/documents_ES/ESDG_2001_004.pdf.
- ERMINE J.-L. (1996). *Les systèmes de connaissance*. Hermès.

- EVANS D., BROWNLOW N., HERSH W. & CAMPBELL E. (1996). Automating concept identification in the electronic medical record : an experiment in extracting dosage information. *Journal of the American Medical Informatics Association*, **3**(suppl), 388–392.
- FAVIER M. (1998). Système d'information. les cinquantes prochaines années pour les SI! In *Journées nationales des IAE*, p. 101–119 : LRSG, Univ. Nantes.
- FENSEL D., ANGELE A. & LANDES D. (1991). KARL : a knowledge acquisition and representation language. In *11th International conference Expert Systems and their Applications*, p. 513–525, Avignon.
- FERNÁNDEZ M., GÓMEZ-PÉREZ A., PAZOS J. & PAZOS A. (1999). Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent System and their Applications*, **14**(1), 37–45.
- FRÉNOT S. & LAFOREST F. (1999). Medical record management systems : Criticisms and new perspectives. *Methods of Information in Medicine*, **38**, 89–95.
- GANASCIA J.-G. (1998). *Le Petit Trésor. Dictionnaire de l'informatique et des sciences de l'information*. Édition Flammarion.
- GANDON F. (2002). *Ontology Engineering : a Survey and a Return on Experience*. Rapport interne 4396, INRIA. 181 p., ISSN 0249-6399.
- GANGEMI A., GUARINO N. & OLTRAMARI A. (2001). Conceptual analysis of lexical taxonomies : The case of WordNet top-level. In *Proc. of FOIS2001*.
- GENESERETH M.-R. (1982). Diagnosis using hierarchical design models. In *AAAI*, p. 278–283.
- GÓMEZ-PÉREZ A. (2000). Développements récents en matière de conception, de maintenance et d'utilisation d'ontologies. *Terminologies Nouvelles*, (19), 9–20. Traduit de l'anglais par S. Descotte.
- GOODY J. (1979). *La raison graphique*. Éditions de Minuit.
- GROS C. & ASSADI H. (1998). Intégration de connaissances dans un système de consultation de documentation technique. In *Actes des premières journées du Chapitre français de l'ISKO*, Lille.
- GROSJEAN M. (2000). Les communications collectives : un mode d'approche des compétences du collectif. l'exemple du collectif hospitalier. *Psychologie du travail et des organisations*, **6**(3-4), 103–31. Numéro spécial compétences collectives au travail.
- GROSJEAN M. (2002). *Connaissances et communication. Conditions d'une efficacité de la production de services*. Rapport interne, DARES. Rapport final.
- GRUBER T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, **5**, 199–220.
- GUARINO N. (1997). Understanding, building, and using ontologies. *International Journal of Human-Computer Studies*, **45**(2/3), 293–310.
- GUARINO N. (1999). The role of identity conditions in ontology design. In V. BENJAMINS, B. CHANDRASEKARAN, A. GOMEZ-PEREZ, N. GUARINO & M. USCHOLD, Coordinateurs, *Proc. of the IJCA'99 Workshop on Ontologies and Problem-Solving Methods*, p. 2/1–2/7, Sweden.
- GUARINO N. & GIARETTA P. (1995). Ontologies and knowledge bases. In *Towards Very Large Knowledge Bases*. Amsterdam : IOS Press.
- GUHA R. V. & LENAT D. B. (1990). Cyc : A mid-term report. *The Artificial Intelligence Magazine*.
- GUIGUE L., RENARD J.-M., BRICON-SOUF N., BEUSCART R. & GEIB J.-M. (2000). Simulation des processus de coopération en USI. *Informatique et santé*, **12**, 133–44.
- GUNDERSEN M. L., HAUG P. J., PRYOR T. A., VAN BREE R., KOEHLER S., BAUER K. & CLEMONS B. (1996). Development and evaluation of a computerized admission diagnoses encoding system. *Computers and Biomedical Research*, **29**(1/2), 351–372.
- HAOUCHE C. (1996). *Utilisation des concepts de l'acquisition des connaissances pour valider un SBC*. Thèse de doctorat, Université de Paris-Dauphine.
- HAOUCHE-GINGINS C. & CHARLET J. (1997). Une méthode pour tester les systèmes à base de connaissances par rapport à leur modèle conceptuel. *Revue d'Intelligence Artificielle*, **11**(4).
- HAOUCHE-GINGINS C. & CHARLET J. (1998). A knowledge-level testing method. *International Journal of Human-Computer Studies*, **49**(1), 1–20.

- HARMELEN F. V. & BALDER J. (1992). (ML)2 : A formal language for kads models of expertise. *Knowledge Acquisition*, **4**, 127–161.
- HARRIS Z., GOTTFRIED M., RYCKMAN T., MATTICK JR P., DALADIER A., HARRIS T. & HARRIS S. (1989). *The Form of Information in Science, Analysis of Immunology Sublanguage*, volume 104 of *Boston Studies in the Philosophy of Science*. Dordrecht & Boston : Kluwer Academic Publishers.
- HATCHUEL A. & WEIL B. (1992). *L'expert et le système. Gestion des savoirs et métamorphoses des acteurs dans l'entreprise industrielle*. Economica. 263 p.
- HATCHUEL A. & WEIL B. (1999). Design-oriented Organizations. towards a unified theory of design activities. In *Proceedings of the 6th International Product development Management Conference*, Churchill College, Cambridge, UK.
- HL7-SD-TC (2001). *Clinical Document Architecture*. HL7. Structured Document Technical Committee. Available at <http://www.hl7.org/Special/committees/>.
- HORROCKS I. (1998). Using an expressive description logic : FaCT or Fiction. In A. COHN, L. SCHUBER & S. SHAPIRO, Coordinateurs, *Proc. of the sixth International Conference on Knowledge representation*, p. 634–47, San Francisco : Morgan Kaufmann.
- ISO/IEC (1997). *Information processing – Hypermedia/Time-based Structuring Language (Hy-Time)*. ISO/IEC. International Standard ISO/IEC JTC 1/SC 18 WG8 N1920rev. Version 1 known as ISO/IEC 10744 :1992.
- JACOB I., KRIVINE J.-P. & MONCLAR F.-R. (2000). De LISA à ELICO : des langages pour le développement de systèmes d'assistance à l'opérateur. In Charlet *et al.* (2000b), chapitre 10, p. 161–75.
- KAMEL BOULOS M., ROUDSARI A. & CARSON E. (2001). Towards a semantic medical web : Healthybermap's dublin core ontology in protégé-2000. In *Proc. of the fifth International Protégé Workshop*.
- KARBACH W., VOB A., SCHUXKEY R. & DROUVEN U. (1991). Model-k : Prototyping at the knowledge level. In *The Expert Systems and their Applications*, p. 501–511, Avignon.
- KARLSSON D. (2001). A design prototype for a decision-support system in the field of urinary tract infection : Application of *openGALEN* techniques for indexing medical information. In V. L. PATEL, R. ROGERS & R. HAUX, Coordinateurs, *Proceedings of the 10th World Congress on Medical Informatics*, p. 479–83, London, UK.
- KASSAR S. E. & CHARLET J. (1997). Représentation de connaissances et aide à la navigation hypertextuelle à partir de cas. In *Actes des 1^{es} Journées Ingénierie des Connaissances*, p. 387–401, Roscoff, France.
- KASSEL G. (2001). *La gestion des concepts et du vocabulaire dans l'entreprise ; Terminologies et Ontologies*. Rapport interne IIIA-01-VT11, IIIA. Rapport de veille technologique IIIA.
- KASSEL G., GRÉBOVAL-BARRY C. & ABEL M.-H. (2000). Programmer au niveau des connaissances en def*. In Charlet *et al.* (2000b), chapitre 9, p. 145–60.
- KAYSER D. (1988). Le raisonnement à profondeur variable. In *Journées du Greco PRC-IA*, p. 103–13, Toulouse, France : Teknea.
- KAYSER D. (1997). *La représentation des connaissances*. Hermès.
- KLINKER G., BHOLA C., DALLEMAGNE G., MARQUÈS D. & DERMOTT J. M. (1991). Usable and reusable programming constructs. *Knowledge Acquisition*, **3**, 117–136.
- KRIVINE J.-P. & DAVID J.-M. (1991). L'acquisition des connaissances vue comme un processus de modélisation : méthodes et outils. *Intellectica*, **2**(12), 101–137.
- KULIKOWSKI C. A. (2002). The micro-macro spectrum of medical informatics challenges : From molecular medicine to transforming health care in a globalizing society. *Methods of Information in Medicine*, **1**, 20–4.
- LAFOREST F., FRÉNOT S. & AL MASRI N. (2002). Dossier médical semi structuré pour des interfaces de saisie multi-modale. In J. CHARLET, Coordinateur, *Dossiers numériques*, volume 6. Paris : Hermès. Numéro spécial de la revue *Document numérique*.
- LAUBLET P., REYNAUD C. & CHARLET J. (2002). Sur quelques aspects du web sémantique. In J. LE MAITRE, Coordinateur, *Actes des 2^{ndes} assises nationales du GDR-I3*, Nancy : Cépaduès, Toulouse.

- LE BOZEC C. (2001). *Gestion des connaissances multi-experts en imagerie médicale. IDEM : images et diagnostics par l'exemple en médecine*. Thèse de doctorat, Univ. Paris 6.
- LE GRAND B. (2001). *Extraction d'information et visualisation de systèmes complexes sémantiquement structurés*. PhD thesis, Univ. Paris 6.
- 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.
- J. LE MAÎTRE, J. CHARLET & C. GARBAY, Coordinateurs (2000). *Le document multimédia en sciences du traitement de l'information*. Toulouse : Cépaduès.
- 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., BOURIGAUULT D. & JAULENT M.-C. (2002). Construction d'une ontologie à partir de corpus : expérimentation et validation dans le domaine de la réanimation chirurgicale. In B. BACHIMONT, Coordinateur, *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 229-38, Rouen, France.
- LENAY O. (2001). *Régulation, planification et organisation du système hospitalier : la place des outils de gestion dans la conception des politiques publiques*. Thèse de doctorat, École des Mines de Paris.
- LENAY O. & MOISDON J.-C. (2000). Croître à l'abri de la gestion ? le cas du système hospitalier public français. *Cahiers de recherche du centre de gestion scientifique, École des Mines de Paris*, (17). ISSN 1268-4317.
- LEROI-GOURHAN A. (1964). *Le geste et la parole*. Albin Michel.
- LEWKOWICZ M. & ZACKLAD M. (2000). *Une approche de la capitalisation des connaissances : l'analyse des processus de prise de décisions collectives*, chapitre 28, p. 451-64. In Charlet et al. (2000b).
- LINDBERG D. A. B. & HUMPHREYS B. L. (1990). The UMLS knowledge sources : tools for building better user interfaces. In *Proceedings of the 14th Annual SCAMC*, p. 121-125, Washington : IEEE.
- LINSTER M. (1992). *Knowledge Acquisition Based on Explicit Methods of Problem Solving*. Ph. d. dissertation, #d 386, Univ. of Kaiserslautern (G).
- 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.
- LÉPINE P. & AUSSENAC-GILLES N. (1996). *Modélisation de la résolution de problèmes : comparaison de KADS et MACAO sur une application juridique*, In N. AUSSENAC-GILLES, P. LAUBLET & C. REYNAUD, Coordinateurs, *Acquisition et ingénierie des connaissances : tendances actuelles*, chapitre 7, p. 131-48. Cepaduès-éditions.
- MANARANCHE R. (1995). Mammifères. In *Encyclopædia Universalis, vol.*, p. 405-13.
- MARCUS S. & McDERMOTT J. (1989). SALT : a knowledge acquisition language for propose-and-revise systems. *Artificial Intelligence*, **39**(1), 1-37.
- MARET P. & GHÉDIRA C. (2002). Dossier médical semi structuré pour des interfaces de saisie multi-modale. In J. CHARLET, Coordinateur, *Dossiers numériques*, volume 6. Paris : Hermès. Numéro spécial de la revue *Document numérique*.
- MÉLÈSE J. (1990). *Au sujet de l'information*. Paris : Les éditions de d'Organisation.
- MUSEN M. A. (1989). Conceptual models of interactive knowledge acquisition tools. *Knowledge Acquisition*, **1**, 73-88.
- MUSEN M. A. (98). Domain ontology in software engineering : Use of protégé with the eom architecture. *Methods of Information in Medicine*, **37**, 540-50.
- NAIDITCH M. (2000). Modélisation des trajectoires : problèmes méthodologiques. *Innovation et technologie en biologie et médecine*, **21**(5), 307-12.
- NELSON S. J., JOHNSTON D. & HUMPHREYS B. L. (2001). Relationships in medical subject headings. In C. A. BEAN & R. GREEN, Coordinateurs, *Relationships in the organization of knowledge*, New York : Kluwer Academic Publishers.

- NEWELL A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87–127.
- NOBÉCOURT J. (2000). Représenter la notion de propriété dans les graphes conceptuels et les logiques de description. In Charlet *et al.* (2000b), chapitre 11, p. 177–94.
- 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.
- PAQUEL N. (2002). *Architecture du système d'information de santé : normalisation de l'interopérabilité*. Rapport interne, EDISANTÉ. Document provisoire.
- PERRIER A. & SIMIŁOWSKI T. (2002). Une série sur la médecine factuelle (« evidence-based medicine ») : mais pour quoi faire ? *Revue des maladies respiratoires*, **19**, 395–398.
- PILLOT F. D. & CUVIER M. F. (1832). *Œuvres complètes de Buffon*. Paris : F. D. Pillot. Vingt-sept volumes augmentés de deux volumes supplémentaires par M. F. Cuvier.
- POITOU J.-P. (1996). La gestion des connaissances comme condition et résultat de l'activité industrielle. *Intellectica*, **1**, 185–202.
- PUERTA A. R., EGAR J. W., TU S. W. & MUSEN M. A. (1992). method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. *Knowledge Acquisition*, **4**(2), 171–96.
- RECTOR A. L. (1998). Thesauri and formal classifications : Terminologies for people and machines. *Methods of Information in Medicine*, **37**(4–5), 501–509.
- REINDERS M., VINKHUYZEN E., VOSS A., AKKERMANS H., BALDER J., BARTSCH-SPÖRL B., BREDEWEG B., DROUVEN U., VAN HARMELEN F., KARBACH W., KARSSSEN Z., SCHREIBER G. & WIELINGA B. (1991). A conceptual modelling framework for knowledge-level reflection. *AI Communications*, **4**(2/3).
- REITER A. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, **32**, 57–95.
- REYNAUD C. (1989). ADELE, un outil d'aide à l'acquisition des connaissances basé sur des justifications. Thèse de doctorat, Université Paris Sud.
- REYNAUD C. (1993). Acquisition and validation of expert knowledge by using causal models. In David *et al.* (1993), chapitre 23, p. 517–540.
- REYNAUD C. (1999). *L'exploitation de modèles de connaissances du domaine dans le processus de développement d'un système à base de connaissances*. Habilitation à diriger des recherches, Université Paris Sud.
- RODRIGUES J.-M., TROMBERT-PAVIOT B., BAUD R., WAGNER J. & MEUSNIER F. (1998). Galen-In-Use : Using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures. In B. CESNIK, C. SAFRAN & P. DEGOULET, Coordinateurs, *Proceedings of the 9th World Congress on Medical Informatics*, Seoul.
- 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.
- ROGERS J., ROBERTS A., SALOMON D., VAN DER HARING E., WROE C., ZANSTRA P. & RECTOR A. L. (2001). GALEN ten years on : Tasks and supporting tools. In V. L. PATEL, R. ROGERS & R. HAUX, Coordinateurs, *Proceedings of the 10th World Congress on Medical Informatics*, p. 256–60, London, UK.
- ROUSSEY C., CALABRETTO S. & PINON J.-M. (2002). Le thésaurus sémantique : contribution à l'ingénierie des connaissances documentaires. In B. BACHMONT, Coordinateur, *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 209–20, Rouen, France.
- SAGER N., HHÀN N., LYMAN M. & TICK L. (1996). Medical language processing with SGML display. *Journal of the American Medical Informatics Association*, **3**(suppl), 547–551.
- SAGER N., LYMAN M., NHÀN N. T. & TICK L. J. (1995). Medical language processing : Applications to patient data representation and automatic encoding. *Methods of Information in Medicine*, **34**(1/2), 140–146.

- SCHREIBER G., AKKERMANS H., ANJEWIERDEN A., DE HOOG R., SHALDBOLT N., DE VELDE W. V. & WIELINGA B. (2000). *Knowledge Engineering and Management : The CommonKADS Methodology*. MIT Press.
- SCHREIBER G., WIELINGA B., AKKERMANS H., DE VELDE W. V. & ANJEWIERDEN A. (1994). Cml : The commonkads conceptual modelling language. In L. STEELS, G. SCHREIBER & W. V. DE VELDE, Coordinateurs, *EKAW'94*, volume LNAI N 867, p. 1–25 : Springer-Verlag.
- 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).
- SIMONDON G. (1958). *Du mode d'existence des objets techniques*. Paris : Aubier.
- SLODZIAN M. (1999). WordNet et EuroWordNet : questions impertinentes sur leur pertinence linguistique. *Sémiotiques*, (17), 51–70. Numéro spécial *Dépasser les sens iniques dans l'accès automatisé aux textes*, coordonné par B. Habert.
- SLODZIAN M. (2000). Wordnet : what about its linguistic relevancy? In R. DIENG, Coordinateur, *Proc. of the EKAW conference*, Juan-les-Pins, France.
- SOUALMIA L., GRÉBOVAL-BARRY C., ABDUKLRAB H. & DARMONI S. (2002). Modélisation et représentation des connaissances dans un catalogue de santé. In B. BACHIMONT, Coordinateur, *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 139–50, Rouen, France.
- SOWA J. F. (1984). *Conceptual Structures : Information Processing in Mind and Machine*. London : Addison-Wesley.
- SPACKMAN K. A., DIONNE R., MAYS E. & WEIS J. (2002). Role grouping as an extension to the description logic of ontylog motivated by concept modeling in snomed. In *Actes du AMIA Annual Symposium 2002*, p. 712–6, San Antonio, Texas.
- SÉROUSSI B., BAUD R., MOENS M., MIKHEEV A., SPYNS P., CEUSTERS W. & ZWEIGENBAUM P. (1996). *Dome final report*. Deliverable report MLAP-Dome 8, DIAM-SIM/AP-HP.
- SÉROUSSI B., BOUAUD J. & ANTOINE E.-C. (1999). Enhancing clinical practice guideline compliance by involving physicians in the decision process. In W. HORN, Y. SHAHAR, G. LINDBERG, S. ANDREASSEN & J. WYATT, Coordinateurs, *Artificial Intelligence in Medicine, Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making, AIMDM'99, Aalborg, DK, June 1999, Proceedings*, number 1620 in Lecture Notes in Artificial Intelligence, p. 76–85. Springer-Verlag.
- SÉROUSSI B., BOUAUD J. & ANTOINE E.-C. (2000). OncoDoc : un système documentaire au service de l'expertise clinique des médecins. In *Actes du 12^e congrès Reconnaissance des Formes et Intelligence Artificielle*, volume 2, p. 111–129, Paris : AFCET.
- STEELS L. (1985). Second-generation expert systems. *Journal of Future Generation Computer Science*, **1**(4).
- STEELS L. (1990). Components of expertise. *The Artificial Intelligence Magazine*.
- STEELS L. (1993). The componential framework and its role in reusability. In David *et al.* (1993), p. 273–278.
- STEFANELLI M. (2002). Knowledge management to support performance-based medicine. *Methods of Information in Medicine*, **1**, 36–43.
- TALBI N. & LAUBLET P. (2000). Construction d'un modèle générique d'expertise à partir d'un exemple : une solution pragmatique. In Charlet *et al.* (2000b), chapitre 3.
- TANGE H. J. (1995). The paper-based patient record : Is it really so bad? *Computer Methods and Programs in Biomedicine*, **48**, 127–131.
- TCHOUNIKINE P. (2002). Pour une ingénierie des environnements informatiques pour l'apprentissage humain. *Information - Interaction - Intelligence*, **2**(1), 59–95.
- TCHOUNIKINE P., ISTENES Z. & TRICHET F. (2000). ZOLA : un langage permettant une approche flexible de l'opérationnalisation du modèle conceptuel d'un système à base de connaissances. In Charlet *et al.* (2000b), chapitre 33, p. 529–42.

- TU S. W., ERIKSSON H., GENNARI J. H., SHAHAR Y. & MUSEN M. A. (1995). Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools : application of PROTÉGÉ-II to protocol-based decision support. *Artificial Intelligence in Medicine*, **7**, 257-289.
- 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.
- WACK M., COTTIN N., MIGNOT B. & EL MOUDNI A. (2002). Dossiers numériques. In J. CHARLET, Coordinateur, *Dossiers numériques*, volume 6. Paris : Hermès. Numéro spécial de la revue *Document numérique*.
- WEIS J.-C. (2002). *Du réseau ville-hôpital au réseau distributionnel*. Rapport interne, STIM/DPA/DSI/AP-HP. DEA informatique médicale.
- WELTY C. & GUARINO N. (2001). Supporting ontological analysis of taxonomic relationships. *Data and Knowledge Engineering*, **39**(1), 51-74.
- WETTER T. (1990). First order logic foundations of the kads conceptual model. In B. WIELINGA, Coordinateur, *Current trends in Knowledge Acquisition*, p. 356-375. Amsterdam : IOS Press.
- WIELINGA B. J., SCHREIBER A. T. & BREUKER J. A. (1992). KADS : a modelling approach to knowledge engineering. *Knowledge Acquisition*, **4**(1), 5-54.
- WOODS W. A. (1991). Understanding subsumption and taxonomy : A framework for progress. In J. F. SOWA, Coordinateur, *Principles of Semantic Networks*, chapitre 1, p. 45-94. San Mateo, Ca. : Morgan Kaufmann Publishers.
- ZOUINAR M. & SALEMBIER P. (2000). Modélisation du contexte partagé pour l'analyse et la des environnements de travail coopératif. In Charlet *et al.* (2000b), chapitre 33, p. 529-42.
- ZWEIGENBAUM P., BACHIMONT B., BOUAUD J., CHARLET J. & BOISVIEUX J.-F. (1995a). Issues in the structuring and acquisition of an ontology for medical language understanding. *Methods of Information in Medicine*, **34**(1/2).
- ZWEIGENBAUM P., BACHIMONT B., BOUAUD J., CHARLET J. & BOISVIEUX J.-F. (1995b). A multi-lingual architecture for building a normalised conceptual representation from medical language. *Journal of the American Medical Informatics Association*, **2**(suppl), 357-361.
- ZWEIGENBAUM P., BOUAUD J., BACHIMONT B., CHARLET J., SÉROUSSI B. & BOISVIEUX J.-F. (1998). From text to knowledge : a unifying document-oriented view of analyzed medical language. *Methods of Information in Medicine*, **37**(4-5), 384-393.
- 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 ».

Table des figures

1.1	L'acquisition des connaissances conduite par les modèles (tiré de (Aussenac-Gilles <i>et al.</i> , 1992)).	5
1.2	Le modèle en couche de KADS avec des exemples médicaux. Tiré de (Schreiber <i>et al.</i> , 2000).	7
1.3	Principe général de la méthode de validation d'AMD.	12
3.1	Un compte rendu de radiologie en XML	32
3.2	Architecture du système HOSPITEXTE.	34
3.3	Visualisation d'un compte rendu d'hospitalisation.	35
3.4	L'architecture générale du système ONCODOC (figure reprise de (Séroussi <i>et al.</i> , 2000)).	38
3.5	(a) Affectation thérapeutique : un découpage de l'approche totalement implémentée.	39
3.5	(b) Affectation thérapeutique : un découpage de l'approche ONCODOC (figure reprise de (Séroussi <i>et al.</i> , 2000)).	39
3.6	Partie des métadonnées de la page < http://www.chu-rouen.fr/ssf/santspe.html >, index thématique des sites en santé du catalogue CISMef.	40
4.1	Les étapes de la méthodologie.	54
4.2	Analyse d'un compte rendu d'hospitalisation.	56
4.3	Connaissances de sens commun sur l'admission.	57
4.4	Trois représentations du genre sexuel en utilisant la représentation graphique des graphes conceptuels où les rectangles sont des concepts et les ovoïdes, des relations.	58
4.5	Le corps de la personne.	58
4.6	Le descripteur « nez » dans le Mesh.	61
4.7	Quelques exemples de couverture du thésaurus de spécialité	67
5.1	Un extrait de la DTD MEDDOS de transport de tout ou partie d'un dossier médical. Le message MEDDOS a été conçu durant le projet européen PROMPT. Originellement implémenté en EDIFACT, il a été ensuite porté en XML en 1998 par ICSF < http://www.icsf.fr >.	78
5.2	Schéma de la <i>Clinical Document Architecture</i>	79
5.3	Structure d'un message ebXML (version 0.98b du 13 mars 2001).	80
5.4	Principes du contenu de l'enveloppe d'échange	82
6.1	Les transformations et le cycle de vie des connaissances médicales.	100

A

acquisition des connaissances, **1**, 15
AMD, 11
analyse distributionnelle, 65
annotation, 36, **40**, 101
apprentissage
 collectif, 99
 organisationnel, **89**, 96
approche médiateur, 85
artefact informatique, 21

B

balisage, 30
balise, 30

C

CDA, *voir Clinical Document Architecture*
CEN TC251, 78
CISMeF, 38
classification, **46**, 59
Clinical Document Architecture, 78
CommonKADS, 7
conception des savoirs, 91
connaissance, 17
connaissances
 du domaine, 3
 pour l'action, 95
contexte, **19**, 21
 partagé, 24
corpus, 16, **50**
 démarche de, 50
CSCW, *voir travail coopératif*

D

définition de type de document, **31**, 77
DICOM, 80

document

 de navigation, 33
 de lecture, 33
 de synthèse, 33
 original, 33
DOME, 31
donnée, 18
dossier médical, **29**, 99, 101
dossier patient, *voir dossier médical*
DTD, *voir définition de type de document*
Dublin core, 40

E

EBM, *voir médecine factuelle*
ebXML, 80
échange de données informatisées, 77
EDI, *voir échange de données informatisées*
EDIFACT, 77
EDISANTÉ, 80
engagement ontologique, 45, **52**, 57
engagement sémantique, **52**, 57
enveloppe d'échange médical, 81
épistémologie, 17
 constructiviste, 19, **25**
 positiviste, 25
Evidence-Based Medicine, *voir médecine factuelle*

F

forme architecturale, 77

G

GALEN, 63
GBP, *voir guide de bonne pratique*

gestion, *voir* sciences de gestion
 gestion des connaissances, **89**, 91
 médicales, 96, 97, **98**
 gouvernementalité, 94
 graphe conceptuel, 55
 guide de bonne pratique, 37

H

Health Level 7, 77
 HL7, *voir Health Level 7*
 HOSPITEXTE, 22, **31**
 HTSC, 35
 hyperdocument, 33
 hyperonymie, 48
 hypothèse
 déterministe, 25
 ontologique, 25
 phénoménologique, 26
 téléologique, 26
 hypothèse d'interaction limitée, 24
 HyTIME, 77

I

index, 68
 indexation, 68
 information, 17
 nominative, 98
 non nominative, 98
 information médicale, 29
 ingénierie des connaissances, 15, **17**, 26,
 109
 intelligence artificielle, 109
 interopérabilité, **76**, 103
 sémantique, 76, 84, 85
 syntaxique, 76, 84
 interprétation, **18**, 19, 51
 pour l'action, 92

K

(KA)², 70

M

médecine factuelle, 99
 médicalisation, 94
 méronymie, 48
 métadonnée, 38
 objective, 41
 subjective, 40
 méthode
 ascendante, 8

 descendante, 8
 méthode de raisonnement, 16
 MENELAS, 55
 Mesh, 40, **60**
 métathésaurus, 61
 modélisation, 2
 langage de, 10
 modèle, 2
 causal, 3
 du domaine, *voir* connaissances du
 domaine
 du raisonnement, 4
 modèle conceptuel, **2**, 16

N

niveau des connaissances, **2**, 21
 normalisation sémantique, **51**, 69

O

ONCODOC, 37
 Ontologie, 45
 ontologie, 3, **43**, 65, 101
 computationnelle, 53
 générique, 68
 régionale, 52
 réutilisation des, 67
 opérationnalisation, **6**, 53
 langage d', 10
 organisation, 83, **95**, 104
 outil de gestion, 93
 outil terminologique, 50

P

P2VIE, 83
 pilotage organisationnel, 93
 planification, 94
 pluridisciplinarité, 109
 PMSI, *voir* programme de médicalisation du sys-
 tème d'information
 primitive, 19, **51**
 programme de médicalisation du système d'in-
 formation, **66**, 93
 propriété, 48

R

référence médicale opposable, 98
 régulation, 94
 réseau de soins, *voir* réseau de santé
 réseau de santé, 83, 94, **105**
 rôle, 7, 53, **57**, 59

raison
 computationnelle, 18
 graphique, 18
Reference Information Model, 78
ressource ontologique, *voir* ressource terminologique
ressource terminologique, 61, 67, **69**
RIM, *voir Reference Information Model*
RMO, *voir* référence médicale opposable

S

sémantique différentielle, 51
sémantique hypertextuelle, **36**, 101
 annotationnelle, 36
SBC, *voir* système à base de connaissances
sciences
 cognitives, 20
 de gestion, 104, 110
 de l'artificiel, 26
 et technologies de l'information et de la communication, 28
 naturelles, 45
science de l'Être, 45
serveur de terminologie, **63**, 102
service Web, 85
Simple Object Access Protocol, 80
SIS, *voir* système d'information de santé
situation d'apprentissage, 91
SNOMED, 61
SOAP, *voir Simple Object Access Protocol*
standard, 85
STIC, *voir* sciences et technologies de l'information et de la communication
sujet connaissant, 95
support, 17
 électronique, 31
 papier, 31
SYNTEX, 65
système à base de connaissances, **15**, 19
système d'information de santé, 84, **103**, 104
système d'usage, 20
système sémiotique, 20

T

tâche générique, 8
TALM, *voir* traitement automatique du langage médical
taxinomie, 46

taxinomie, *voir* taxinomie
terminologie, 59
thésaurus, 59
thésaurus sémantique, **70**, 72, 102
Topic Maps, 71, **73**
traçabilité de l'échange, 81
traitement automatique du langage médical, **36**, 101
trajectoire de patient, *voir* trajectoire de soins
trajectoire de soins, 82, **94**, 106
travail coopératif, 104

U

UMLS, 61
UPERY, 65
urbanisation, 104

V

validation, 6, 11

W

Web sémantique, **40**, 72, 84, 103
WORDNET, 59

X

XML, 30, 77

RÉSUMÉ

Le travail présenté dans ce mémoire d'habilitation à diriger des recherches cherche à faire le point sur le domaine de l'Ingénierie des connaissances et à étudier ses apports, présents et futurs, à la gestion des connaissances médicales. L'Ingénierie des connaissances ne peut pas résoudre à elle seule les problèmes de la gestion des connaissances médicales. Un certain nombre d'années passées à travailler dans ces domaines fondamentaux – pour l'Ingénierie des connaissances – et appliqués – pour la gestion des connaissances médicales – nous ont convaincu qu'il était nécessaire de mobiliser d'autres domaines dans une approche pluridisciplinaire. C'est que nous avons fait en entreprenant des recherches qui ressortissent à l'Ingénierie des connaissances mais aussi au domaine des systèmes d'information et à la gestion. C'est cette démarche que nous développons dans ce mémoire.

Pour ce faire, nous proposons d'abord une introduction à l'acquisition des connaissances et à la façon dont le domaine rencontre les applications médicales. Nous y abordons les concepts et les modèles de ce domaine et y rappelons rapidement les résultats que les travaux ayant concouru à ces recherches nous ont permis d'obtenir. Nous poursuivons par une analyse épistémologique de l'Ingénierie des connaissances pour en argumenter les apports et en définir un cadre de travail.

Nous abordons ensuite plus précisément des applications médicales et constatons que le caractère contextuel de l'information et ses rapports avec son support (ici informatique) nous amène à mettre l'accent sur l'intérêt du document textuel comme support et véhicule de la connaissance. Le projet de dossier médical informatisé HOSPITEXTE développé au sein du service est présenté dans ce paradigme et la problématique de l'indexation des documents médicaux est spécifiquement étudiée.

La notion d'ontologie est abordée de façon approfondie en justifiant et prenant une position affirmée sur ce que nous pensons que c'est et que ça n'est pas. Cette position nous semble nécessaire à la compréhension de ce que l'on peut faire ou non avec les ontologies comme avec de nombreux « produits » terminologiques tels que les thésaurus ou les bases de données lexicales. Nous défendons, à l'occasion de la description de travaux développés en collaboration, l'utilisation d'outils du traitement du langage naturel comme mode d'accès privilégié aux concepts d'une ontologie. Des ontologies en médecine, construites durant ces travaux, servent d'exemples à nos propos.

Les systèmes d'information en médecine, sont abordés au regard du support de la mise en œuvre de la gestion des connaissances médicales et vis-à-vis des questions d'interopérabilité. Ces systèmes dont le champ d'action grandit, de l'établissement de soins à la médecine de ville, impliquent l'ensemble des professionnels de santé pour une pathologie ou une situation donnée. Cette situation crée de nouvelles contraintes pour la mise en œuvre de ces larges systèmes d'informations des réseaux, en particulier au niveau de l'interopérabilité qui doit obligatoirement passer par des standards qui se construisent et s'imposent lentement.

La dernière partie du mémoire s'appuie sur les précédentes pour, d'abord rechercher un certain nombre de points communs entre la gestion et l'Ingénierie des connaissances et, ensuite, proposer un certain nombre d'indications sur l'élaboration d'un système de gestion des connaissances en santé. À partir de là, nous repositionnons les enjeux de la gestion des connaissances médicales ainsi que leur spécificité au regard des problématiques développées dans les chapitres précédents. Les projets comme les développements normatifs sont resitués dans un contexte d'intégration. Enfin, nous essayons de tirer les conséquences de ces travaux dans le contexte des développements futurs attendus, en particulier au regard des nouvelles structures de soins qui se font jour.

Mots-clés : Ingénierie des connaissances, épistémologie, théorie du support, ontologies, informatique médicale, gestion, thésaurus, dossier médical, interopérabilité, gestion des connaissances, Web sémantique.