



HAL
open science

Pourquoi n'y a-t-il pas d'expérience en ingénierie des connaissances ?

Bruno Bachimont

► **To cite this version:**

Bruno Bachimont. Pourquoi n'y a-t-il pas d'expérience en ingénierie des connaissances ?. 15èmes Journées francophones d'ingénierie des Connaissances, May 2004, Lyon, France. pp.53-64. <hal-00348389>

HAL Id: hal-00348389

<https://hal.science/hal-00348389v1>

Submitted on 26 Jan 2009

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.



HAL Authorization

Pourquoi n’y a-t-il pas d’expérience en ingénierie des connaissances ?

Bruno Bachimont^{1,2}

¹ Institut National de l’Audiovisuel, Recherche et Expérimentation
4, Avenue de l’Europe, 94366 Bry sur Marne, bbachimont@ina.fr

² Heudiasyc UMR CNRS 6599, Université de Technologie de Compiègne
BP 20259 60205 Compiègne cedex, Bruno.Bachimont@utc.fr

Résumé : L’ingénierie des connaissances modélise les connaissances d’un domaine pour les opérationnaliser dans un système destiné à assister une tâche ou le travail intellectuel dans ce domaine (résolution de problème, aide à la décision, consultation documentaire, etc.). Le problème posé dans cet article est le statut qu’il faut conférer aux modèles élaborés : sont-ils à considérer comme les modèles scientifiques qui selon une épistémologie standard doivent décrire et prédire les phénomènes ? Quelles expériences viennent les confirmer et les infirmer ? Selon nous, il n’y a pas d’expériences en ingénierie des connaissances. Cette dernière est une ingénierie des inscriptions numériques des connaissances dont la modélisation formelle et syntaxique doit soutenir une interprétation critique, imprédictible et contextuelle. L’interprétation s’effectue selon des normes et conventions permettant d’aborder l’usage concret dans sa singularité et sa différence. Le calcul formel organise les inscriptions et assure la médiation entre les différentes interprétations. L’ingénierie des connaissances est donc un art interprétatif mobilisé sur une technique formelle.

Mots-clés : Épistémologie, ingénierie documentaire, représentation des connaissances, sciences nomologiques, sciences idiographiques

1 Introduction

L’ingénierie des connaissances est encore une jeune discipline qui, bien que principalement issue de l’intelligence artificielle, entretient des relations de collaboration voire de filiation avec bien d’autres disciplines : linguistique, psychologie, sociologie, logique formelle, etc. Ce pluralisme interdisciplinaire implique un questionnement épistémologique récurrent et nécessaire. Ce dernier renvoie non seulement à des problèmes philosophiques mais aussi, et surtout, à des enjeux méthodologiques. En effet, selon la fondation que l’on voudra donner à l’ingénierie des connaissances et donc le statut qui en découlera, l’objectif de sa pratique et les modalités de ses travaux seront largement différents.

Ainsi, si l'ingénierie des connaissances (IC) se revendique, contrairement à ce que laisse entendre son libellé, comme une science visant la modélisation des connaissances, on lui demandera d'explicitier les outils théoriques d'où elle tire ses modèles, et de préciser selon quels protocoles empiriques les prédictions ou hypothèses formulées à l'aide des modèles sont évaluées. Dans ce cas, l'ingénierie des connaissances se distingue bien mal des sciences cognitives et de l'intelligence artificielle, dont elle revendique pourtant son émancipation même si elle se félicite de leur cousinage. Mais si l'IC ne veut pas se confronter à l'évaluation empirique, elle peut trouver sa scientificité dans le mathématisme et le formalisme de ses modèles. On lui demandera alors de montrer ses théorèmes et de justifier son indépendance vis-à-vis de la logique et des mathématiques. En outre, dans ces deux cas, quel statut donner à l'ingénierie qu'elle met en œuvre : une simple application de ses modèles, mais alors toute application ou réalisation doit être motivée par un modèle théorique montrant ses hypothèses et démontrant ses conclusions ; ou bien une technique possédant son propre savoir, mais alors on s'éloigne de la justification scientifique au profit de l'efficacité empirique.

En revanche, si l'ingénierie des connaissances est une ingénierie comprise comme un répertoire de procédés divers permettant d'instrumenter des systèmes informatiques, on lui demandera de prouver son efficacité et de préciser en quoi elle se distingue de disciplines voisines, notamment l'informatique de gestion et le génie logiciel (cf. (Bézivin, 2001)). Selon ces deux positions extrêmes, l'IC doit choisir entre sa récupération dans des disciplines scientifiques rattachées aux sciences cognitives (sciences expérimentales) ou aux mathématiques (sciences formelles), ou sa dissolution dans les techniques informatiques (recherche d'efficacité empirique).

Ces deux positions, pour caricaturales qu'elles soient, n'expriment pas moins l'alternative dans laquelle la plupart des chercheurs de l'IC se sentent plongés. Si bien que régulièrement le doute s'empare des praticiens de cette discipline qui reconduisent alors la discussion épistémologique à propos des fondements. On citera pour mémoire, parmi les travaux récents, (Charlet, 2001; Charlet, 2002), où les questions suivantes sont abordées : l'IC est-elle une science, une technologie, un art pratique ? Peut-on l'enseigner ? Comment articuler et distinguer l'IC des disciplines voisines : les sciences de l'organisation, les sciences de la gestion ?

Pour analyser cette question et élaborer notre contribution, nous nous proposons de repartir d'un constat qui ne laisse pas de nous surprendre et de nous plonger dans la perplexité : il n'existe pas d'expérimentation en IC. Par expérimentation, nous entendons le processus largement idéalisé des sciences dites nomologiques (i.e. reposant sur des lois) selon lequel une théorie est élaborée et sa validité est établie par la confrontation de ses prédictions avec l'expérience. L'expérience a donc un rôle de confirmation ou de falsification d'une loi ou une d'une théorie qui ont pour but de représenter le fonctionnement naturel des choses : la variété des phénomènes est rapportée à l'unité et l'idéalité d'une loi.

L'IC n'attend pas de ses modèles de prédire quoi que ce soit sur le monde ni sur la connaissance. Elle vise plutôt à instrumenter le travail intellectuel, l'exercice de la pensée, le travail de la connaissance. Son ambition est empirique et vise à l'efficacité. Cependant, l'IC élabore ses propres modèles et ne se réduit pas à mettre en œuvre des modèles issus d'autres disciplines théoriques ou modélisatrices. Si l'IC élabore des

modèles, mais si ces derniers n'ont pas à se confronter à la prédiction et la falsification expérimentale, quel est leur statut et leur rôle ? Comment caractériser l'ingénierie des connaissances et lui donner un objet et une méthode qui permettent non seulement de comprendre le positionnement revendiqué par ses praticiens mais aussi de préciser sa méthodologie et de proposer des axes de recherches ? C'est l'ambition, sans doute démesurée, de cet article.

Pour cela, nous revenons sur le constat qu'il n'y a pas d'expérience en IC. L'expression volontairement provocatrice renvoie à ce qui nous semble être une propriété profonde de l'IC et c'est la raison pour laquelle nous motivons cette thèse et l'explicitons. Bien évidemment, il ne faudra pas attendre de nous une démonstration ni un examen empirique exhaustif de la discipline : notre propos n'est pas de décrire sociologiquement l'ensemble de notre discipline, mais de nous appuyer sur notre propre pratique en tant qu'ingénieur et celle de nos collègues pour comprendre l'ingénierie des connaissances dans ce qui nous semble être son essence.

Après ce constat, nous élaborons deux niveaux d'analyse. Selon le premier, nous argumentons que l'IC ne peut avoir de validation par elle-même : soit elle revendique une validation formelle et devient une logique appliquée, soit elle revendique une validation empirique et doit alors emprunter les modalités d'expérimentation à la discipline qui lui a fourni les connaissances qu'elle a modélisées. L'IC est alors prise entre une logique qu'elle applique et une pratique ou discipline empirique qu'elle modélise.

Si cette analyse permet de comprendre le malaise permanent dans lequel semble plonger les praticiens de l'IC dès lors qu'ils s'interrogent sur la nature de ce qu'ils font, elle ne permet pas de construire une solution. C'est pourquoi nous abordons un deuxième plan d'analyse, où nous abordons l'IC non comme une ingénierie modélisant des connaissances, mais comme une ingénierie des inscriptions numériques de connaissances, qui vise à instrumenter le travail cognitif associé à ces inscriptions.

L'idée est que la connaissance ne peut s'appréhender techniquement, c'est-à-dire via une ingénierie, qu'à travers les inscriptions qui l'expriment. Les inscriptions possèdent alors deux dimensions : une dimension matérielle manipulable techniquement (le « signifiant » en quelque sorte), une dimension interprétative abordée sémiotiquement (le « signifié »). La connaissance s'aborde à travers des symboles physiques, sachant que le sens des symboles physiques, qui leur est associé en tant qu'ils sont des symboles, n'est pas directement corrélé aux propriétés matérielles ou syntaxiques qui leur sont associées en tant qu'ils sont matériels. Bref, si les connaissances ne s'abordent qu'à travers leur inscription, l'inscription n'est pas une connaissance mais seulement son interprétation. Et c'est là qu'un travail est nécessaire pour articuler la matérialité de l'inscription à la connaissance qu'elle exprime, et c'est donc là que l'IC intervient.

Le rapport de l'IC à la connaissance n'est qu'indirect : l'ingénierie comme technique ne manipule que des objets matériels qui sont des inscriptions. Si elle peut élaborer des modèles où elle prédit le comportement formel des inscriptions, elle ne peut prédire ce qu'il en sera de leur interprétation comme connaissances. Elle décrit la singularité d'une situation d'usage en décrivant sa différence aux normes et conventions interprétatives associées aux inscriptions. Ainsi, l'IC est à comprendre comme étant à la fois une modélisation (au sens nomologique classique) et une interprétation (au sens sémiotique et critique). Comme modélisation, l'IC prédit et valide le comportement for-

mel des systèmes qu'elle élabore : il s'agit de modèle de système informatique, et non de modèle de connaissances. Comme interprétation, l'IC se confronte à l'interprétation des inscriptions manipulées selon les normes associées à ses modèles et les ajuste aux usages et aux pratiques. Les normes interprétatives s'accumulent et se sédimentent pour devenir une mémoire de l'usage et fournir un cadre d'intelligibilité pour les usages futurs. L'IC interprétative rend compte de la singularité (*l'idiotisme*) de chaque situation comme écart ou différence à des normes ou conventions accumulées à travers le système. L'IC se fait alors idiographique. L'IC est donc une technologie, dont la technique (*techno-*) procède de deux savoirs scientifiques (*-logie*), un savoir nomologique de modélisation formelle, un savoir idiographique critique de l'interprétation sémiotique.

2 La démarche de l'ingénierie des connaissances

Dans cette section, nous revenons sur la pratique standard de l'ingénierie des connaissances pour expliciter ce que nous entendons par l'absence d'expérience en IC, avec les réserves faites plus haut. L'ingénierie des connaissances se présente habituellement comme une démarche de modélisation. La modélisation connaît essentiellement deux tendances, qui se traduisent par deux types essentiels de modèles, *les modèles de raisonnement* et *les modèles du domaine*. Selon la première approche, on abstrait et généralise ce que savent faire les programmes logico-formels pour proposer des méthodes de résolution de problèmes ; ces méthodes sont des moules génériques, des schémas conceptuels, dans lesquels on peut exprimer les données tirées du domaine : on obtient un modèle instancié qu'il suffit d'opérationnaliser en un programme informatique pour obtenir le système à base de connaissances (David *et al.*, 1993). Ces méthodes permettent de guider l'acquisition des connaissances puisque, dès lors qu'on les a choisies, il suffit de mettre en correspondance les termes du domaine avec les éléments de la méthode. Selon la seconde approche, on se fonde sur l'application qu'il faut réaliser et la tâche du domaine plutôt que sur les méthodes de raisonnement que les programmes permettent de mener. Il s'agit alors de modéliser les objets pertinents du domaine, les concepts permettant de les considérer et les connaissances permettant de les manipuler. Cette démarche est dite « ascendante » car il s'agit davantage de rendre compte du problème (avec le risque d'aboutir à un modèle complexe peu opérationnalisable) que d'y appliquer un modèle générique (en ayant un modèle opérationnalisable, mais ne répondant pas à la question posée). « coller au problème », ou y « plaquer un modèle » est l'alternative habituelle de la conception d'un système à base de connaissances.

Mais, s'il faut modéliser, il faut également déterminer comment le modèle s'articule à la réalité informatique qui l'opérationnalise. On connaît la proposition de (Newell, 1982) de distinguer un niveau des connaissances où l'on modélise les connaissances portant sur un problème à traiter ou un domaine à caractériser, et un niveau des programmes où l'on opérationnalise le modèle en un programme effectif. L'un des nombreux intérêts de cette proposition est de pouvoir rendre compte du fait que différents programmes peuvent opérationnaliser un même modèle, et par ailleurs qu'un programme introduit des détails et informations endogènes à l'opérationnalisation mais extrinsèques à la modélisation (ces détails sont arbitraires par rapport à la réalité modélisée).

On retiendra qu'un modèle d'IC exprime les connaissances d'un domaine dans un langage de modélisation et l'opérationnalise en un système. L'objectif est d'ordre pratique et vise l'efficacité empirique : le système d'IC sera validé par son utilité et sa capacité à être correctement utilisé en situation pour résoudre les problèmes pour lesquels il a été conçu. Le modèle ne sera donc pas évalué par rapport à une expérience en situation de laboratoire, mais par rapport aux situations effectives de son usage. Le modèle n'exprime pas un savoir ou une théorie sur le monde, mais constitue un outil pour agir dans le monde.

3 Modèles scientifiques ou applications techniques

Cela étant dit, si un modèle d'IC n'est pas une théorie mais un outil, comment aller plus loin dans sa caractérisation et mesurer les conséquences méthodologiques qu'elle entraîne ? La question qui est la nôtre est donc de comprendre quelle est la nature du modèle construit par l'IC et en quoi il est possible de le valider. La modélisation en IC se caractérise par le fait qu'elle renvoie à une réalité externe au modèle et par le fait qu'elle s'opérationnalise par un programme. Mais, la modélisation ne vise pas directement une réalité externe, mais les connaissances s'y rapportant, si on veut prendre au sérieux le fait qu'il s'agit d'une modélisation de connaissances. Or, les connaissances en question ne sont pas élaborées par l'IC mais modélisées et opérationnalisées par elle. Il en résulte que le lien empirique à la réalité n'est pas assumé par le modèle construit par l'IC mais par la connaissance représentée.

Comment alors envisager une validation et évaluation du modèle construit ? Comme peut-il prédire des phénomènes qui seraient infirmés ou confirmés par l'expérience ? Autrement dit, le terme de « modèle » n'est-il pas impropre ? En effet, un modèle de connaissances mécaniques n'est pas un modèle mécanique du monde. Autrement dit, la vérité revendiquée par le modèle d'IC n'est pas de décrire ou prédire le monde, mais les connaissances qui s'y rapportent. Dans ces conditions, toute application, tout montage empirique qui mobilisera un modèle d'IC ne permettra pas de valider un modèle, mais d'évaluer comment le travail des connaissances est facilité, amélioré par le système. Mais cette caractérisation va à l'encontre des revendications de l'IC comme science (par exemple (Charlet, 2001)). Globalement, nous pensons que, si l'IC veut être une science et donc valider ces modèles selon une attitude dite « scientifique », trois postures sont possibles :

La validation mondaine : le modèle de l'IC ne peut être validé empiriquement sinon en s'en remettant à la discipline ayant élaboré les connaissances modélisées par l'IC. En effet, en s'axant sur les connaissances, l'IC s'est en quelque sorte coupée du monde et ne peut y revenir directement, mais seulement médiatement. Le risque est alors que l'IC ne deviennent qu'une opérationnalisation de modèles venant d'autres disciplines, ses outils de modélisation étant progressivement intégrés et récupérés par les disciplines étudiant le monde.

La validation cognitive : le modèle de l'IC a pour ambition de refléter et se conformer à l'activité cognitive ; il prévoit, prédit et accompagne les connaissances qu'il faut

mettre en œuvre, sa pertinence découlant du fait qu'il reproduit le fonctionnement cognitif de l'utilisateur. Dans cette optique, l'IC opérationnalise des modèles issus des sciences cognitives au service desquelles elle doit se mettre.

La démonstration formelle : le modèle de l'IC a pour ambition d'être informatiquement opérationnalisé. Les modèles de l'IC doivent par conséquent, être nécessairement formel à un moment ou un autre. La validité des modèles de l'IC peut alors porter sur la puissance formelle de ses modèles, à la fois en termes de cohérence syntaxique, de complétude sémantique et de tractabilité computationnelle. Cet aspect de l'IC est évidemment très important, mais débouche sur le risque de voir l'IC devenir une province de la logique appliquée ou de l'informatique théorique.

En reprenant une distinction aristotélicienne matière/forme, l'IC est prise dans l'alternative selon laquelle la matière empirique qu'elle tire du monde ne lui est pas propre, et que la forme à laquelle elle recourt pour modéliser cette matière est empruntée aux disciplines formelles. Ainsi, si l'IC prétend faire des modèles, ce ne sont pas les siens et elle ne peut les valider. Elle devient une discipline ancillaire d'opérationnalisation informatique, la maîtrise technique des outils informatiques n'étant alors plus que le seul argument pour justifier la constitution d'une discipline indépendante et autonome comme l'IC. L'IC pourrait prétendre proposer des outils formels et non des modèles mais, on l'a déjà souligné, elle ne peut dès lors pas justifier sa différence à une logique appliquée. Dans ce contexte, l'IC ne peut faire d'expérience, car elle n'a rien à expérimenter. Elle ne fait que des applications. Applications dont la pertinence empirique lui échappe et dont l'intelligibilité théorique ne vient pas d'elle.

4 L'ingénierie des inscriptions numériques

Et pourtant, ce qu'on appelle les modèles de l'IC ne sont pas de simples applications. Bien peu de praticiens de l'IC se reconnaîtraient dans de simples utilisateurs de modèles ou formalismes importés. Il nous semble que leur point de vue est non seulement légitime comme objectif à atteindre, mais surtout fondé dans la réalité effective. Mais pour cela, il nous semble qu'il faut considérer pas tellement l'IC dans ce qu'elle dit mais plutôt dans ce qu'elle fait.

L'ingénierie des connaissances est une expression métonymique : l'IC n'est pas une technologie des connaissances mais une ingénierie des inscriptions numériques des connaissances. Nous justifions ce point de vue par deux arguments. Premièrement, une connaissance ne coïncide pas avec des objets matériels, elle reste une instance idéalisée qui permet de rendre compte par exemple que plusieurs textes expriment un même objet idéal, le théorème de Pythagore. La connaissance n'est donc pas un objet matériel sur lequel une technique peut avoir *directement* prise. En revanche, les connaissances ne s'appréhendent qu'à travers des inscriptions matérielles qui les expriment, et dont elles sont l'interprétation. Ainsi, pour atteindre la connaissance, il faut en passer par la médiation de son inscription. Deuxièmement, le numérique permet de conférer à un ensemble d'inscriptions données, les contenus numériques, une cohérence et une autonomie constituant un objet d'étude particulier et important. Le numérique est en effet un

support universel et homogène : universel car tout contenu peut être produit ou transféré sur un support numérique, homogène car il constitue un système technique homogène dont l'intégration va croissante, c'est la (trop) fameuse convergence.

Dans cette perspective, l'IC trouve un objet, les inscriptions numériques de connaissance. Elle trouve également un objectif : instrumenter le travail intellectuel ou cognitif, l'exercice des connaissances associées aux inscriptions manipulées par ses techniques. Autrement dit, l'IC ne cherche pas à modéliser le monde ou la pensée, ni à opérationnaliser des modèles venus d'ailleurs, mais à instrumenter et outiller le travail de la connaissance. Dans ce cadre, comment qualifier et caractériser les « modèles » élaborés par l'IC et les systèmes qui les opérationnalisent ? Pour cerner cette question, nous proposons de revenir à la notion d'ingénierie et de technique pour comprendre ce qu'implique une ingénierie de la connaissance comprise comme une technique manipulant des inscriptions numériques devant s'interpréter comme des connaissances.

L'ingénierie des connaissances propose des techniques qu'elle thématise et systématise comme une technologie, en comprenant par cette notion un ensemble de techniques gouverné par un savoir scientifique démontré et théorisé (le fameux *logos*). La technologie est donc la technique vue depuis le *logos* scientifique. Traditionnellement, le *logos* mobilisé par la technologie est un savoir scientifique issu des sciences de la nature : sciences physico-mathématiques (modélisation de la nature), sciences cognitives (naturalisation de la pensée). Cette posture, quand l'objet de la technique est la manipulation d'inscriptions, est incomplète car les inscriptions ne peuvent se réduire à une manipulation formelle prescrite par des sciences analogues aux sciences de la nature. Elles impliquent également une *interprétation* dans la mesure où il faut les aborder comme des inscriptions de connaissance. Les inscriptions numériques sont donc à la rencontre de deux paradigmes, que nous conceptualisons de la manière suivante :

Paradigme de la répétition : ce paradigme correspond aux sciences de la nature (physique, chimie, etc.) et aux disciplines formelles (logique et mathématique). Dans ce contexte, les phénomènes pris en compte sont théorisés et rapportés à des occurrences ou instances de types théoriques qu'ils matérialisent. Les types sont articulés en lois formelles dont le fonctionnement et l'application obéissent à des critères non-contextuels, universels et nécessaires. Le fait empirique est donc toujours compris comme la manifestation d'une loi déjà connue dont il répète l'instanciation. Les variations qui peuvent exister par rapport aux faits empiriques déjà connus sont inessentiels et sont rapportées à la variabilité prévue et autorisée par la loi théorique. Si la variation en revanche sort du cadre prévu par la loi théorique, le modèle théorique rentre en crise et une élaboration nouvelle est nécessaire. On parle, dans ce paradigme, de lois *nomographiques*¹, c'est-à-dire de lois qui synthétisent dans leur formulation tout ce qu'il faut savoir des faits qu'elles décrivent. Par exemple, la loi de la chute des corps synthétisent toutes les chutes passées et à venir et il n'est pas nécessaire de les décrire pour elles-mêmes. Contrairement au paradigme suivant.

Paradigme de la caractérisation : les faits pris en compte ne peuvent être ni décrits

1. Littéralement, décrire (*graphique*) les lois (*nomos*), en prenant la notion de loi au sens fort, c'est-à-dire prescrivant totalement le fait ou phénomène qu'elle subsume.

ni prédits par des lois formelles qui les synthétisent et qui en assument la variabilité et leur relation. Chaque fait doit être décrit pour lui-même, dans sa singularité. Mais, puisqu'une singularité absolue n'est pas connaissance, on la caractérise à partir d'un type théorique, d'une catégorie dont on montre en quoi elle en est différente et s'en écarte. Alors que les lois nomologiques réduisent les différences et les considèrent comme des instanciations variables d'une loi toujours identique, les lois *idiographiques*² propres au paradigme de la caractérisation décrivent les faits pour eux-mêmes, dans leur singularité, mais par rapport à une norme dont ils sont une déviation, une particularisation, mais jamais une simple instantiation. Les sciences idiographiques sont typiquement les sciences de la culture (Rastier & Bouquet, 2002) et les sciences historiques (Veyne, 1971).

Sciences nomographiques et idiographiques s'opposent dans la mesure où le fait « nomographique » est indépendant du contexte alors qu'en dépend le fait « idiographique ». En effet, si pour chaque fait, il faut prendre en compte le contexte, ce dernier étant à chaque propre à la situation, il en ressort qu'il est impossible de subsumer le fait sous une loi à chaque fois identique dont il répète l'instanciation. Un fait dont la description ne peut être isolée du contexte conduit donc nécessairement à une description idiographique. Or, l'interprétation des inscriptions numériques est à chaque fois contextuelle. Par définition, les inscriptions que manipule l'IC doivent s'interpréter comme des connaissances dans le contexte d'usage et la situation concrète des utilisateurs. Par conséquent, l'IC doit rapporter les inscriptions numériques à un double point de vue ; (i) selon un point de vue *nomographique*, les inscriptions sont des symboles manipulés par des techniques informatiques et formelles. La formalisation permet d'encadrer et de contraindre la manipulation symbolique et de prédire son comportement (telle expression se ré-écrit comme telle ou telle expression : inférence formelle et syntaxique). Selon un point de vue *idiographique* (ii), les inscriptions s'interprètent en fonction du contexte ; il n'est donc possible que de décrire le sens du comportement formel du système par rapport au système de normes, conventions, concepts du domaine qui le rendent signifiant par rapport à la pratique.

Ce double point de vue implique que l'ingénierie des connaissances est une technologie qu'il faut rapporter d'une part à un *logos* nomologique portant sur la formalisation logique des inscriptions et de leur manipulation et un *logos* idiographique portant sur la description de leur interprétation dans la singularité des situations effectives et concrètes d'utilisation. Cette description est une *critique*, au sens étymologique du terme³ : il

2. Littéralement, décrire (*graphique*) le singulier (*idio*). On parle aussi de sciences *nomologiques* ou de sciences *idiologiques* quand on considère les lois associées. Cette terminologie fut introduite pour la première fois par Wilhelm Windelband dans son discours de 1894 pour le rectorat de Strasbourg (Windelband, 1915). Cette distinction nomologique/idiographique fut approfondie par Heinrich Rickert qui lui substitua l'opposition entre les méthodes généralisantes des sciences de la nature et les méthodes individualisantes des sciences de la culture, en particulier l'histoire (Rickert, 1997). Ces propositions furent critiquées à leur époque par (Cassirer, 1977) puis condamnées sans ambage par (Aron, 1970). Malgré cela, elles gardent leur intérêt et peuvent être reprises. Nous détaillerons dans un travail futur comment dépasser les critiques auxquelles elles s'exposent.

3. Critique, *criticus* en latin, vient du grec *kritikos*, *krinein* : juger comme décisif. La critique est l'activité de l'esprit consistant à repérer et déterminer les facteurs décisifs, ce qui rend la situation et son évolution intelligibles. En ce sens, toute connaissance est critique : connaître, c'est savoir rapporter l'infini complexité d'une

s'agit de sélectionner, trier et hiérarchiser ce qui est pertinent et décisif dans la situation concrète ; cette sélection repère ce qui dans la situation est différente par rapport aux autres situations : ce qui, s'il était différent, ferait que la situation serait différente⁴. Dans cette optique, **l'ingénierie des connaissances est une technique des inscriptions formelles et une critique de leur interprétation.**

5 La médiation formelle de contenus documentaires

5.1 Interprétation typique et inscription idiographique

Si l'ingénierie des connaissances est une ingénierie des inscriptions numériques des connaissances, conjuguant technique formelle et interprétation critique, il est nécessaire de déterminer les orientations méthodologiques et thématiques qu'ouvrent cette perspective. La principale difficulté que doit affronter l'IC est d'articuler la technique formelle de manipulation syntaxique des inscriptions à leur interprétation. Dans cette optique, deux stratégies apparaissent. (i) La **représentation formelle des connaissances** où l'inscription est exprimée dans un langage logico-formel dont la syntaxe permet de contrôler la sémantique. Une sémantique formelle assure une correspondance nomologique entre la génération syntaxique des formules du langage et la composition sémantique de leur interprétation. (ii) L'**expression documentaire des connaissances** où l'inscription adopte une forme traditionnelle mêlant langues naturelles et formats documentaires qui régulent les modes de production, transmission et interprétation des contenus exprimés. Dans ce cadre, la forme matérielle et syntaxique des inscriptions ne permet pas par elle-même de déterminer l'interprétation qui en sera faite ; il faut lui adjoindre pour cela des considérations issues du contexte, tant dans la production que la lecture des inscriptions.

Le problème que rencontre la représentation formelle des connaissances est que l'interprétation faite des inscriptions ne peut prendre place que dans un modèle (au sens de la théorie modèle), qui est une construction mathématique et idéalisée (Bachimont, 2004). Autrement dit, le modèle qu'est l'inscription formelle n'est interprétable que dans un modèle théorique qui ne permet pas de déterminer par lui-même son rapport à une situation concrète effective. Il n'y a de sémantique formelle que parce que l'interprétation ne se fait pas dans le monde réel, mais dans des modèles mathématiques construits par la théorie des modèles. Le problème reste donc entier pour l'IC de rapporter ces modèles à la situation d'usage où un système à base d'inscriptions est utilisé pour instrumenter le travail intellectuel. Si bien que la représentation formelle des connaissances n'est pas une alternative à l'expression documentaire, mais seulement une étape préalable, pas toujours nécessaire, à la production d'inscription dont l'interprétation sera exogène et arbitraire par rapport à leur forme syntaxique.

Ainsi, en donnant une formalisation logique des inscriptions numériques des connaissances, sera-t-il possible de prédire non pas l'interprétation mais le comportement formel (les réécritures) qui seront faites. Mais il ne sera pas possible d'en prédire et

situation à ses éléments essentiels, c'est-à-dire ce qui régule et détermine sa nature et son comportement.

4. Ainsi, supposer que la température de la pièce est différente ne change rien au sens d'une lecture effectuée dans cette pièce. En revanche, changer ce que l'on sait de l'auteur influe le sens de la lecture.

déterminer l'interprétation. L'enjeu de l'interprétation ne sera pas d'en prédire les modalités, mais de permettre la description idiographique et critique de l'usage, en le caractérisant comme conformité et différence à des types, normes et conventions. Autrement dit, un système d'IC sera interprétable s'il fournit les éléments permettant de repérer ces écarts à la norme attendue (par exemple, le sens d'un mot dans le modèle du système par rapport à son sens usuel dans une pratique linguistique donnée).

L'IC et ses systèmes sont donc entre norme et calcul : norme sémiotique et culturelle à partir de laquelle mener l'interprétation critique concernant les différences « décisives » dans la compréhension et l'utilisation, calcul formel à partir duquel déterminer les actions du système et prédire son comportement (et non son usage).

Quel sens alors donner à la confirmation expérimentale d'un système d'IC ? Puisque l'objectif d'un système d'IC est d'instrumenter le travail intellectuel, il sera d'autant plus utilisable qu'il permettra l'interprétation critique des inscriptions. Par conséquent, la confirmation expérimentale sera donnée par la capacité du système d'intégrer et d'objectiver les écarts d'interprétation qui surviennent lors de chaque utilisation effective. Le système d'IC doit donc vérifier deux conditions. (i) **Expliciter les normes d'interprétation** dans la mesure où les inscriptions du système doivent comporter une explicitation des conventions et normes qu'il faut suivre pour les interpréter : thesaurus, terminologies, ontologies, etc. Ces normes précisent à l'utilisateur comment aborder la situation concrète d'utilisation et l'aident à cerner les différences pertinentes. (ii) **Inscrire idiographiquement les différences**, dans la mesure où les différences entre la situation effective et les conventions adoptées n'invalident pas le système, mais permettent de l'enrichir de la prise en compte des facteurs décisifs du nouveau cas : annotations, indexations, etc., et d'ajuster l'interprétation des inscriptions. Le système d'IC permet la mémoire de ses utilisations. Cette mémoire correspond à une sédimentation (ou accumulation dont on ne retient que le caractéristique) documentaire qui constituera une norme à laquelle confronter les futurs usages. L'explicitation des normes assurent l'intelligibilité du système, l'inscription idiographique son utilisabilité.

5.2 Le système typique d'ingénierie des connaissances

Ceci nous amène à brosser en guise de conclusion ce qui, selon nous, doit constituer un système d'IC type. Un système d'IC doit manipuler des inscriptions numériques en les instrumentant des différentes caractérisations critiques ou décisives de leurs usages. Par conséquent, un système d'IC est un système de *sources* documentaires, explicitant des normes d'interprétation, dont on sédimente les circonstances d'exploitation et dont on formalise les réécritures ou transformations possibles. *Le paradigme documentaire* nous fournit une illustration de cette approche. En proposant d'enrichir un contenu par des balises formelles contrôlées par une grammaire, l'ingénierie documentaire ne remplace pas une connaissance par une représentation formelle censée et destinée à en ternir lieu, mais conserve le contenu initial dans toute sa richesse interprétative propre, et propose d'instrumenter un usage donné par un registre de balises adéquates. Que ce soit avec HTML et XML, les balises ajoutent à un contenu des éléments permettant de manipuler la source pour un usage donné. Autrement dit, on peut considérer que chaque DTD ou schéma XML correspond à la formalisation d'un usage donné qui vient s'ajou-

ter au contenu et l'instrumenter sans prétendre s'y substituer. *Alors que* le paradigme de la représentation des connaissances entend *substituer* au contenu sa représentation formelle, le paradigme de l'expression documentaire des connaissances entend *enrichir* le contenu sans jamais le supprimer. Les balises sont alors les traces de l'usage que l'on veut sédimer pour le faciliter et préserver ses modalités d'exécution. Un système d'IC, dans ces conditions, pourra être vu en 3 couches :

La couche des sources documentaires : l'inscription n'est jamais brute et renvoie toujours à une expression située culturellement et socialement ; c'est ainsi que nous entendons ici la notion de document : une inscription résultant d'un processus social conditionnant sa production (un auteur faisant un roman, un médecin un compte rendu, etc.), sa transmission (le support, son organisation, par exemple un dossier patient en médecine), sa consultation ou lecture. La source documentaire possède avec elle les éléments permettant de la recontextualiser et de l'interpréter car elle appartient à une tradition sociale (les médecins apprennent à lire les dossiers, donc les inscriptions produites et transmises comme des dossiers patients sont lisibles).

La couche d'instrumentation documentaire : les sources peuvent être enrichies d'annotations ou balises consignnant des traces de l'usage, gloses (annotations) pour la lecture, instructions pour la transformation ou la publication, etc. C'est habituellement le niveau où s'exerce l'ingénierie documentaire. Mais on y intégrera également les formalismes comme les *Topic Maps* (Park & Hunting, 2003) qui permettent d'associer à des sources des concepts organisés entre eux. Également, on associera à ce niveau les outils de la documentation : thesaurus, index, terminologie, etc. Ces outils instrumentent le contenu sans le formaliser ni ne sont eux mêmes formels. Cette couche est cruciale : elle permet d'exprimer les normes que l'on associe aux sources pour les interpréter ; elle permet à l'utilisateur de cerner son usage par rapport à celui prévu et d'annoter sa propre interprétation en ajoutant ses différences.

La couche de formalisation du contenu : les annotations, gloses et balises peuvent être formalisées pour permettre des réécritures automatiques associant les sources et facilitant le travail de leur exploitation. On retrouvera à ce niveau les outils comme les ontologies et de manière générale les approches habituellement mobilisées par le Web Sémantique (Fensel *et al.*, 2002; Davies *et al.*, 2003).

Ces couches sont ordonnées, et nous plaçons qu'il ne faut pas envisager la formalisation pour elle-même, mais comme une instrumentation du contenu. Donc, si on a une couche de formalisation du contenu, c'est qu'on est passé par les deux précédentes. En revanche, on peut avoir des sources documentaires sans instrumentation, une instrumentation sans formalisation.

6 Conclusion

Les modèles de l'ingénierie des connaissances permettent d'explicitier des points de vue interprétatifs proposés à un utilisateur pour mobiliser des sources documentaires ou

contenus. Les modèles ne modélisent pas le monde, mais proposent une instrumentation pour exploiter des sources que l'utilisateur mettra lui-même en rapport à sa situation d'usage. Contextuelle, imprédictible, l'interprétation des sources n'est pas modélisable. Exploitant des sources instrumentées, l'utilisateur inscrit son propre usage en annotant le contenu des différences qu'il fait avec le point de vue proposé. Le système est une mémoire d'usages, qui sédimente les expériences passées pour penser les usages à venir. L'IC n'aura donc pas d'expérience scientifique, mais des expériences humaines d'utilisation dont il faut garder la mémoire. En s'intéressant à la connaissance, l'IC doit prendre acte qu'elle est critique, contextuelle et variable. Le formel n'est pas une réponse, mais une instrumentation.

Références

- ARON R. (1970). *La philosophie critique de l'histoire : essai sur une théorie allemande de l'histoire*. Paris: Seuil.
- BACHIMONT B. (2004). *Arts et Sciences du numérique : Ingénierie des connaissances et critique de la raison computationnelle*. Mémoire pour l'habilitation à diriger les recherches, Université de Technologie de Compiègne.
- BÉZIVIN J. (2001). From object composition to model transformation with the mda. In *TOOLS*, Santa Barbara: IEEE.
- CASSIRER E. (1977). *Substance et fonction : éléments pour une théorie du concept*. Paris: Les éditions de minuit. *Substanzbegriff und Funktionbegriff: Untersuchungen über die Grundfragen der Erkenntniskritik*. Berlin 1910.
- CHARLET J. (2001). Ingénierie des connaissances : un domaine scientifique, un enseignement? In J. CHARLET, Ed., *IC'2001*, p. 233–252, Grenoble.
- CHARLET J. (2002). *L'ingénierie des connaissances : Développements, résultats et perspectives pour la gestion des connaissances médicales*. Mémoire d'habilitation à diriger des recherches, Université Pierre et Marie Curie.
- J.-M. DAVID, J.-P. KRIVINE & R. SIMMONS, Eds. (1993). *Second Generation Expert Systems*. Springer-Verlag.
- J. DAVIES, D. FENSEL & F. V. HARMELEN, Eds. (2003). *Towards the Semantic Web : Ontology-Driven Knowledge Management*. Wiley.
- D. FENSEL, W. WAHLSTER, H. L. LIEBERMANN & J. HENDLER, Eds. (2002). *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press.
- NEWELL A. (1982). The Knowledge Level. *Artificial Intelligence*, **18**, 87–127.
- J. PARK & S. HUNTING, Eds. (2003). *XML Topic Maps : Creating and Using Topic Maps for the Web*. Addison Wesley.
- F. RASTIER & S. BOUQUET, Eds. (2002). *Une introduction aux sciences de la culture*. Formes sémiotiques. Paris: Presses Universitaires de France.
- RICKERT H. (1997). *Science de la culture et science de la nature*. Bibliothèque de philosophie. Paris: Gallimard. *Kulturwissenschaft und Naturwissenschaft*, J.C.B. Mohr (Paul Siebeck), 1926.
- VEYNE P. (1971). *Comment on écrit l'histoire*. Paris: Seuil.
- WINDELBAND W. (1915). *Geschichte und naturwissenschaft. discours de rectorat, strasbourg, 1894*. In *Präludien. Aufsätze und Reden zur Philosophie und ihre Geschichte*, p. 136–160. Tübingen: J.C.B. Mohr. *Histoire et Sciences de la nature*. Les études philosophiques, 2000(1).