



HAL
open science

Le métier : son savoir, son parler

Caroline Djambian

► **To cite this version:**

Caroline Djambian. Le métier : son savoir, son parler. TOTh 2011 - Terminologie & Ontologie : Théories et applications, May 2011, Annecy, France. pp.75-92. hal-00805596

HAL Id: hal-00805596

<https://hal.science/hal-00805596>

Submitted on 10 Apr 2013

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.

Djambian, C. (2011). Le métier: son savoir, son parler. In *TOTH Terminologie et Ontologie: Théorie et Application: actes de la V^e conférence internationale TOTH* (pp.75-92). Annecy: Institut Porphyre, *Savoir et Connaissance*. En ligne <http://www.porphyre.org/toth/files/actes/TOTH-2011-actes.pdf>

Le métier : son savoir, son parler

Caroline Djambian*

* LSIS, Faculté des Sciences et Techniques
Université Paul Cézanne Aix-Marseille III
cdjambian@yahoo.fr

Résumé. Le savoir de métier est ici au centre de nos réflexions. Connaissance expérientielle collective formée et formulée par le biais de concepts propres aux métiers, sa transmission est aujourd'hui en mutation dans les entreprises, posant des problématiques complexes comme dans le cas de la Division Ingénierie Nucléaire d'EDF. La réponse est dans la mise en forme de ce savoir désormais médiaté : le langage. Nous en arrivons donc naturellement à la terminologie qui, en tant que représentation linguistique de la *science* d'un champ social, en fixe le langage. Les ontologies qui modélisent et représentent formellement le système notionnel des terminologies, en sont la suite logique.

Nous illustrons notre réflexion en présentant une application concrète sur un domaine de la Division Ingénierie Nucléaire d'EDF, par la construction d'une base de connaissances constituée d'une terminologie et d'une ontologie centrée sur le *sens métier*.

Le métier : son savoir, son parler

1. Introduction

Métier et savoir cultivent une relation intrinsèque qui implique l'ensemble des connaissances nécessaires à une pratique de travail. La transmission du savoir s'est traditionnellement faite par une longue période d'apprentissage au contact des pairs, pour atteindre, par une pratique répétée, la maîtrise de l'art. Mais les situations de travail ont changé. Aussi comment s'intègre aujourd'hui cette dimension expérientielle dans l'acquisition de la compétence de métier ? La Division Ingénierie Nucléaire (DIN) d'EDF se trouve aujourd'hui confrontée à cette demande. Entre renouvellement générationnel des experts métiers où le manque d'anticipation entraîne la déperdition des savoirs d'ingénierie ; problématiques de durée de vie du parc nucléaire où les experts ayant conçu les matériels aujourd'hui en fin de vie ne sont plus dans l'entreprise ; et gestion de la documentation technique sensible où perdurent les connaissances requises mais à l'heure inexploitable, inorganisées et mal valorisées. Notre réponse consiste en des orientations opérationnelles pour les métiers cœurs de l'entreprise afin de faciliter la transmission des savoirs d'Ingénierie Nucléaire portés par les documents techniques.

2. Métier et savoir

2.1 L'expérience

L'expérience fonde le savoir d'Ingénierie Nucléaire dans des métiers où les compétences ne sont acquises qu'au sein de l'entreprise par une formation sur le tas d'au minimum un à deux ans. Dans cette longue tradition de compagnonnage, l'expérience, socle du savoir de métier, est aujourd'hui tronquée par le non-croisement des générations. Or, que recouvre cette notion ? L'expérience s'acquiert et sous-entend une temporalité, un vécu qui nous prémunit de l'aléa. Située comme l'une des trois dimensions de la vie humaine, avec l'action et la pensée, elle passe par la réceptivité, la perception (acquisition au contact des choses : la *métis*) et aboutit à ce que A. Barberousse (1999) nomme le « *jugement* », résultat de la pensée active et initiateur de l'action. Qu'elle soit perçue comme « *l'expérience instructrice* » nous permettant d'utiliser des concepts et de former des théories (l'empirisme moderne de Locke (1989) la voit comme seule fondatrice de la connaissance) ou comme l'expérience en tant que « *passivité féconde* » (activité créatrice du jugement où nous recevons passivement de l'information par l'intermédiaire de l'expérience), de nombreux textes depuis l'antiquité se préoccupent de son articulation à la connaissance. Il en ressort que l'expérience peut légitimement se voir attribuer un rôle primordial dans la formation de cette dernière.

Le point de vue d'Aristote (1986) dans sa *Métaphysique* est que « *c'est de la mémoire que provient l'expérience pour les hommes : en effet, une multiplicité de*

souvenirs de la même chose en arrive à constituer finalement une seule expérience ; et l'expérience paraît bien être à peu près de même nature que la science et l'art, avec cette différence toutefois que la science et l'art adviennent aux hommes par l'intermédiaire de l'expérience, ... ». Cette expérience, appréhension du général, aboutit à l'*emperia*, la connaissance empirique, celle de ce qui nous entoure immédiatement et qui peut devenir *science* grâce à l'exercice du raisonnement. A ses côtés Aristote place l'*art*. Il est ici celui du métier, de la technique. Ainsi, peut-on l'assimiler à une mémoire collective technique quand « *d'une multitude de notions expérimentales, il se dégage un jugement universel, applicable à tous les cas semblables* ».

Connaissance collective	$\left\{ \begin{array}{l} \text{Emperia + Raisonnement = Science} \\ \text{Emperia + Jugement universel = Art} \end{array} \right.$
Connaissance individuelle	
	Pratique

Face à ces deux notions se dresse la *pratique* où l'on ne peut tout simplement pas se passer de l'expérience. Ici on ne parle plus d'une connaissance collective, mais de la connaissance individuelle de l'homme de métier, puisque la connaissance de l'universel n'est rien sans l'expérience. Ainsi, les ingénieurs de la DIN se trouvent aujourd'hui armés d'un bagage théorique reçu en école. Il ne suffit en aucun cas à acquérir les compétences de métier d'Ingénierie Nucléaire que seule la pratique au contact des pairs peut façonner. La qualification théorique s'articule au savoir-faire pratique : c'est la notion de compétence. La pratique est donc un passage indispensable pour maîtriser la connaissance collective et par là-même accéder à une légitimité dans le groupe. L'expérience atteint le statut d'*art* au moment où l'individu devient capable de transmettre sa connaissance. Les experts de métier qui ont construit leur expérience sur l'ancien système de compagnonnage grâce à un long apprentissage pratique, aboutissent sur une base théorique à une nouvelle théorie améliorée de leur expérience. Ils sont le point central de la transmission de l'expérience technique.

2.2 Formation et formulation des connaissances expérientielles : les concepts

La relation connaissance et expérience est aussi un thème majeur de *La critique de la raison pure* de Kant (1997), où nous retrouvons la notion de « *connaissance empirique* » (cf. *l'emperia*) que l'auteur oppose à celle de « *connaissance pure* » (le raisonnement ou ici, entendement). Pour Kant, nous pouvons connaître de deux façons : avec ou sans l'aide de l'expérience. Donc, si l'expérience n'est pas le seul ingrédient nécessaire à la formation de la connaissance dans le système de Kant, il convient de préciser comment interviennent les processus issus de l'entendement, ce qu'il nomme « *concepts* » ou encore « *catégories* ». Il décompose ainsi la connais-

Le métier : son savoir, son parler

sance en deux éléments complémentaires qui sont le concept et l'intuition (pure ou empirique, produite dans notre rapport expérientiel au monde). « *Se forger la pensée d'un objet et connaître un objet, ce n'est donc pas la même chose. A la connaissance appartiennent en effet deux éléments : premièrement le concept, par lequel en général un objet est pensé (la catégorie), et deuxièmement l'intuition, par laquelle il est donné ; car si une intuition correspondante ne pouvait aucunement être donnée au concept, il serait formellement une pensée, mais dépourvue de tout objet, et par son intermédiaire ne serait possible absolument aucune connaissance d'une quelconque chose,...* ». L'expérience au vrai sens du terme est effectivement *l'emperia*, c'est-à-dire l'application de concepts à des intuitions empiriques. Le concept est donc inscrit au cœur de l'expérience et s'il est essentiel dans sa formation, il l'est aussi dans son expression.

De façon générale, la question de la relation de l'expérience aux concepts a été étudiée selon des perspectives très diverses : soit que l'on considère que les concepts déterminent le contenu même des expériences perceptives ; soit que l'on cherche à repérer des expériences ayant un contenu non conceptuel ; soit que l'on étudie la façon dont on exprime nos expériences par le langage afin d'y déterminer le rôle des concepts comme Carnap (1981) ou Quine (1980)... Wittgenstein (1961) dans ses *Investigations philosophiques* s'est également intéressé à l'expression de l'expérience par le langage. Il s'oppose à l'idée selon laquelle un langage privé, destiné entre autre à porter sur nos diverses expériences, serait possible et prône un langage avant tout caractérisé par son « *usage public* ». Cela implique que ce système de concepts partagés réfère à un système de connaissances elles aussi partagées par des individus ayant une base expérientielle assimilable, une culture commune, notion que l'on retrouve chez Husserl (1970).

3. Le savoir de métier diffusé

3.1 La « science » médiatée

Ce que B. Lamizet (1992) présente sous le terme de « *science* » (à rapprocher de « *l'art* » d'Aristote) est une culture collective construite et validée au sein de l'institution et sur laquelle le sujet s'appuie pour authentifier son discours dans le champ social que représente l'entreprise. L'individu y perd sa subjectivité pour n'être reconnu dans l'espace social de la diffusion d'information que comme porteur de savoir. Certes, un document, une information sont toujours reconnus comme émanant d'un auteur, mais cet auteur l'est dans son champ institutionnel et sa production d'information n'est justifiée que par le savoir dont il est porteur et qui lui est reconnu. Il parle « au nom de » l'institution. Dans l'échange d'information, le savoir émane d'un individu pour être ensuite réapproprié par un autre et par la diffusion de ce savoir, il entre dans un champ collectif où il devient *science*.

Le savoir est donc approprié par et attribué à un sujet. Il est essentiel d'avoir conscience de ces trois étapes où, dans l'information diffusée, le savoir part d'un sujet pour rentrer dans le domaine institutionnel sous forme de *science* et être ensuite réapproprié par un tiers. Dans la tradition de compagnonnage dont faisait l'objet la transmission de savoir au sein de la DIN d'EDF, le savoir n'était pas médiaté et passait directement d'un expert vers un jeune ingénieur. Cette pratique étant remise en question par les changements internes à l'institution, la transmission du savoir repose de plus en plus sur cette étape transitoire jusque là quasi inexistante, de mise en suspension dans l'espace public en l'absence du sujet connaissant. Les échanges très informels qui avaient cours jusqu'alors grâce à une identification claire des partenaires de la communication, sont aussi progressivement troublés et amenés vers des modes de communication médiatés. C'est toute une culture, des structures d'échange et transmission des savoirs établies au sein d'une communauté d'usage, qui sont modifiées.

Au sein de notre structure, la médiation s'opère par les textes techniques. Le contexte d'Ingénierie Nucléaire (de par les enjeux et les exigences réglementaires imposés à ce domaine) sous entend un consensus fort sur les références communes, leurs règles et normes de transmission, fixées par l'usage social. Le fait qu'elles soient portées dans le discours de la *science*, démontre qu'elles ont été éprouvées dans leur potentiel de « réappropriabilité ». C'est ce qui fait leur intérêt pour nous.

3.2 Le langage, cristallisation du savoir

Ce processus de médiation des savoirs collectifs fait évidemment appel au langage. Comment pourrait-il en être autrement ? Pour B. Lamizet (1992) « *Le langage est un code de signification : il s'agit d'un système qui produit des représentations du réel et qui met en œuvre des significations dans le cadre de relations d'équivalence et de représentation, conventionnellement établies et stabilisées. Les langages de l'information vont occuper, au sein de la communication, la place d'une médiation généralisée. Les langages de l'information constituent le lieu de la communication où s'opère la médiation qui donne lieu, à partir du réel dont elle se soutient, à l'avènement de la symbolisation qui rend possible la structuration des échanges de sens* ».

La mise en œuvre des langages passe bien évidemment par l'énonciation qui couvre deux caractéristiques : l'individualisation, partant d'un sujet et reçue par un autre ; l'actualisation, puisqu'émetteur et récepteur lui donnent sens par des références à leur réel. En ce qu'il est, par l'information, la mise en forme du réel dans le champ du symbolique (la communication), le langage est un sujet d'étude particulièrement intéressant. Puisqu'il est relatif à un réel donné, il est représentatif de la structure dans laquelle il a été énoncé. Il est le témoin d'un champ défini et de sa culture puisqu'il est la matérialisation de l'information qui y circule. Si le savoir représente la trace d'un champ social, le langage est la cristallisation de cette trace.

Le métier : son savoir, son parler

Selon les situations de communication durant lesquelles ils sont actés, les langages de l'information peuvent prendre trois formes. Les langages de symbolisation produisent des systèmes de symboles et de représentation. Nous nous y intéresserons peu, bien qu'ils soient extrêmement présents dans les documents techniques de l'Ingénierie Nucléaire (ex : sous la forme de schémas). Les langages de représentation rendent compte du réel. Cela peut être sous la forme de la description (ex : dans le discours scientifique) pour décrire le réel ou pour aller vers un consensus symbolique de l'information diffusée dans le champ social. Cela peut être sous la forme de l'appropriation de l'information par les partenaires de l'échange. Ce sont donc eux qui vont faire émerger l'opinion et le savoir. Les langages d'opération sont quant à eux les principaux langages exprimés dans les textes techniques de l'Ingénierie Nucléaire. Ils font de la référence un outil opératoire puisqu'ils visent, grâce à cet outil, l'obtention de résultats et de nouvelles références. Ce sont donc par essence, les langages de la constitution du savoir tel que nous l'avons défini plus haut. Leur but de création de nouvelles références les distingue des langages de description qui se contentent de représenter le réel. La référence n'est donc plus qu'une représentation du réel, mais devient également un objet de *science* sur lequel on pourra produire des opérations. L'enjeu est donc de capter le savoir de métier à travers le langage et d'en faire émerger les références communes suffisamment stables pour être partagées par tous. Dans un cadre technique comme le nôtre, à la culture métier excessivement prégnante et en réponse aux diverses problématiques de gestion et transmission des informations et savoirs du domaine, l'étude de ce langage d'opération que nous nommerons à compter de maintenant « terminologie métier », paraît cruciale.

3.3 La terminologie métier

La terminologie est l'outil qui fixe le langage spécialisé d'une communauté de pratique. Le travail terminologique oriente pour S. Lainé-Cruzel (2006) « l'interprétation qui doit être associée à chaque terme, et définit le terme à utiliser pour désigner un concept ou un objet », soit la manifestation linguistique à associer à ce que C. Roche (2005) nomme la « *réalité extralinguistique partagée* » par un métier, une communauté. La consensualité autour de cette terminologie, intrinsèquement liée à un contexte spécifique, assure une bonne transmission, réception et appropriation de l'information et des savoirs qu'elle véhicule. Appliqué au champ industriel, le travail terminologique vise la normalisation et la clarification du langage d'un domaine et de ses significations afin de réduire la marge d'arbitraire inhérente au langage et d'aller vers un système purifié et donc plus efficient.

Gouadec (1990) rappelle que « l'élaboration d'une terminologie est une opération intellectuelle, humaine et extrêmement complexe » dont chaque étape pose de multiples questions pour lesquelles les avancées de l'Intelligence Artificielle (IA) apportent aujourd'hui une assistance précieuse. En s'appuyant sur ces techniques on trouve schématiquement depuis 2005 deux voies en terminologie. L'une est une démarche purement onomasiologique, se basant sur les experts du domaine. L'autre,

dite sémasiologique, relève de la sémantique distributionnelle et vise la classification de gros volumes d'informations. La différence d'approche revient pour N. Ausse-nac-Gilles et A. Condamines (2000) à la distinction entre « *la conception de connaissances formelles structurées dans les ontologies, et la conception de ressources terminologiques issues de textes* » : les résultats en sont-ils assimilables, les méthodes en sont-elles unifiables ?

Mais le travail terminologique est en réalité la part préliminaire incontournable et indissociable du travail ontologique. La terminologie est trop souvent perçue comme ne s'intéressant aux mots que pour eux-mêmes. Elle vise en fait les notions que ces mots désignent. C. Roche (2005) l'évoque comme un « *système de termes reflétant une modélisation conceptuelle* ». Deux modèles se superposent, l'un linguistique (l'objet) et l'autre extralinguistique (le concept), ce dernier pouvant recouvrir plusieurs modèles linguistiques lorsque plusieurs communautés de pratiques (ayant chacune sa langue d'usage) partagent la même conceptualisation du monde. Le langage, cristallisation du savoir, n'est alors pas dissociable du concept qui est au centre de la formation de ce savoir. Afin de modéliser et représenter formellement le système notionnel des terminologies, soit la couche extralinguistique, la notion d'ontologie est à ce jour l'une des approches les plus intéressantes. Le système notionnel qu'elle représente, associé à un vocabulaire de mots constitué des noms des concepts, forme « le pile et le face » d'un même travail.

L'extraction à base de textes est le socle de nombreux travaux de construction de terminologies et d'ontologies, les textes étant la trace souvent la plus exploitable et stabilisée des connaissances d'un domaine. Cette méthode repose sur des procédés statistiques exploitant par exemple une analyse distributionnelle de type Harris (1968) et/ou linguistiques. Le résultat est le lexique de termes et de mots d'usage de la langue de spécialité qui sert de base à l'élaboration de la structure conceptuelle de l'ontologie. Une fois validée par les experts, elle pourra être érigée en ontologie du domaine. Pour ne citer que quelques exemples, les travaux de BCT (Bases de Connaissances Terminologiques) portés par A. Condamines et N. Ausse-nac-Gilles (2000), se basent sur des techniques de linguistique de corpus, dans une influence terminologique. Mais les « *ontologies régionales* » de B. Bachimont (2000) sont sans nul doute la démarche à laquelle nous nous sommes le plus référés. Cette approche a influencé les réflexions issues du groupe TIA (Terminologie et Intelligence Artificielle). Les textes y sont effectivement considérés comme la source presque exclusive des connaissances, le problème étant alors de construire des modèles de connaissances à partir de l'expression linguistique de ces connaissances. B. Bachimont propose des réponses en plusieurs étapes, dont la normalisation fondée sur des principes de différenciation, en référence aux principes différentiels d'Aristote. D'autres travaux menés au sein d'EDF ont également servi d'appui à notre méthode qu'ils visent la mise en cohérence de l'information technique comme ceux d'H. Boccon Gibod (2006) ou la réappropriation des connaissances avec C. Roche (2007).

Le métier : son savoir, son parler

Notre expérience, développée dans C. Djambian (2010), nous a prouvé, comme dans ces travaux, que la structure conceptuelle construite à partir d'un corpus est effectivement une très bonne base de travail pour la construction d'une ontologie, mais qu'elle ne peut en aucun cas constituer une ontologie en elle-même. Le passage de l'une à l'autre nécessite au contraire de lourds remaniements puisque, comme l'affirme Rastier (2004) « *le lexique des langues ne reflète pas la conception scientifique du monde* ». Ce sont les connaissances extralinguistiques qui vont permettre de lier ces deux plans et de comprendre les textes produits, diffusés, appropriés au sein d'une communauté de métier. En effet, l'usage de figures de style y est courant pour exprimer les concepts du domaine : nous les nommerons « parler métier ». Elles expriment de façon tacitement conventionnelle une conceptualisation qui est en fait l'ontologie du domaine¹. L'aide des experts est par conséquent indispensable pour déduire l'ontologie de la structure lexicale issue de l'extraction à partir des textes. Cette base permet de faire exprimer les concepts réellement désignés par les tropes et d'évoluer progressivement vers une représentation conceptuelle. C. Roche (2007) conseille cependant de toujours garder en mémoire que « *si les notions de langue de spécialité (ou langues spécialisées), terminologie et ontologie entretiennent certains rapports, elles ne se recouvrent pas. Ainsi, si la langue s'intéresse prioritairement aux relations entre signifiants et signifiés, la terminologie et l'ontologie s'intéressent principalement aux rapports entre concepts et objets : le signifié n'est pas un concept* ».

4. La méthode : une base de connaissances centrée sur le sens métier

4.1 Délimitation du domaine

Nous souhaitons démontrer sur un domaine particulier de l'Ingénierie Nucléaire ce qui pouvait être réalisable et élargi ultérieurement. A partir de compétences stratégiques de l'entreprise, comme la Sûreté Nucléaire, transversale et animée d'une véritable culture de métier, nous avons choisi le domaine « Accidents Graves », où de forts besoins en termes d'appropriation des connaissances étaient exprimés par les nouveaux arrivants. L'identification du domaine, une fois confirmée par les responsables concernés, s'est basée sur des entretiens semi-directifs, menés auprès de l'ensemble des ingénieurs de la compétence. Nous avons ainsi confirmé les besoins et recensé des ressources informationnelles formant une collection de documents de référence (environ 700), documentation alors noyée dans la masse de gestion courante et nécessitant expressément d'être épurée et valorisée.

¹ Les traces de ce parler métier se retrouvent essentiellement dans le langage oral et connaissent des variantes mêmes infimes, entre les diverses communautés d'un même métier. On se trouve face à un langage à plusieurs strates : officiel et usuel avec des variantes dans les usages.

4.2 Choix du corpus de textes

A compter de cette étape, notre travail s'est basé sur l'échange constant avec un expert du domaine. Après lui avoir clairement exposé les finalités du choix du corpus, l'expert a commencé à sélectionner parmi la documentation de référence, une vingtaine de documents techniques. Puis rapidement, il a préféré recentrer son choix sur 8 documents qu'il estimait couvrir 90% des concepts métier. Il s'agit de documents stratégiques, états de l'art de recherche et développement, dossiers destinés à l'Autorité de Sûreté Nucléaire et procédures, tous été cités par les ingénieurs lors des entretiens, comme ayant une forte légitimité et valeur pour le domaine (contenu, pérennisation et richesse des concepts métier exprimés). Nous avons proportionnellement accordé un temps relativement long à cette étape amont, pour cibler un corpus très pertinent de 570 pages.

4.3 Extraction des lexiques

Traitements initiaux

L'analyse linguistique du corpus a été réalisée par l'Equipe Condillac avec l'outil LCW. Le logiciel effectue dans un premier temps une lemmatisation des textes, puis génère un lexique de termes (mots simples et expressions). L'identification des expressions se fait à partir d'un dictionnaire de patrons lexico-syntaxiques : (SBC = substantif, SBP = nom propre, PREP = préposition, DTN, DTC = déterminant, ADJ = adjectif, ADJ2PAR = participe passé, INC = mots inconnus du dictionnaire). Dans un premier temps, un lexique de 20 026 syntagmes nominaux a été généré. Les termes ayant une fréquence d'apparition dans le corpus inférieure ou égale à 10 n'étant conservés, le lexique se composait au final de 770 syntagmes nominaux.

Mais les constats suivants se sont d'emblée imposés :

1/ La non prise en compte des acronymes s'est avérée très gênante, puisqu'elle éliminait des notions cœurs du domaine (voire les plus importantes), concernant les projets, outils, matériels, phénomènes physiques... toutes exprimées par de longues propositions (ex : EPS (Etude Probabiliste de Sûreté), GIAG (Guide d'Intervention en Accident Grave), DCH (Direct Containment Heating), circuits RIS-EAS (injection de sécurité - aspersion enceinte) ...). Les cas étaient nombreux et les termes formant ces propositions ne se retrouvaient pas de façon isolée dans le lexique.

2/ Il en était de même pour toutes les expressions métier contenant des chiffres ou caractères spéciaux (ex : mode α , filtre U5, combustible O2, Bugey post-VD3, WASH 1400, dispositifs H4-U3 ...).

3/ Beaucoup de termes au score inférieur à 10 et donc non retenus, étaient pertinents et des expressions plus claires y apparaissaient.

Le métier : son savoir, son parler

Le point 3 s'explique évidemment par la petite taille du corpus. Pour le point 1, les exemples évoqués commencent par une majuscule et sont donc étiquetés par le logiciel comme "SPB" (nom propre). Or les patrons lexico-syntaxiques utilisés commençaient tous par "SBC" (pour nom commun). Nous avons donc décidé de générer un nouveau lexique contenant uniquement les termes en majuscule, en conservant les chiffres. Ce deuxième lexique de 1710 acronymes a été extrait à partir du même corpus. Tous les niveaux d'occurrence ont été conservés et aucun traitement n'a été effectué par l'Equipe Condillac (contrairement au premier lexique), celui-ci nécessitant une connaissance plus approfondie de l'entreprise et du domaine.

Pour les deux lexiques, il aurait été nécessaire de mieux cibler les traitements automatiques (ex : éliminer certaines parties des documents) pour éviter de lourds re-traitements manuels. Dans le cas du premier lexique, ce tri a posteriori a abouti à 746 syntagmes nominaux. Le lexique d'acronymes comportant quant à lui énormément de bruit, a nécessité une contextualisation minutieuse de chaque acronyme dans les textes pour ne conserver que 1298 acronymes.

Il est intéressant de noter que les deux lexiques correspondent à des pratiques différentes et complémentaires : le lexique des syntagmes nominaux fait appel à des notions globales ou stratégiques, alors que celui des acronymes est plus proche du « terrain » et nomme essentiellement des matériels, actions, processus, projets...

Validation par les experts

Sur cette étape nous avons mis à contribution un deuxième expert et présenté nos deux lexiques de 2044 termes pour validation. L'objectif a été de ne garder que les expressions relatives aux Accidents Graves et de compléter les manques éventuels.

Le lexique des syntagmes a été validé par le second expert ayant une vision plus stratégique. Il a consacré 3 heures de travail à cette tâche. 62% des termes pré-triés ont été supprimés et 63 nouvelles expressions ont été ajoutées. Plusieurs syntagmes ajoutés existaient dans le lexique original avec une fréquence inférieure à 10 et donc non retenus. Le résultat est un lexique de 344 termes. Un travail similaire a été réalisé par l'expert principal, plus proche de la « pratique », sur le lexique des acronymes. Une heure de travail a été consacrée à cette tâche. Un nombre important d'acronymes ont été supprimés (76%) car ne relevant pas spécifiquement du domaine. 23 nouveaux acronymes ont été ajoutés. Le résultat est un lexique de 333 acronymes. L'autre expert a également été consulté et y a aussi consacré 1h. Avec 76% de suppressions et 13 ajouts (correspondant à une vision plus théorique du domaine), il restait 325 acronymes. Une suppression des doublons aboutit à un lexique de 264 acronymes. Il a été décidé de concaténer les deux validations et de conserver tous les termes y apparaissant. Après un dernier traitement des doublons, on obtient un lexique de 343 acronymes. Il ne reste donc que 26% du lexique tel que présenté aux experts (après deux tris des traitements automatiques) et 20% du lexique initial qui de toute évidence aurait pu être plus affiné.

4.4 Principes de classification des Accidents Graves (AG)

Etape intermédiaire

A partir de ces lexiques et des indications données par les experts, l'ébauche d'un réseau lexical a été réalisée à base de relations linguistiques (synonymie, d'hyponymie, ...). Un tel réseau devient très vite inextricable. De plus, il paraît difficile de demander aux experts de construire un réseau sémantique directement à partir de 687 termes.

Cet exercice nous a confrontés à la nécessité d'acquérir une connaissance suffisamment solide du domaine, que nous avons fondée sur l'étude approfondie et systématique des textes du corpus et du sens des termes manipulés. Nous avons souhaité dans un premier temps, structurer le vocabulaire relatif aux Accidents Graves pour dessiner progressivement les grandes notions du domaine, autour de trois modélisations principales : le déroulement d'un AG ; les dispositions AG ; la performance des équipements. Ce travail a permis de s'interroger sur la signification des termes des lexiques et de comprendre à quel niveau ils intervenaient dans le déroulement de l'accident grave. Des termes intervenant dans de nombreuses désignations d'AG n'étaient définis d'aucune part, comme si l'usage répété dispensait d'une définition officielle. Les interviews menés précédemment auprès des ingénieurs avaient permis de situer les spécialistes des diverses branches du domaine. Ils ont eux-mêmes éclairci les notions clés en les replaçant dans le contexte général de l'accident grave. Il est ainsi apparu que la majorité des notions s'organisaient d'un point de vue phénoménologique.

Construction du réseau conceptuel

L'approfondissement des notions clés a permis de construire une modélisation « déroulement d'un Accident Grave » composée des sept vues, que nous avons soumise à un ingénieur du métier. Il a jugé les concepts et leur organisation pertinents, mais le découpage en diverses parties de l'accident grave difficile à aborder et trop détaillé. Il a donc proposé de synthétiser le tout sur un grand modèle offrant une vision globale. Il s'est en fait dirigé d'emblée vers une organisation des connaissances de type expertes. En logique épistémologique on distingue en effet, trois niveaux :

1. la logique des termes, ontologique et terminologique, niveau de définition des concepts et de la façon de les nommer.
2. la logique des jugements, niveau relationnel : définition des relations (autres que spécifiques) entre les concepts.
3. la logique du raisonnement, niveau de déduction, exprimant les connaissances expertes de type si...alors.

En suivant le mode de raisonnement des ingénieurs, la redistribution des vues précédentes s'est faite facilement pour réaliser au final une vue générale d'un AG en fonction des divers phénomènes pouvant y intervenir. D'autres vues ont été créées

Le métier : son savoir, son parler

en complément. Nous avons donc à ce point les vues suivantes : déroulement d'un AG ; événements initiateurs ; perte du confinement ; rejets ; dispositions AG. Nous avons pris soin de distinguer dès cette étape, les termes d'usages des termes normés et les dénominations des concepts, pour nous concentrer sur le niveau conceptuel en sortant de la langue (liée aux lexiques). Les Relations créées sont multiples : est soit ; et/ou (relation demandée par l'expert, en remplacement de « est soit » dans certains cas) ; si (type de séquence mais avec option) ; séquence (séquence logique sans option) ; voir (pour les éléments plus anecdotiques). Les concepts ont été organisés selon plusieurs Ensembles : pré-AG ; progression en cuve ; progression hors cuve ; parade ; exigences de sûreté (ajouté par l'expert).

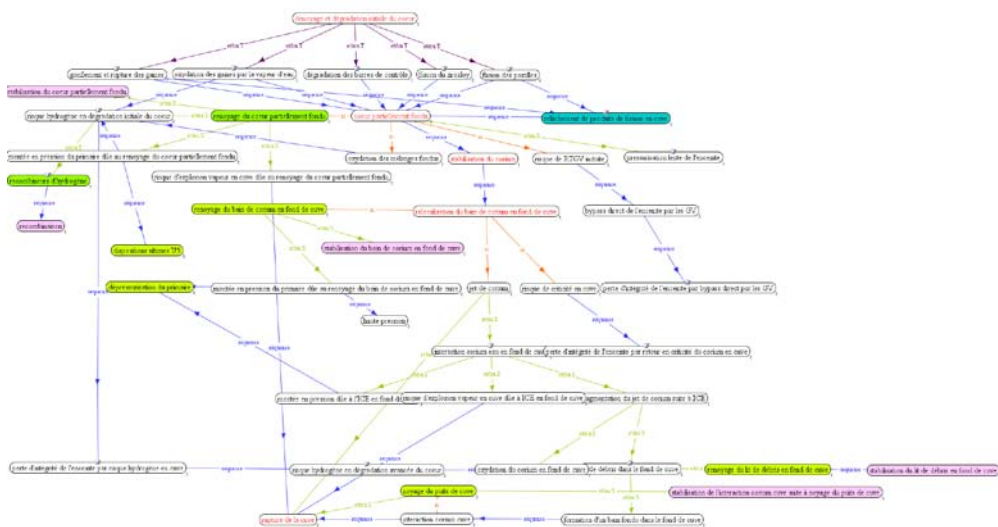


Fig. 1 – Ensemble « progression en cuve ».

A notre grande satisfaction les modélisations se sont avérées très pertinentes, ce qui a nettement soulagé la validation de l'expert : une heure a suffi pour contrôler l'ensemble des concepts et leurs relations. L'expert s'est basé sur les phénomènes physiques qu'il avait identifiés pour son propre compte dans un état de l'art du domaine, pour ne négliger aucune notion. Il a souhaité conserver nos dénominations, jugées plus génériques, au lieu de celles utilisées dans son document, ce qui prouve qu'elles lui parlaient indépendamment de ses usages. Tous les concepts étant identifiés et clairement nommés, nous avons complété certaines dénominations pour qu'elles suffisent à situer le concept dans le réseau notionnel.

4.5 Définition de l'ontologie depuis le réseau conceptuel

Puisque nous étions à ce stade déjà dans le conceptuel, la suite logique était de regrouper de manière plus spécifique les relations pour faire émerger ce qui est de nature ontologique. Nous avons commencé par clarifier les relations et/ou en distinguant deux catégories : et/ou T (pour « type de ») ; et/ou S (pour « séquence »). Nous sommes ensuite repartis de nos modélisations pour restructurer tous les concepts en ensembles sémantiquement liés avec des relations de type uniquement générique/spécifique : « est une sorte de », ce qui a demandé une réorganisation complète des concepts. Il a aussi été nécessaire de créer de nouveaux concepts plus généraux pour catégoriser les phénomènes physiques intervenant lors d'un AG.

En nous aidant de nos premiers modèles beaucoup plus détaillés, nous obtenons au final 11 ontologies AG : évènement initiateur ; dénoyage et dégradation initiale du cœur ; dégradation avancée du cœur ; rupture de la cuve ; progression de l'AG hors cuve ; perte d'intégrité de l'enceinte ; rejets radioactifs ; comportement des produits de fission ; dispositions AG ; parades ; exigences de sûreté.

Les ontologies ont été construites selon la méthode de différenciation spécifique et traduites en OWL en utilisant l'outil OCW de l'Equipe Condillac. L'expert a ensuite validé les ontologies sur l'organisation des relations et la présence de l'ensemble des concepts précédemment identifiés dans les réseaux conceptuels.

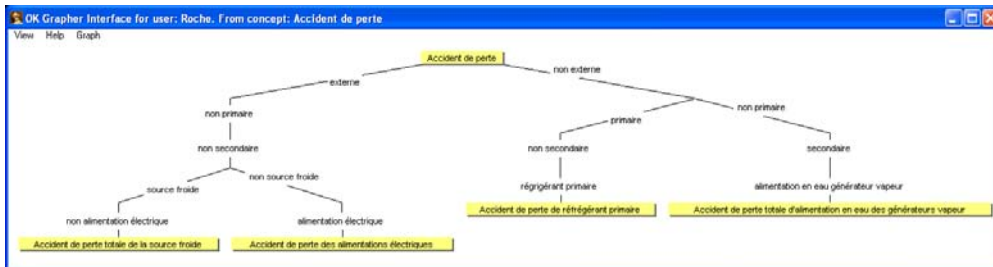


Fig. 2 – Détail d'« Evénements initiateurs » : les « Accidents de perte ».

Le métier : son savoir, son parler

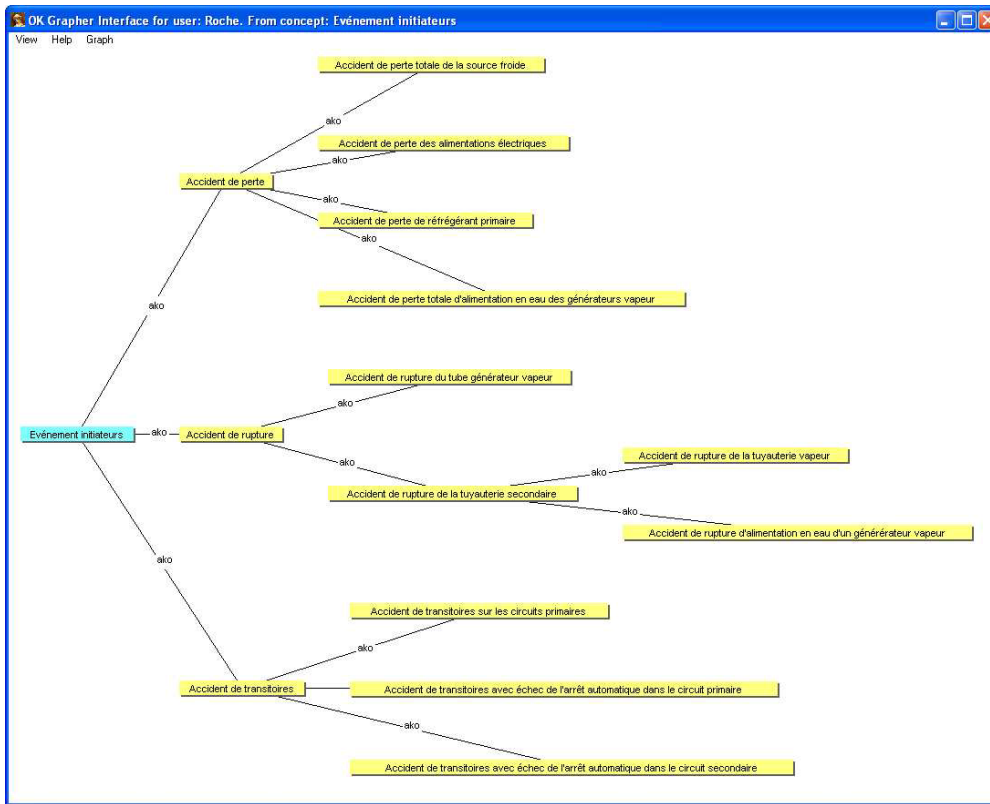


Fig. 3 – Vue globale d'« Événement initiateur ».

4.6 Définition de la terminologie

Cette étape est nécessaire pour faire le lien entre l'ontologie et les documents. Elle consiste à associer chaque terme des lexiques (687 termes) aux concepts de l'ontologie, identifier les synonymes, les acronymes et leurs développés... Suite à l'ultime validation de l'expert, cette terminologie a été convertie par l'Equipe Condillac au format XML, compatible avec leur environnement TCW et la plupart des outils utilisés aujourd'hui dans les entreprises.

La collection de documents précédemment recensés a été retriée pour en conserver au final 415 dont 333 au format image (inexploitables pour l'utilisation escomptée). Leur océrisation a été réalisée dans le cadre d'une démarche globale de numérisation de la documentation technique de l'entreprise, de sorte à permettre leur indexation. Une plateforme de test, présentée dans la figure suivante, a été mise à disposition des ingénieurs du domaine « Accidents Graves ». L'indexation des documents sur l'ontologie permet une recherche par navigation dans les concepts (accès à la documentation plus interactif) ou en langage naturel. Les termes d'usages et

leurs relations linguistiques sont utilisés pour l'indexation et l'expansion de requête. Ainsi, par exemple, dans le cas de la requête « *EDE* », l'outil signalera une ambiguïté et proposera les notions de « *mise en dépression de l'espace entre enceinte* » ou « *échauffement direct de l'enceinte* ». Dans ce second cas, l'outil remontera également des documents traitant de « *DCH* » ou « *direct containment heating* », équivalent anglais utilisé de façon indifférenciée dans la langue d'usage du métier.

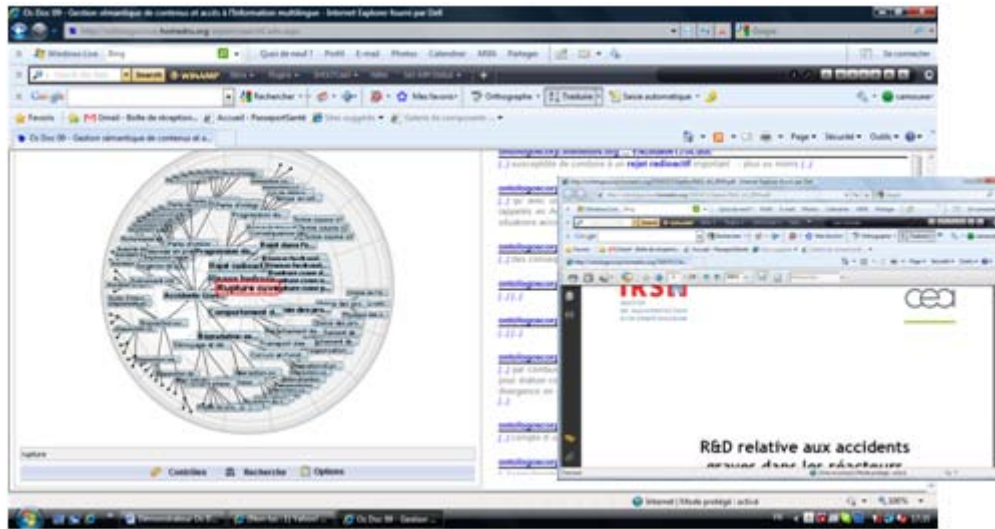


Fig. 4 – Plateforme de test « Accidents Graves ».

5. Conclusion

La construction de la connaissance expérientielle connaît de fortes mutations dans nos contextes de travail actuels. Sa formation, sa transmission se trouvent totalement modifiées : les traditions de compagnonnage historiques dans les entreprises, foncièrement informelles et intersubjectives sont tronquées et seules restent aujourd'hui les traces formelles et médiatées... Il manque donc une étape dans la formation de l'expérience de métier et il convient d'y apporter des solutions très opérationnelles puisque les besoins sont eux bien concrets. Nous abordons alors le problème comme un système de poupées russes dont la couche supérieure est la notion de métier. Elle englobe celle de connaissance collective, la *science* du domaine, diffusée dans les documents et objectivée par le langage. Les concepts étant au cœur du processus de formation des connaissances expérientielles, c'est de leurs manifestations en langue, les termes, qu'il faut partir. Concepts et mots des concepts sont donc à exploiter dans les formes médiatées des savoirs : la documentation technique. C'est à partir d'un corpus de textes que peut être construite une structure con-

Le métier : son savoir, son parler

ceptuelle qui ne pourra, en aucun cas, constituer une ontologie en elle-même, mais sera la base de travail de sa construction. Pour ce faire, le recours aux connaissances extralinguistiques s'impose. C'est pourquoi l'*expert* (celui qui détient l'*expérience*) doit être placé au cœur de la réflexion. Il a naturellement trouvé sa place comme pivot de notre méthode, dont nous décrivons ici chaque étape : extraction des lexiques, prise de distance pour construire progressivement les modélisations conceptuelles, passage vers les ontologies, puis retour aux mots pour définir la ou plutôt les terminologies de métier. En partant des documents et avec l'appui continu de l'expert, il est possible de reconstituer les modèles linguistiques et extralinguistiques du domaine et de substituer les maillons manquants dans la transmission des connaissances en faisant émerger le *sens métier*. La base de connaissances que nous avons construite sur le domaine restreint des « Accidents Graves » nucléaires, a trouvé une réalisation concrète dans une plateforme de test soumise aux usagers. L'étape suivante de nos travaux sera maintenant d'en amorcer l'évaluation.

Références

- Aristote, (1986). *Métaphysique*. Paris: Vrin
- Aussenac-Gilles, N., Condamines, A., (2000). « Entre textes et ontologies formelles : les bases de connaissances terminologiques ». *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*. Paris: Eyrolles
- Bachimont, B., (2000). « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances ». *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*. Paris: Eyrolles.
- Barberousse, A., (1999). *L'expérience*. Paris: Flammarion
- Boccon Gibod, H., (2006). *Application de méthodes et outils de Web sémantique pour la gouvernance d'un système d'information industriel*. EDF R&D
- Carnap, R., (1981). « Protocol Statements in the Formal Mode of Speech ». *Essential readings in Logical Positivism*. Oxford: Blackwell
- Djambian, C., (2010). *Valorisation d'un patrimoine documentaire industriel et évolution vers un système de gestion des connaissances orienté métiers*. Thèse de doctorat, Université Jean Moulin Lyon3.
- Gouadec, D., (1990). *Terminologie, Constitution des données*. Paris: Afnor
- Harris, Z. S., (1968). *Mathematical Structures of Language*. R.E. Krieger Publishing Company, reprint 1979
- Husserl., (1970). *Expérience et Jugement*. Paris: PUF
- Kant., (1997). *Critique de la raison pure*. Paris: Aubier

- Lainé-Cruzel, S., (2006). « Terminologie et intelligence artificielle ». *Encyclopédie de l'informatique et des Systèmes d'Information*. Paris: Vuibert
- Lamizet, B., (1992). *Les lieux de la communication*. Liège: Mardaga
- Locke, (1989). *Essai philosophique concernant l'entendement humain*. Paris: Vrin.
- Quine, (1980). « Les deux dogmes de l'empirisme ». *De Vienne à Cambridge*. Paris: Gallimard
- Rastier, F., (2004). « Ontologie(s) ». *Revue d'Intelligence Artificielle* 18
- Roche, C., (2005). « Terminologie et ontologie ». revue *Langages* 157
- Roche, C., (2007). « Dire n'est pas concevoir ». *Actes IC 2007*. Toulouse: Cépaduès éditions
- Wittgenstein., (1961). *Investigations philosophiques*. Paris: Gallimard

Summary

The paper focuses on the craft know-how, understood as a collective experiential knowledge, formed and formulated through the craft concepts. It now sees its transmission modes changing in companies, posing complex problems as in the case of the EDF Nuclear Engineering Division. The answer is in the formal manifestation of this knowledge, henceforth mediated: the language. This brings us naturally to the terminology, linguistic representation of a social field's *science*, which sets the language of this field. Ontologies that model and formally represent the notional system of terminologies are the logical next step.

We illustrate our discussion by presenting a practical application on a specific domain of the EDF Nuclear Engineering Division, building a knowledge database that includes a terminology and an ontology centered on the *craft meaning*.