Savoir technique en mouvement : les connaissances expérientielles face au numérique

Caroline Djambian

▶ **To cite this version:**

Caroline Djambian. Savoir technique en mouvement : les connaissances expérientielles face au numérique. 2014. hal-00978722v2

HAL Id: hal-00978722

https://hal.science/hal-00978722v2

Preprint submitted on 11 Dec 2014 (v2), last revised 24 Jul 2022 (v3)

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.

Le *sens collectif* au cœur de la valorisation de la documentation technique

Caroline Djambian
GRESEC (Groupe de REcherche Sur les Enjeux de la Communication)
caroline.djambian@upmf-grenoble.fr

RÉSUMÉ. A ce jour, les mutations signifiantes autour des technologies du numérique, observables tant dans les pratiques que dans la technique, s'inscrivent dans le « temps long ». Notre étude prendra comme base et témoin de ces mutations le document technique à travers le terrain empirique de l'ingénierie nucléaire. Nous constaterons que l'immatérialisation rapide et peu contrôlée du travail, mêlée à des évolutions contextuelles, a entièrement remis en cause le système info-communicationnel des organisations. De nouveaux rapports à l'information et à la communication des savoirs techniques émergent sous une double influence du numérique : la personnalisation des dispositifs documentaires par les utilisateurs et une médiation croissante des échanges. Dans un nouveau modèle communicationnel porté par le numérique, l'utilisateur devient acteur des pratiques de médiation. Les usages prennent désormais appui sur le principe de communautés de pratiques, faisant apparaître un autre type de compétence qui centralise l'information parcellaire et hétérogène, noyée dans la masse de données qui nous est aujourd'hui fournie. Notre but est ici de tenter d'accompagner l'ensemble de ces changements. Nous partons du fait qu'au sein des communautés de pratiques - qui ne sont en réalité que la continuité exacerbée par le numérique, des anciens modes communicationnels – le rapport à l'autre s'est déplacé dans l'échange médiaté, au rapport à la référence. Ces références, propres à un réseau, sont objectivées par les langues d'usages fixées dans les documents techniques. Nous proposons ici une méthode de valorisation de ces documents visant à en faire émerger le « sens » propre aux communautés d'un domaine technique, par la construction d'une terminologie et ontologie de domaine, rendues accessibles par une base de connaissances répondant aux nouvelles pratiques.

MOTS-CLÉS : document numérique, information scientifique et technique, pratiques informationnelles, médiation, culture collective, langues de spécialité, base de connaissance métier, terminologie, ontologie de domaine.

Introduction

A ce jour, des mutations signifiantes sont induites par une relation bilatérale entre l'évolution de l'information numérique et les nouvelles pratiques info-communicationnelles qu'elle implique. Cette évolution, qui s'est faite de manière exponentielle et plus ou moins contrôlée, demande d'ouvrir des réflexions sur l'accompagnement que nous pouvons apporter à ces bouleversements. Notre étude se base sur le cadre empirique de l'ingénierie nucléaire. Si nous présentons d'abord les spécificités de son information, les évolutions qu'elle connait sont elles, bien généralisables. Nous observerons progressivement les mutations de son fonctionnement documentaire, puis de l'ensemble des modes info-communicationnels et d'appropriation des savoirs de métier. Ainsi, les changements dus au numérique sont maintenant liés à l'émergence d'une nouvelle forme de compétence, que nous nommerons la *compétence unique*. L'usager devient alors acteur des pratiques de médiation, face à des connaissances maintenant représentées par les *documents media* porteurs, par le biais du langage, des références stabilisées des savoirs d'un domaine. Pour accompagner cette évolution il faut répondre aux nouvelles pratiques. C'est ce que nous tentons de réaliser en proposant une méthode de valorisation de la documentation technique pour en faire émerger le *sens collectif* aux communautés d'un domaine technique, par la construction d'une base de connaissances métier axée sur une terminologie et une ontologie dudit domaine.

1. Evolutions contextuelles

1.2. Spécificités de l'information et de la documentation technique traitées

Comme le rappelle Yves Gambier : *« à l'industrie traditionnelle consommatrice d'énergie et de main d'œuvre succède une industrie dont la matière première devient d'avantage le flux d'information »* (Gambier, 1991). Le document technique occupe aujourd'hui dans les structures de tous types, un rôle pivot pour leur activité et la pérennité de leurs savoirs, et dont l'importance est nettement accrue par le numérique. Nous pouvons commencer par rappeler quelques unes des caractéristiques principales de ce document technique, reprises par M. Holzem et J. Labiche dans l'ouvrage « Document et organisation » (Holzem, Labiche, 2004). Selon eux, sa lecture ne constitue qu'une étape et non une finalité qui est en fait l'action (Heurley, 2001 ; Fayol, 2002) ; le document technique suit également des règles rédactionnelles tacites à un contexte, un collectif donné, et visant à en orienter le comportement ; enfin, le document technique est la jonction entre le langage (le monde de l'information et de la communication) et le monde réel (Adam, 2001). Ces diverses caractéristiques émergeront au fil des observations que nous allons exposer dans cet article.

Le cadre empirique qui sert de fil conducteur à notre étude est celui de l'ingénierie nucléaire. Il possède comme tout contexte, des spécificités du point de vue de l'information et des documents qu'il traite. Qui peut nier que le nucléaire est sans nul doute un objet particulier ? La durée de vie des centrales durée de vie n'est en rien celle de produits classiques, aujourd'hui souvent éphémères. Or, les outils qui traitent l'information de cet artefact n'en suivent pas les temporalités. Leur existence n'est qu'un passage dans le cycle de conception et d'exploitation des centrales. Les systèmes informationnels doivent portant jouer le rôle de relais du savoir de l'ingénierie nucléaire et répondre, dans ces circonstances particulières, à des exigences réglementaires fortes, notamment liées aux enjeux de sécurité. *« Pour des installations industrielles dont les risques doivent être impérativement maîtrisés, la cohérence des représentations a toujours été une obligation réglementaire, autrement plus complexe à établir »* que pour des objets courants (Boccon-Gibod, 2006).

Or, il faut comprendre qu'une réalisation aussi conséquente qu'une centrale nucléaire s'appuie sur une production d'information très conséquente et jouant un rôle primordial. Les systèmes et processus d'information qui la régissent sont calqués sur les objets qu'ils traitent. Ainsi, la classification des documents suit l'organisation des centrales par paliers (répartition par puissance en MWe : 900, 1300, 1450, 1600 (EPR)). Les centrales sont elles-mêmes découpées en maillons dénommés systèmes élémentaires et reliés entre eux. La structure de la documentation suit essentiellement l'organisation de ces systèmes à travers les référentiels du parc nucléaire dont le plus important, le DSE (Dossier de Système Elémentaire), est lui-même la base de tout un réseau de documents. La documentation de l'ingénierie nucléaire représente dans notre cadre et à l'heure actuelle, un volume de 10 millions de documents numérisés (avec une grande hétérogénéité de types et formats), centralisés dans une GED (Gestion Electronique des Documents). Pour donner un ordre de grandeur de cette documentation de référence, un seul de ces documents, nommé Rapport de Sûreté Nucléaire, peut facilement atteindre le millier de pages.

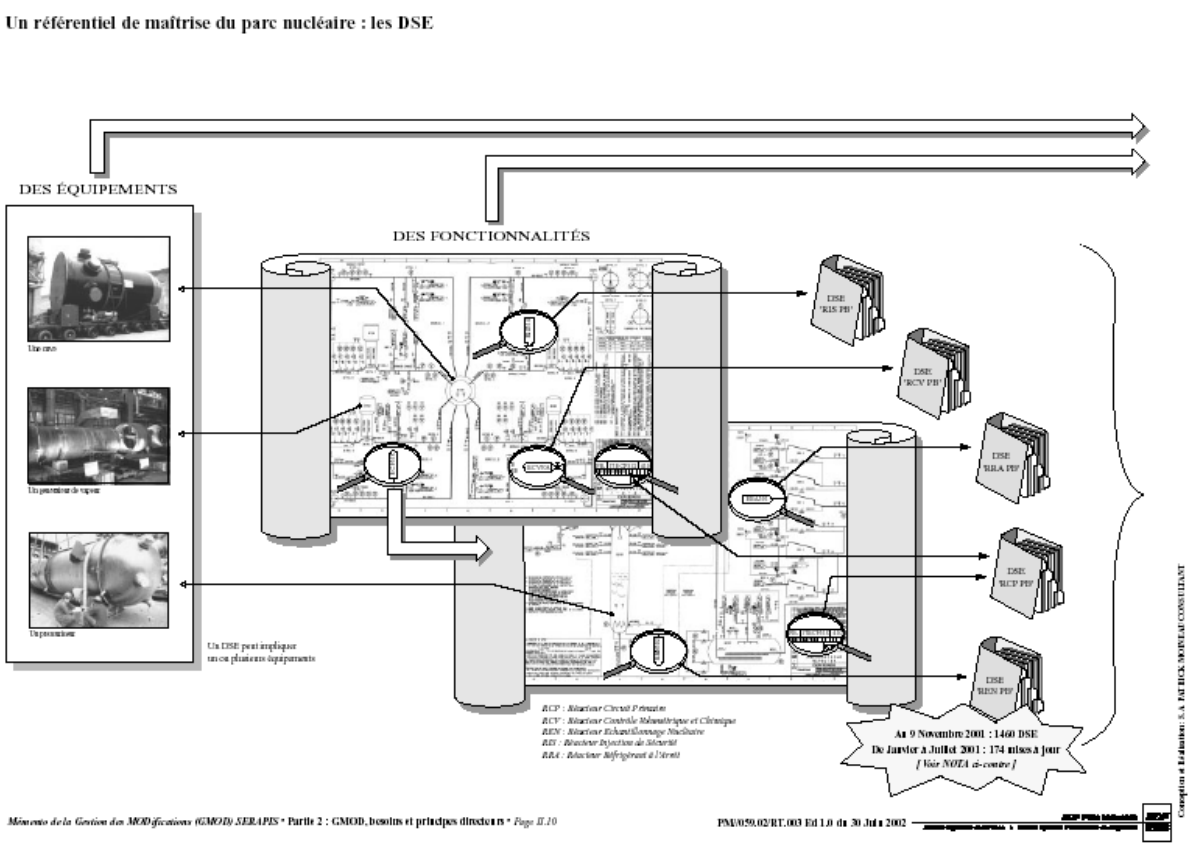

Fig. 1 - *Le DSE (Dossier de Système Elémentaire).*

Une centrale est un gigantesque réseau fonctionnel matérialisé par des schémas et décomposé en systèmes élémentaires. Le DSE (Dossier de Système Elémentaire) est un référentiel et correspond à une fonctionnalité de la centrale. Il est un maillon d'un ensemble de documents : les référentiels de maîtrise du parc nucléaire.

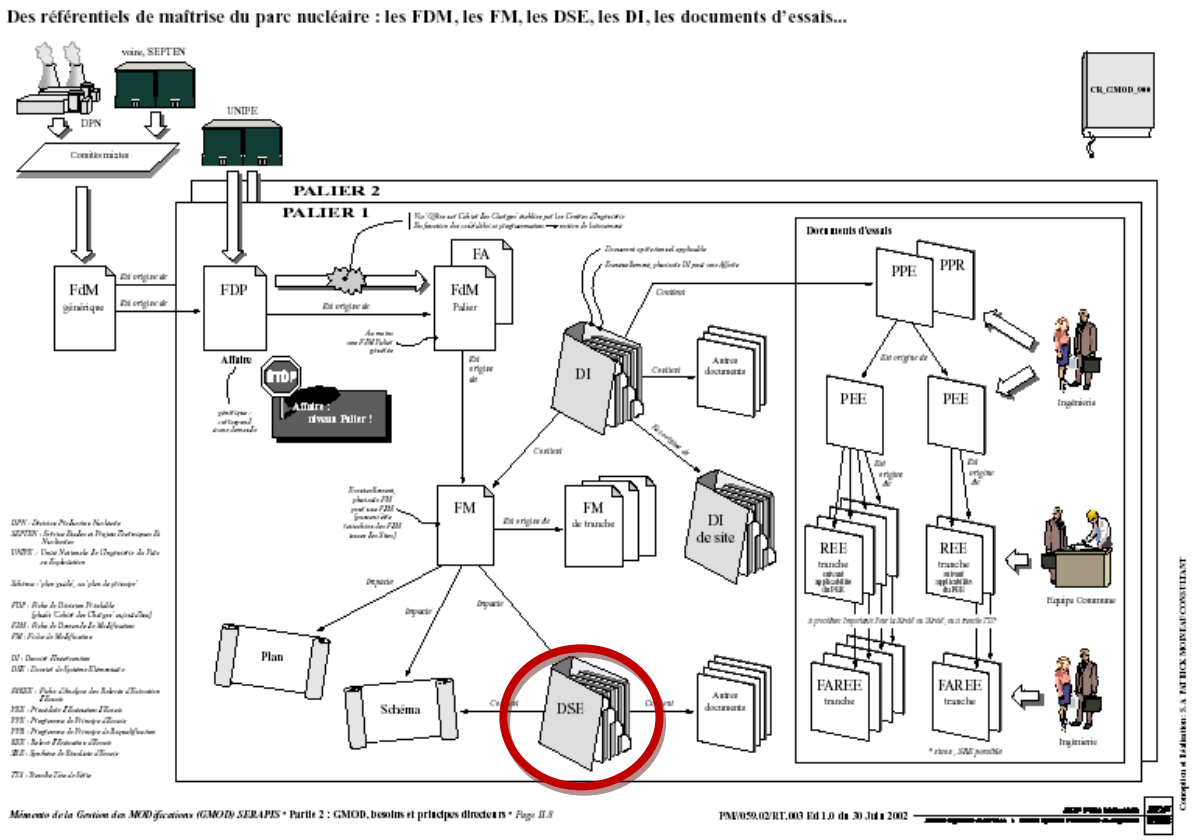

Fig. 2 - *Les référentiels de maîtrise du parc.*

Le DSE (Dossier de Système Elémentaire) parmi ces référentiels.

1.2. Mutations des fonctions informationnelles

La complexité de cette documentation s'explique par la multiplicité des métiers qui interviennent dans l'œuvre nucléaire et le croisement de leurs activités en constante évolution. Cette complexité et des contextes réglementaires lourds ont engendré des exigences et mécanismes imposés bien avant l'informatisation massive qu'a connue le champ industriel. Cet ensemble de paramètres a généré des processus documentaires rigoureux attachés à une logique matérielle. Par la suite, l'arrivée de l'immatérialisation du travail a facilité le suivi de ces exigences réglementaires. Elle s'est cependant réalisée dans une logique encore prégnante où tous les objets métiers reliant les activités de la vie d'une installation nucléaire étaient centralisés dans des documents papiers. La problématique de l'interopérabilité des systèmes n'a alors guère été une priorité, tout comme celle de leur appropriation par les utilisateurs. L'introduction de nouveaux dispositifs d'information a été faite, comme dans tant de structures, selon une logique purement technicienne, au détriment des aspects socio-organisationnels qu'elle a bouleversés.

Plus spécifiquement, le passage progressif de cette documentation au tout numérique a foncièrement modifié le cycle documentaire et surtout fait évoluer le rôle des acteurs dans les activités et pratiques informationnelles. Premièrement, la maîtrise documentaire, aujourd'hui régie par un processus qualité respectant les exigences que nous venons d'évoquer, a été décomposée en six sous-processus présentant des interactions entre eux et associés à des acteurs. Mais comme le souligne Bernard Miège, les *« changements de systèmes techniques »* sont souvent *« l'occasion »* ou l'alibi « *de la mise en œuvre de nouvelles formes de*

management u travail » (Miège, 2004). Autrefois centralisée par le service de documentation et animée par des professionnels de l'information, la maîtrise documentaire s'est donc trouvée entièrement décentralisée entre ces acteurs et place aujourd'hui l'ingénieur métier, non plus comme destinataire final de la chaîne documentaire, mais au centre même du processus comme producteur, indexeur et utilisateur de l'information. Ainsi, les départements techniques gèrent l'élaboration et l'évolution des listes des documents de doctrine pour chaque Collection dont ils ont la responsabilité et les secrétariats réalisent l'enregistrement des documents qu'ils émettent, suite à une première indexation réalisée par les ingénieurs eux-mêmes. Les documentalistes n'interviennent plus dans la maîtrise documentaire, que pour vérifier le respect du processus qualité.

D'autre part, mais également conséquence de ce premier point (puisque l'utilisateur a été rendu autonome malgré lui), l'informatisation rapide et peu contrôlée des postes de travail a fait fleurir une multitude d'applications détournant le Système d'Information officiel pour s'adapter progressivement aux mutations ambiantes. Venant d'initiatives éparpillées des utilisateurs eux-mêmes, elles se sont agrégées autour du système central (la GED) de manière connexe mais non reliée, en dehors de toute cohérence. On dénombrait en 2009, pour la seule ingénierie nucléaire, soit une structure de 3500 personnes, 400 bases de données de formats variés (Notes, Oracle, Access, Excel, Filemaker,…) et 162 intranets. Cette quantité impressionnante d'outils créés par les métiers eux-mêmes, répond en réalité à des mutations du rapport à l'information. Amorcées depuis les années 90 elles furent induites par un changement des modes de travail. Elles sont ainsi passées de pratiques informationnelles verticales, où la gestion de l'information suit une autorité descendante lourde et cloisonnée, encadrée par des obligations réglementaires, leurs procédures et mesures de traçabilité organisées autour d'une production documentaire papier très hiérarchisée ; à une pratique horizontale répondant aux exigences de concertation, de réactivité, de coopération entre acteurs et de recherche d'information plus que de documents. On évolue alors de modes informationnels diachroniques à un fonctionnement synchronique répondant à de nouveaux besoins de travail en temps réel (Pintea, 1995).

Ce glissement a nettement été influencé par le travail en « mode projet » (synchronique par excellence et qui s'oppose à une division traditionnelle des tâches par métiers, purement diachronique), ainsi que par les pratiques des nouvelles générations qui ont grandi avec l'outil informatique. Une telle modification des pratiques s'observe essentiellement dans des contextes d'expertise généralisée ou du moins, dans des contextes où les niveaux de formation et profils des utilisateurs sont assez homogènes (ce qui pose également la question d'une potentielle sélectivité sociale face à de nouveaux systèmes informationnels). Ainsi, on assiste par ces stratégies d'adaptation individuelles et collectives, bien au-delà d'un simple bricolage informationnel, à une véritable personnalisation par les utilisateurs des dispositifs documentaires.

2. Influence des technologies numériques entre individualisation et médiation des pratiques : une construction résolument sociale des savoirs

2.1. Un déplacement de la communication des savoirs porté par le numérique

En réalité, ces mutations ne concernent pas seulement les pratiques documentaires mais, de façon plus profonde, les modes de communication des savoirs. Lorsque l'ensemble du travail de conception des centrales était fixé dans une documentation papier régie par un système unique et centralisé, mais lourd et procédurier, ce système n'occupait qu'un rôle de stockage.

Paradoxalement, la diffusion de l'information et des savoirs entre ingénieurs était alors, proprement informelle et intersubjective, basée sur le réseau interne. Les ingénieurs ayant connu cette époque font perdurer ces usages : ils ne recherchent pas l'information dans la GED mais savent « qui travaille sur quoi » et font directement appel à la personne concernée dans leur réseau. La transmission des savoirs s'appuyait quant à elle, sur une tradition de compagnonnage où des « anciens » inculquaient le métier *in situ* au nouvel arrivant pendant un à deux ans. La problématique majeure pour cette structure comme pour d'autres, est d'être aujourd'hui confrontée à des départs en retraite massifs et mal anticipés de ces « anciens » et de leurs connaissances informelles. Les enjeux sont donc forts et laissent présager d'une déperdition conséquente de savoirs, elle-même renforcée par les politiques de mobilité interne courantes de nos jours.

S'il n'y a donc pas eu de croisement générationnel, les structures sont de plus confrontées à la dissonance qui s'est installée entre fonctionnements anciens et actuels. L'informatisation généralisée, croisée au départ des anciens et de leurs pratiques info-communicationnelles centrées sur l'intersubjectifs, ont provoqué de réels glissements. Le rapport à l'autre qui régissait les échanges de savoirs est maintenant reporté sur le rapport à la référence. L'information, dans un cadre de communication désormais médiatée, dépasse maintenant la relation intersubjective où les acteurs sont nettement identifiés l'un par l'autre. Elle rentre dans un champ où les acteurs de la communication sont abstraits. Par conséquent, outre cette concentration sur la référence, l'identité du sujet avec lequel on échange est aussi assimilable à *« une consistance sociale construite dans les rapports sociaux dont le savoir est la trace. »* (Lamizet, 1992). L'information se déploie dans le champ social qu'est l'institution, dans notre cas, une entreprise ou un domaine métier (groupe d'individus formant une compétence spécifique autour de références communes), et est le mode de production et de diffusion des savoirs portés par les sujets. L'information apparaît alors comme un champ de pratiques où vont se construire, par la communication, les références qui formeront le savoir, la culture collective d'un groupe donné. Dans ce contexte, la seule trace des savoirs accumulés dans l'entreprise est portée par la documentation technique. L'information diffuse ce savoir dans l'institution par le biais des documents *media* qui véhiculent de façon parcellaire et éparse, les références établies par la génération précédente et dont l'ensemble forme la culture collective de la communauté (ici, du métier). L'acquisition de la compétence sans « ancien » passe donc aujourd'hui par la maîtrise de ces références communes contenues dans les documents.

Or, la génération des nouveaux arrivants a, quant à elle, grandi avec l'informatisation. Elle a d'autant plus des pratiques d'appropriation du savoir diverses, que ses modes informationnels se sont créés avec le web. Les jeunes ingénieurs confrontés à leur nouvelle thématique de travail (dont ils n'ont acquis qu'une connaissance très partielle en école) recherchent ainsi spontanément l'information sur Google, Wikipedia et valident ces premières bribes d'information auprès de réseaux spécialisés sur internet, au lieu de se tourner vers la documentation interne qui ne leur parle tout simplement pas. Ils reproduisent en quelque sorte le fonctionnement par réseau des anciens, mais de manière médiatée. On peut aller jusqu'à affirmer que les pratiques reprennent des schémas identiques à l'avant numérique, mais exacerbés par ce dernier. Etant comme on l'a vu, contemporains de nouvelles organisations de travail orientées sur le transversal et l'interactif, la recherche de document brut ne répond en effet plus à leurs logiques et à leurs besoins : ils fonctionnent par notions. Leurs usages passent de modes de recherche documentaires, prépondérants avec la GED par exemple (recherche d'informations secondaires comme une référence de document, pour aboutir à l'information primaire), à un mode de recherche contextuel (en partant d'un mot ou groupe de mots du langage usuel pour aboutir à l'information) (Lefèvre, 2000). Nous assistons donc à une double influence des nouveaux dispositifs d'information par, à la fois

« l'individualisation des pratiques informationnelles et communicationnelles » inscrite dans l'évolution des « logiques sociales » que nous évoquions plus haut, et la *« tendance à ce que les échanges sociaux interpersonnels se voient adjoindre des échanges médiatisés à distance »* (Miège, 2004). Le constat qui s'impose est que le rôle de l'usager évolue résolument dans les pratiques informationnelles. Au sein d'un *« nouveau »* ou *« cinquième modèle de communication »* profondément porté par les nouvelles technologies numériques, il n'est plus observateur mais acteur des pratiques de médiation.

Cet acteur de la communication médiatée, *« consistance sociale »* qui se construit *« dans les rapports sociaux »,* évoque pour nous les travaux de Milgram, puis Guare et Watts. Sur la base de l'*expérience du petit monde* menée par Stanley Milgram (1967), puis de la *loi des Six degrés de séparation* (Guare, 2010 ; Watts, 2003), on observe effectivement que les nouveaux usages informationnels prennent appui sur les principes des communautés de pratique ou de métier (sur le modèle des réseaux sociaux). Elles sont animées par des échanges qui s'organisent autour de règles, références, terminologies et conceptions du monde communes, grâce à un consensus tacitement établi. Les échanges qui s'y opèrent, sont ainsi le moteur de la création des connaissances individuelles en passant par une appropriation des savoirs collectifs de la communauté. Les interactions entre membres du groupe renforcent des comportements sociaux stéréotypés (Harris, 1995) et les règles et normes instaurées dans le groupe orientent l'opinion des individus (Pettigrew, 1958). La façon dont la construction de l'espace social est vécue par la communauté semble être une condition de participation et d'apprentissage pour ses membres (Kreijns et al., 2004 ; Preece, 2000). Le groupe forme donc un espace social virtuel de communication au sein du champ plus large de l'institution. Appréhender ses règles, références, langues, implique de prendre part à l'espace d'échange et de formation de connaissances. Or, ces dernières ne sont aujourd'hui plus disponibles au travers d'une source unique (document ou personne), mais d'une mosaïque de données parcellaires, éparses et hétérogènes dont l'assemblement formera l'information. Ce processus de reconstitution et appropriation des savoirs engendre un nouveau type de compétence lié aux nouvelles pratiques. Cette compétence globalisante ou centralisatrice, nous la nommerons la *compétence unique* (Djambian & Agostinelli, 2013). On assiste donc bien, comme le soulignait B. Miège, à la construction de nouvelles normes d'action communicationnelle, incontestablement centrées sur une communication médiatée et profondément liées à de nouvelles formes de compétences (Miège, 2003).

2.2. Connaissances représentées sans sujet connaissant : le rôle des langages de l'information portés par la documentation technique

Comme nous venons de l'exposer, les références communes portées par la documentation de l'institution sont les marques stabilisées du savoir des experts aujourd'hui absents. Ce sont elles qui permettront de faire le lien entre ancien et nouveau système. Dans notre cadre comme dans de nombreux autres aujourd'hui, *« l'acquisition des connaissances ne se réalise pas à partir d'un expert ou de plusieurs, mais de la connaissance existante dans les bases de données, sorte de mémoire exosomatique, ... La connaissance que l'ingénieur a ici à traiter n'est pas de la connaissance en action à travers l'expertise des sujets connaissants, mais de la connaissance produite par eux et stockée dans les bases de données »* (Polanco, 1999). On parle désormais de *« connaissance représentée sans sujet connaissant »* (Popper, 1979), traduite en langages écrits. En ce que ces langages sont les témoins de leur champ social et la matérialisation des informations qui y circulent, la réappropriation des savoirs par les nouveaux arrivants implique, comme nous l'exposions plus haut, qu'ils les individualisent et

actualisent par leur maîtrise des références au réel. C'est sur ces références reconnues au sein du champ social de façon consensuelle, que s'organise l'énonciation et la réception de l'information, soit, les modes de transmission des savoirs et leur compréhension entre acteurs. Ces références échangées au sein d'un groupe doivent de plus, être légitimes pour les acteurs, sans quoi l'information elle-même ne sera pas valide. Elles sont objectivées dans notre cas, par le langage de l'ingénierie nucléaire - *« les langages de l'information vont occuper, au sein de la communication, la place d'une médiation généralisée »* (Lamizet, 1992). Nos travaux (Djambian, 2010) recoupent effectivement la thèse de Vygotski (1985) selon laquelle les connaissances sont socialement élaborées grâce au langage et aux autres systèmes de sens qui les représentent. Ainsi, par l'exploitation du langage il est possible de faire émerger le *sens métier* propre à une communauté technique et de pérenniser le savoir outre l'absence de ses détenteurs originels.

Ce langage est fixé dans les documents techniques qui jouent donc un rôle primordial dans la médiation et la représentation des connaissances collectives *« sans sujet connaissant »*. C'est ainsi que R. T. Pédauque (Pédauque, 2006) évoque le document tout d'abord comme *signe*, unité de sens faisant appel à la compréhension et composé des signifiants que sont les termes propres à une communauté, auxquels correspondent les connaissances. Il introduit également la notion de sociabilité et culture collective dans laquelle s'inscrit un document, qui n'a de raison et de légitimité que dans un contexte donné. Le document *medium* est donc la trace des relations sociales, alors que l'organisation est fondée sur la production et l'échange de documents. Ce sont les communautés de métier qui forment ce contexte en tant que réseau social et patrimoine commun. L'appropriation est alors à appréhender comme la somme de la *compréhension* (document signe), de la *sociabilité* (document medium) et de la *lisibilité* (document forme). A nous de construire ce troisième aspect qui englobe les signes vecteurs de connaissances, celle du document *forme*, composé d'une structure dans laquelle s'imbriquent des données et que nous voulons rendre lisible pour rendre les connaissances qu'il véhicule appropriables. Ce besoin est prégnant dans l'ingénierie nucléaire qui, comme tout domaine foncièrement technique, a une forte nécessité opérationnelle. Elle s'inscrit dans les secteurs scientifiques et techniques nécessitant une conceptualisation du monde et des dénominations, univoques, soit *« un moyen d'expression qui permette à la fois de prévenir les erreurs d'interprétation et d'empêcher les fautes de raisonnement »* (Frege, 1971). Garantir la lisibilité et donc l'appropriation réside ainsi dans le fait de valoriser, au-delà d'une langue de spécialité et de sa conceptualisation, un patrimoine d'information et de connaissances des domaines techniques de la structure, c'est-à-dire, de viser la création d'un véritable *« format d'échange »* (Roche, 2007).

3. Valorisation de la documentation technique vers l'émergence du « sens collectif » : proposition de méthode

3.1. Les trois strates media des connaissances : textes, termes, concepts

Si le langage représente et objective les savoirs, la terminologie en tant que science, est le reflet de la culture collective d'un domaine, en fixe le langage spécialisé et assure par des références stables, une communication fiable. Elle oriente l'interprétation qui doit être associée à chaque terme et définit le terme à utiliser pour désigner un concept ou un objet. 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 dans son champs social. Face à l'ambiguïté du langage courant, elle vise à

réduire les problématiques d'interprétations limitées à un domaine donné au sein duquel l'usage est stabilisé. Cependant, ces ambiguïtés persistent par l'évolution du langage dans le temps ou par la diversité des pratiques langagières entre communautés d'usages.

La création de terminologies dans le champ industriel ou scientifique, a pour but d'éliminer ces problèmes dans un domaine spécifique dont la délimitation rend plus aisée la construction d'un langage normé. Il s'agit donc de figer les usages et de normer, d'officialiser, un langage de spécialité de sorte à garantir au mieux la transmission et l'appropriation des références communes. Les termes qui constituent une terminologie, spécifiques à un métier, à une communauté, renvoient ainsi à une réalité extralinguistique partagée, qu'il s'agisse d'artefacts, de pratiques, méthodes, ou encore processus, qui sont les références communes à transmettre. Nous verrons donc que dans notre méthode, le travail terminologique est la première étape du travail ontologique. Nous présentons ici la création d'une base de connaissances métier d'un domaine circonscrit de l'ingénierie nucléaire, composée d'une terminologie et d'une ontologie centrées sur les références communes du domaine et qui forment le *sens métier* (Djambian, 2010). Cette base de connaissances est construite autour des trois strates media des connaissances : les textes, les termes, les concepts (Agosti et al., 1996).

3.2. Choix d'un corpus documentaire et langagier d'un domaine technique

Dans un domaine où la production documentaire est historiquement ancrée, les textes techniques peuvent être le socle stable pour construire une base de connaissances métier. Ces textes expriment essentiellement des langages d'opération. Ils font de la référence un outil opératoire visant l'obtention de résultats et de nouvelles références. Ce sont par essence, les langages de la constitution du savoir, sur lesquels nous pourront travailler. Comme dans tout projet basé sur les textes, la constitution du corpus occupe une importance majeure. Suite à une circonscription du domaine au sujet des « accidents graves » nucléaires, notre sélection fût accompagnée d'un expert du domaine, dans le souci de respecter l'application visée, la couverture maximale du domaine, l'homogénéité du corpus (du point de vue des sources et de l'usage de la langue et des concepts) et la légitimité accordée par les métiers. Après avoir choisi dans un premier temps un vingtaine de documents, l'expert s'est recentré sur 8 documents techniques qu'il estimait couvrir 90% des concepts métier, soit un corpus de 570 pages. Ce corpus recouvrait des documents stratégiques, notes de synthèses, référentiels métier ou plans de Recherche et Développement. Cependant, *« aucun corpus textuel n'explicite toutes les connaissances que sa lecture présuppose »* (Lerat, 2008). Il est donc improbable de construire la base de connaissances métier sans l'intervention d'un expert à chaque étape du travail. Le traitement des ressources spécialisées requiert des ressources externes (L'Homme, 2004, 2008). Plus centrée sur le travail humain que les méthodes sémasiologiques[1] de linguistique de corpus (Bourigault *et al.*, 2004), nos travaux empruntent ainsi aux méthodes onomasiologiques la relation à l'expert et la prise de distance aux mots, pour capter la *réalité extratextuelle partagée*.

[1] En linguistique, la sémasiologie est l'« *étude sémantique qui consiste à partir du signe linguistique* (dans notre cas, les termes) *pour aller vers la détermination du concept ».* Elle s'oppose à l'onomasiologie qui « *consiste, à partir du concept, à rechercher les signes linguistiques, l'expression qui lui correspond* » (Petit Larousse en couleur).

Il s'agit pour nous, dans cette méthode, d'exprimer la tension constante observable au sein du triptyque document-termes-concepts. Le travail terminologique n'est donc pour nous, que la part préliminaire incontournable et indissociable du travail ontologique. En effet, les travaux de terminologie ne s'intéressent trop souvent qu'aux mots comme finalité, sans aller jusqu'aux notions qu'ils désignent. Or, *« le lexique des langues ne reflète pas la conception scientifique du monde »* (Rastier, 2004). Ce sont les connaissances extralinguistiques qui vont permettre de lier ces deux plans et de comprendre les textes produits, diffusés, réappropriés au sein d'une communauté de métier. En effet, l'usage de figures de style y est courant pour exprimer les concepts du domaine : c'est le *parler métier*. Elles expriment de façon tacitement conventionnelle une conceptualisation qui est en fait l'ontologie du domaine. Les traces de ce parler 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. L'aide des experts est par conséquent indispensable pour exprimer les concepts réellement désignés par les tropes et évoluer progressivement vers une représentation conceptuelle.

Ainsi, 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. C'est l'usage qui va valider les termes au sein d'une communauté (Depecker et Roche, 2007), un domaine métier étant composé de plusieurs communautés apportant des variantes terminologiques associées à un même concept. Voici un exemple, dans le domaine des accidents graves qui nous intéresse, des diversités langagières des métiers au sein de ce même domaine. Pour nommer la notion représentant les processus qui interviennent dans le déroulement d'un accident grave (AG) nucléaire, les métiers appartenant à une première communauté de pratique parlent de *« physique des AG »* et les métiers d'une seconde communauté parlent de *« dynamique des AG »*, ce au sein de la même institution à deux étages de distance. Des objets linguistiques divers servent donc à nommer le même concept. Or ces langues d'usages sont encore différentes du langage normé utilisé dans la rédaction de documents officiels par exemple, où l'on parlera du *« déroulement d'un AG »*.

On imagine donc à quel point, la multiplicité de ces modèles linguistiques pose donc un réel problème dans la recherche d'information en langage naturel par exemple, ou dans l'appropriation des connaissances par les nouveaux arrivants. D'où la nécessité de recenser, clarifier et délimiter ce langage, (la notion de normalisation est d'ailleurs liée à celle de terminologie par l'origine même du mot, puisque le latin *« terminus »* désigne ce qui limite, ce qui borne), tout en gardant la richesse des langues d'usages des diverses communautés de pratique, en ne se limitant pas au niveau linguistique mais en remontant jusqu'à la strate extralinguistique, qui seule permet de prendre en compte cette diversité. Le langage dans toute sa richesse et son ambiguïté, cristallisation du savoir de chaque communauté d'un même domaine, n'est donc pas dissociable du concept, de la référence unique, 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 différents noms des concepts, forme « le pile et le face » d'un même travail.

3.3. Génération des lexiques et création des réseaux conceptuels

Notre méthode part des techniques de linguistique de corpus essentiellement portées par la société scientifique de TIA (Terminologie et Intelligence Artificielle), qui reviennent à des techniques sémasiologiques. A partir de notre corpus, une analyse linguistique à partir de patrons lexicaux-syntaxiques préétablis, a été réalisée avec l'outil LCW de l'Equipe Condillac (université de Savoie). Un premier lexique de 20 026 syntagmes nominaux a été généré, ramenés à 770 en éliminant les syntagmes de fréquence d'apparition inférieure ou égale à 10. Ce lexique a cependant soulevé un problème de taille, à savoir que les métiers nomment les concepts par de longues propositions qu'ils réduisent dans le langage courant sous la forme d'acronymes, comportant également parfois des chiffres ou caractères spéciaux. Un lexique complémentaire de 1710 acronymes a donc été généré à partir du même corpus, ne conservant que les valeurs numériques et noms propres. Les notions cœurs du domaine (projets, outils, matériels, phénomènes physiques…) étaient ainsi représentées dans les lexiques. Des retraitements manuels et de la validation d'experts ont résulté deux lexiques de 344 syntagmes nominaux et 333 acronymes.

L'ébauche d'un réseau lexical a été réalisée à partir de ces lexiques, à base de relations linguistiques (synonymie, d'hyperonymie, …), mais est rapidement devenu inextricable. Ainsi, *« il convient de se rappeler que tout travail terminologique devrait être fondé sur les notions et non sur les termes »* (Felber, 1984). Effectivement, face à la difficulté d'élaborer une structure lexicale directement à partir des mots du domaine, il convient de sortir de la langue et de se diriger directement vers le conceptuel. A cette étape, notre travail se dissocie des méthodes de linguistique de corpus pour se diriger vers des techniques onomasiologiques. Il est effectivement impératif de conserver tout au long du travail ce *« souci constant de s'extraire de tout discours pour se référer à un « socle » stable de connaissances »* (Roche, 2007), cette part extralinguistique, externe et complémentaire aux textes.

Nous avons alors souhaité suivre le raisonnement de la logique experte (logique du raisonnement exprimant les connaissances de type « si…alors ») par laquelle les ingénieurs communiquent spontanément leur savoir en organisant des concepts en séquences. En dessinant progressivement les grandes notions du domaine pour les identifier et les classer dans des modélisations simples autour du processus du déroulement d'un Accident Grave, soit d'un point de vue phénoménologique[2], nous avons obtenu des *Vues* réparties en *Ensembles* (Figure 3), composées de relations multiples suivant les séquences de la dynamique des Accidents Graves. Ce travail a été réalisé sur la base de l'étude systématique et approfondie des textes du corpus et des échanges répétés avec les métiers. Il est important dès cette étape, de se concentrer sur le niveau conceptuel en sortant de la langue des lexiques, en distinguant les termes d'usages des termes normés et des dénominations des concepts (pour reprendre notre exemple cité plus haut, les termes d'usages sont « physique des AG » ou « dynamique des AG », le terme normé « déroulement d'un AG » et nous avons choisi de dénommer le concept par « déroulement d'un accident grave » de sorte à que cette dénomination seule permette de situer le concept sans ambiguïté dans le réseau notionnel).

Lors de la validation par l'expert, la vision globale du déroulement d'un Accident Grave lui a immédiatement parlé et a permis de passer rapidement l'ensemble des concepts en revue. Ainsi, tous les concepts avaient été identifiés et clairement nommés. Il a également souhaité de conserver les dénominations de la modélisation, les jugeant plus significatives, ce qui implique que ces notions lui parlaient indépendamment de ses usages. La mouvance du

[2] Le but ici à été de dessiner l'ensemble des phénomènes intervenant dans un Accident Grave suivant leurs séquences (causes, effets, etc).

système linguistique a effectivement été démontrée à maintes reprises dans notre expérience, nous obligeant à constater comme de nombreux auteurs (Rastier, 1995 ; Slodzian, 1999 ; Condamines, 2003) que le système notionnel d'un domaine n'est pas représenté par une terminologie unique. La figure 3 présente l'un des Ensembles formant le réseau conceptuel validé par l'expert.

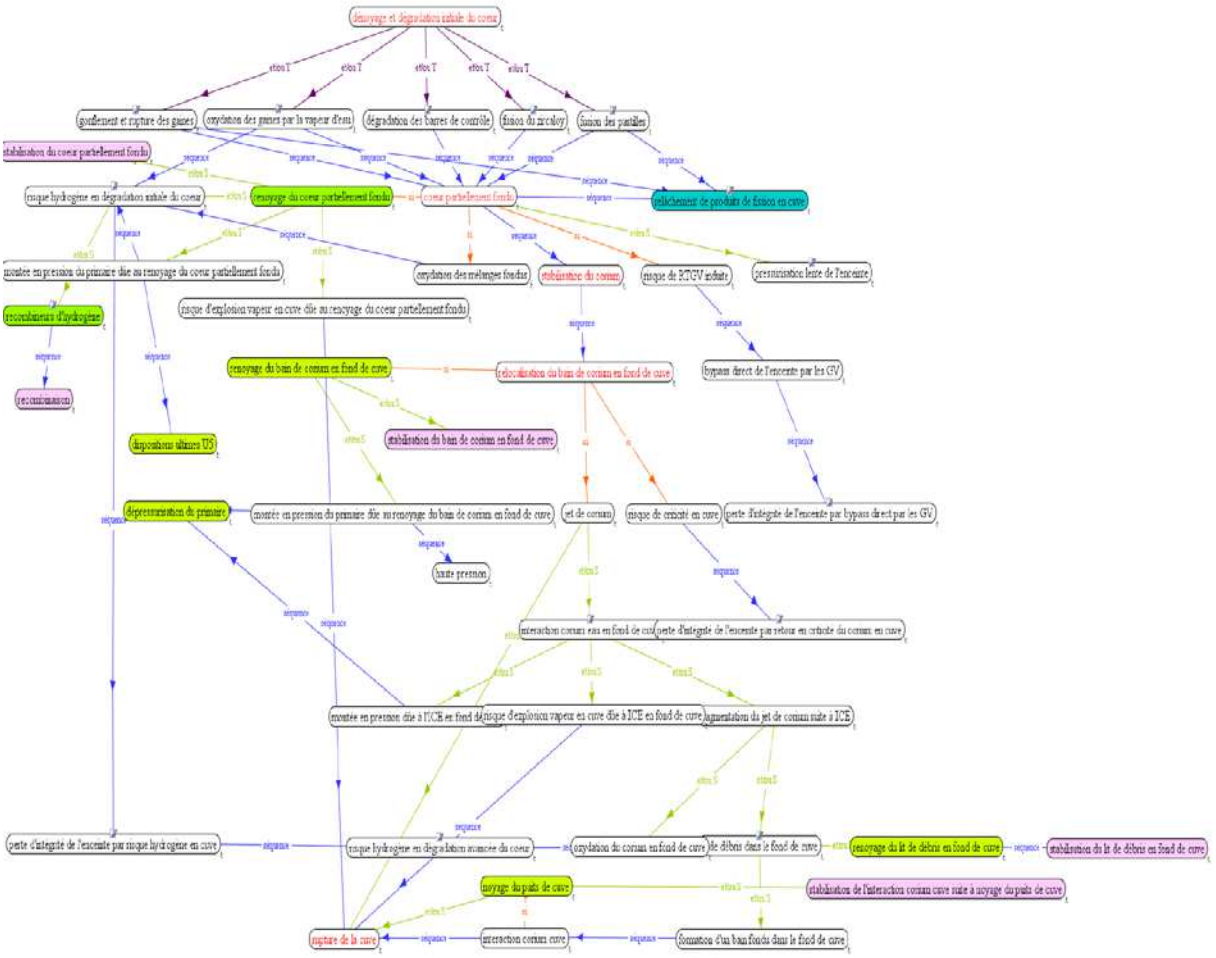

Fig. 3 – Ensemble « *progression* en cuve ».

3.4. Définition de l'ontologie et de la terminologie

La figure ci-dessus démontre la complexité de la réalité physique du déroulement d'un accident grave par la multiplicité des relations entre chaque concept. Or une ontologie est une représentation conceptuelle destinée à être comprise et exploitée par un ordinateur. Ce dernier ne peut en aucun cas exploiter des relations aussi complexes. Il faut donc transformer ce réseau notionnel en un réseau liant uniquement les concepts par relation de subsomption (de type générique/spécifique). Pour ce faire nous avons utilisé la méthode de différenciation spécifique, inspirée de l'arbre dichotomique de Porphyre (cf. *L'Isagogè* de Porphyre). Il répartit les concepts par relations « est une sorte de » ou « n'est pas une sorte de », comme présenté dans la figure suivante. Il faut signaler que, dans le cas d'un réseau sémantique construit directement à partir des lexiques, une étape intermédiaire serait ici nécessaire pour passer des termes du réseau (appartenant encore au(x) modèle(s) linguistique(s)) à un système

notionnel représentant les concepts du domaine. Mais, puisque nous étions déjà à cette étape au niveau conceptuel, nous avons directement pu clarifier les relations multiples et restructurer tous les concepts en ensembles sémantiquement liés par des relations générique/spécifique. Comme le rappelle Rastier, les *« principes de la construction ontologique restent invariables : il s'agit de subsumer la diversité des étants sous l'unité des concepts hiérarchiquement supérieurs »*. (Rastier, 2004). Cependant, il a parfois été nécessaire de créer des concepts génériques qui n'apparaissaient pas dans le réseau notionnel, comme par exemple, pour catégoriser certains phénomènes physiques. La figure ci-dessous montre l'un de ces concepts artificiels : « accident de perte ». Il permet, parmi les événements initiateurs d'un accident grave, de différencier ce type d'accident d'un accident de rupture ou de transitoires. Ces dénominations permettent de catégoriser et situer d'emblée le type d'accident dans l'ensemble du réseau. Ils n'apparaissent pas dans la vision finale de l'ontologie, notamment celle fournie à l'utilisateur. Après validation par l'expert, les onze sous-ontologies obtenues ont été converties au format OWL avec l'outil OCW de l'Equipe Condillac.

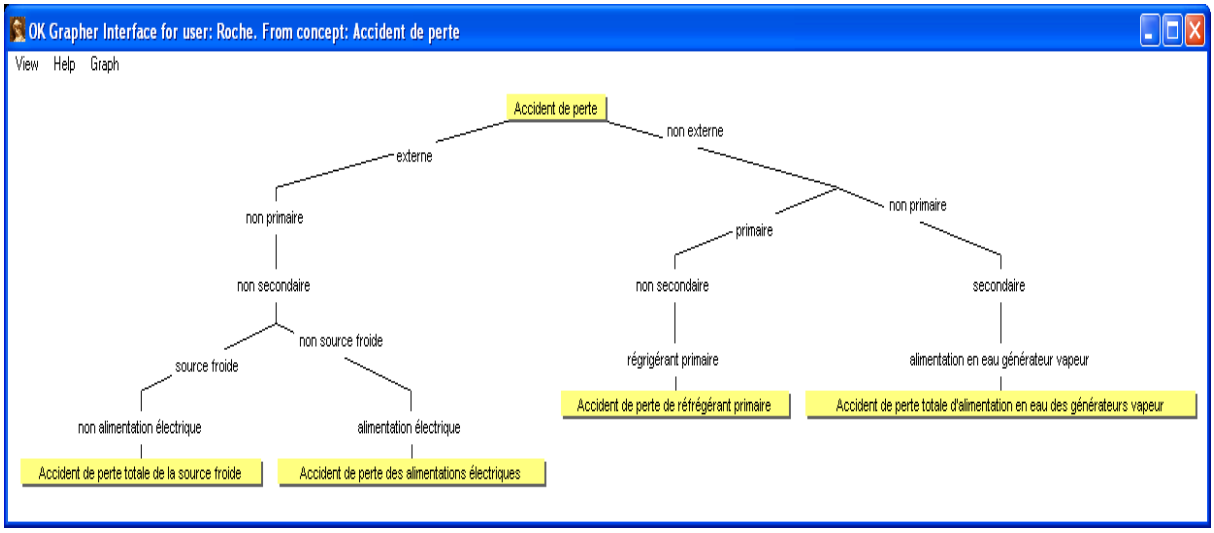

Fig. 4 – *Le principe de différenciation spécifique. Détail de la construction de la sous-ontologie « Evénements initiateurs » : les « Accidents de perte ».*

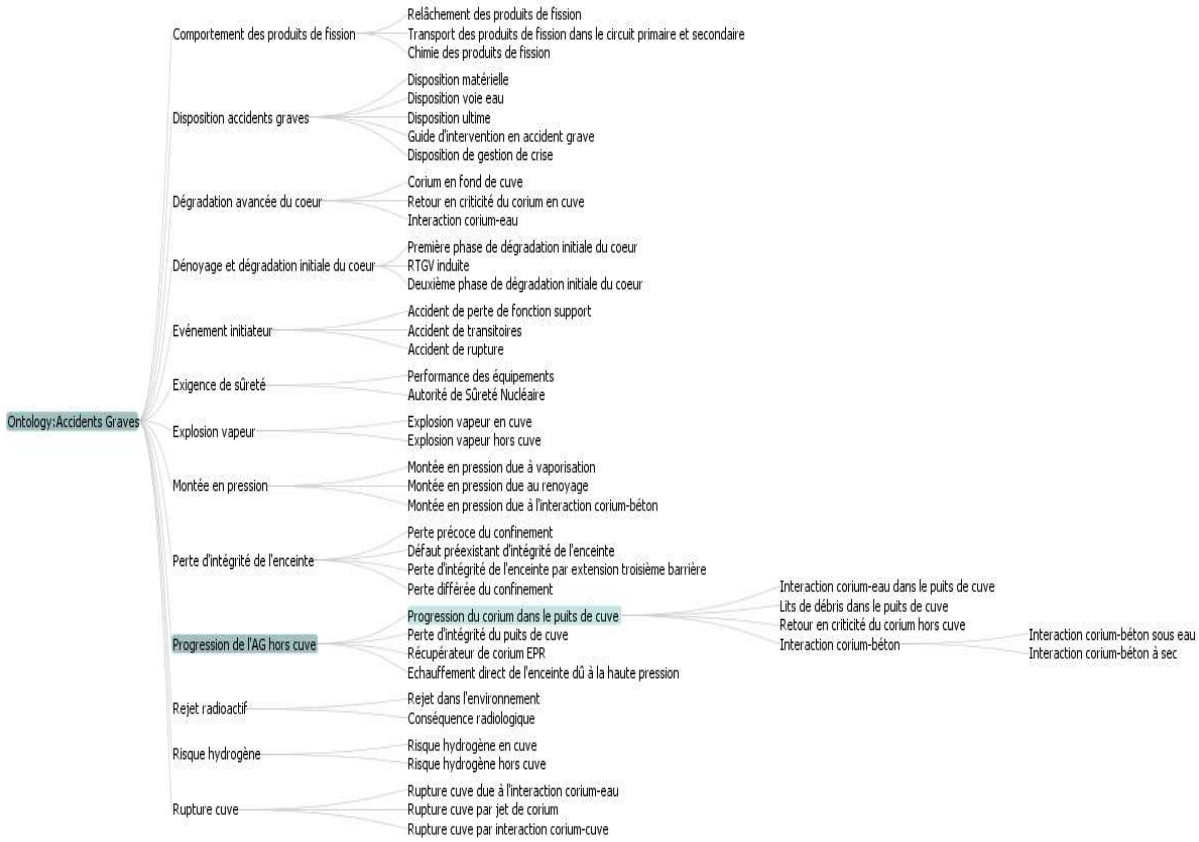

Fig. 5 – L'une des vues finales de l'ontologie sous forme de « Treeview ».

Mais *« ..., les ontologies sont des réseaux de concepts et n'ont donc pas de signifiants. Comment alors les articuler aux langues ? Ici intervient l'articulation de la synonymie et de l'hyperonymie : la première permet de réduire la diversité des signifiants (en fait des signes) ; la seconde permet de les subsumer sous des concepts. On passe ainsi des signes aux choses (par la référence) puis aux concepts (par la subsomption), ... »* (Rastier, 2004). Il s'agit ici de retourner aux mots des lexiques pour exploiter les relations d'hyperonymie et synonymie permettant d'articuler extralinguistique et linguistique. Cette étape est nécessaire pour faire le lien entre l'ontologie et les documents et permettre à terme la recherche et gestion (indexation automatique) dans le système d'information. La définition de la terminologie consiste à associer chaque terme des lexiques aux concepts de l'ontologie, identifier les synonymes, les acronymes et leurs développés, etc. Suite à une ultime validation par l'expert, la terminologie a été convertie au format XML, compatible avec la majorité des outils actuels de recherche et gestion de la documentation.

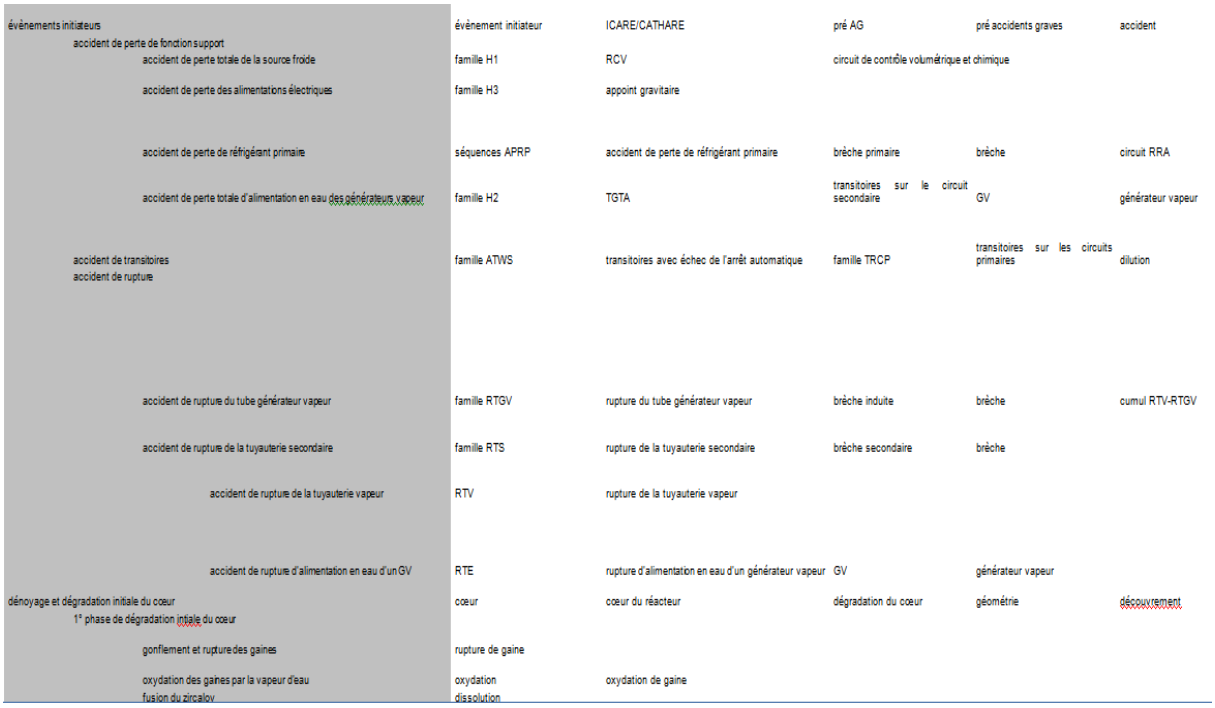

Fig. 5 – Définition de la terminologie « Accidents Graves ».

A l'issue, une plateforme de test a été mise à disposition des métiers. La Base de connaissances Accidents Graves (terminologie et ontologie) a ainsi été appliquée à la gestion et recherche de la documentation technique de référence du domaine et à la représentation de ses connaissances. Les documents sont automatiquement indexés sur l'ontologie en prenant en compte les variations d'usages des termes entre communautés de pratiques et permettant une recherche par concepts (accès à l'information plus adapté aux nouvelles pratiques informationnelles) ou en langage naturel avec expansion de requêtes. Par exemple, dans le cas de la requête « *EDE »,* l'outil signale une ambiguïté et propose les notions de *« mise en dépression de l'espace entre enceinte »* ou *« échauffement direct de l'enceinte »*. Dans ce second cas, l'outil propose également des documents traitant de *« DCH »* ou *« direct containment heating »,* équivalent anglais utilisé de façon indifférenciée dans la langue d'usage du métier. Ce système offre une vision globale de l'environnement informationnel et une synthèse de l'information contenue. Il permet de passer de la fourniture d'une information brute à une information implicite de niveau plus élevé (Polanco, 1999). Sont ainsi valorisés et pérennisés le patrimoine documentaire et les savoirs techniques qu'il recèle.

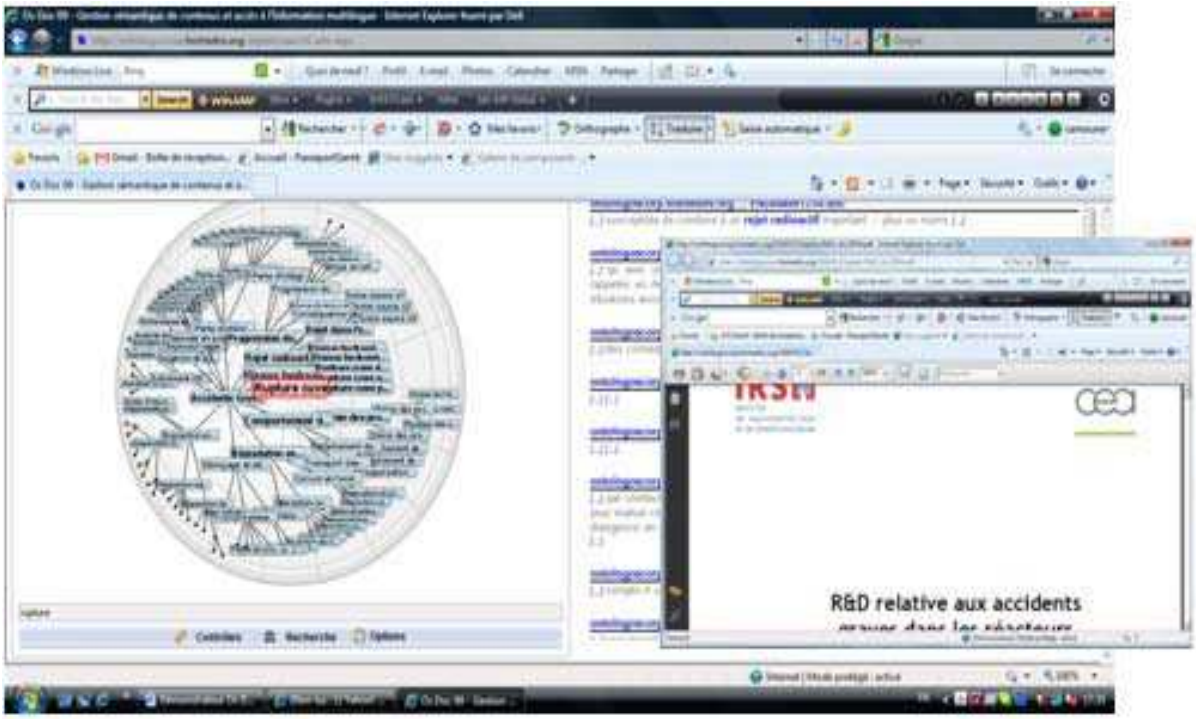

Fig. 6 – *Plateforme de test « Accidents Graves ».*

Conclusion

L'immatérialisation du travail s'est faite, sous couvert d'une apparente facilité, de manière rapide et peu contrôlée, selon des logiques techniciennes omettant les problématiques d'interopérabilité et d'appropriation. Les modifications dues aux technologies du numérique ont d'ailleurs servi d'alibi à des remaniements organisationnels plus profonds. On les retrouve à travers les modifications du management informationnel dans les structures, ayant souvent fait l'objet d'une décentralisation de la maîtrise documentaire. L'utilisateur rendu autonome malgré lui face aux Systèmes d'Information a mis en place des stratégies d'adaptation en créant ses propres outils, qui répondent à ses besoins (coopération, réactivité) et à son nouveau rapport à l'information. On assiste ainsi à une réelle personnalisation des dispositifs documentaires par les utilisateurs, qui deviennent acteurs d'échanges désormais médiatés. Cette double influence de l'immatérialisation aboutit à un nouveau modèle de communication porté par le numérique. Les nouveaux usages s'organisent en communautés de pratiques et la compétence de travail se veut à présent sociale et centralisatrice des d'information cachées dans la masse des données qui nous est accessible. Les modes de communication des savoirs techniques sont donc bouleversés. Le rapport à l'autre qui régissait les échanges de savoir s'est déplacé sur le rapport à la référence. Les documents techniques deviennent alors les média des références établies par la communauté (ex : les anciens experts) et qui forment la culture collective d'un domaine technique. Acquérir la compétence technique revient à maîtriser ces références. Ces dernières étant objectivées par les langues d'usage des communautés, elles sont à appréhender comme étant le point de départ pour la création de systèmes capables d'accompagner ces nouvelles pratiques et de faire le pont avec les anciennes. Au milieu de la profusion créée par le numérique la valorisation du *sens collectif* à ces communautés peut permettre de rétablir une communication fiable des savoirs.

Bibliographie

Adam, J-M. (2001). Types de textes ou genres de discours ? Comment classer les textes qui disent de et comment faire ?. *Langages*, 141, 10-27.

Agosti, M., F. Crestani, et M. Melucci (1996). Design and implementation of a tool for the automatic construction of hypertexts for information retrieval. *Information Processing & Management,* 32, 4:459-476.

Boccon-Gibod, H. (2006). *Application de méthodes et outils du Web sémantique pour la gouvernance d'un système d'information industriel*. Nantes: Semaine de la Connaissance.

Bourigault, D., N. Aussenac-Gilles, et Charlet J. (2004). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas. *Revue d'Intelligence Artificielle* X.

Condamines, A. (2003) *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*. Mémoire d'habilitation à diriger des recherches, Université Toulouse Le Mirail.

Depecker, L., Roche C. (2007). Entre idée et concept : vers l'ontologie. *Langages 168*, pp. 106-114.

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 Lyon 3.

Djambian, C., Agostinelli, S. (2013). De la métis au e-learning : la médiation du rapport au savoir. *Distances et médiations des savoirs.* A paraître.

Fayol, M. (2002). Les documents techniques : bilan et perspectives. *Psychologie Française*, 47-1, 9-18.

Felber, H. (1984). *Manuel de terminologie*. Paris: Unesco.

Frege, G. (1971). *Ecrits logiques et philosophiques*. Paris: Seuil.

Gambier, Y. (1991). Présupposés de la terminologie : vers une remise en cause. *Cahiers de Linguistique Sociale : Terminologie et sociolinguistique*, 18, 31-58.

Guare, J. (2010). *Six degrees of separation*. London: A&C Black.

Harris, J.R.(1995). "Where is children environment? A group socialization theory of development". *Psychological review*, 102, 458-489.

Heurley, L. (2001). Du langage à l'action : le fonctionnement des textes procéduraux. *Langages*, 141, 64-79.

Holzem, M., Labiche, J. (2004). Conclusion : Un essai de définition conceptuelle. In Holzem, Maryvonne & Labiche, Jacques (Eds). *Document et Organisation : forum document et organisation, semaine du document numérique, La Rochelle, Juin 2004.* Paris : Europia.

Kreijns, K., Kirschner, P. A., Jochems, W., & van Buuren, H. (2004). "Measuring perceived quality of social space in distributed learning groups". *Computers in human behavior*, 20, 607-632.

Lamizet, B., (1992). *Les lieux de la communication.* Liège: Mardaga.

Lefèvre, P. (2000). *La recherche d'informations, du texte intégral au thésaurus*. Paris : Hermès Science.

Lerat, P. (2008). Propositions pour un réseau conceptuel des instruments de mesure œnologiques. *Actes TOTh 2008*. Annecy: Institut Porphyre, 73-90.

L'Homme, M. C. (2004). *La terminologie : principes et technique.* Montréal: Les presses de l'Université de Montréal.

L'Homme, M. C. (2008). Ressources lexicales, terminologiques et ontologiques : une analyse comparative dans le domaine de l'informatique. *Revue française de linguistique appliquée* XIII-1, 87-118.

Miège, B. (2004). *L'information-communication, objet de connaissance*. Bruxelles: De Boeck

Miège, B. (2003). « La société conquise par la communication ? Oui, mais selon quelles modalités ? ». *Actes Pour une refondation des enseignements de communication des organisations 2003*. Paris : direction de l'Enseignement scolaire.

Milgram, S. (1967). The small world problem. *Psychology today*, 1(1), 60-67.

Pedauque, R. T. (2006). *Le document à la lumière du numérique.* Caen : C&F éditions.

Pettigrew, T.F. (1958). « Personnality and social cultural factors in intergroup attitudes: a crossnational comparison". *Journal of conflict resolution*, 2, 29-42.

Pintea, J. (1995). *Reengineering des systèmes documentaires*. Paris: Les éditions d'organisation.

Polanco, X. (1999). « Extraction et modélisation des connaissances : une approche et ses technologies » (EMCAT). In J., Maniez, & W., Mustafa El Hadi, (Eds.), *Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche d'information*. Lille: Travaux et recherches.

Popper, K. (1979). *Objective knowledge: an evolutionary approach.* Oxford: Oxford University Press.

Preece, J (2000). *Online Communities: Designing Usability, Supporting Sociability*. New York: John Wiley & Sons.

Rastier, F. (1995). Le terme : entre ontologie et linguistique. *La banque des mots* 7, 35-65.

Rastier, F. (2004). Ontologie(s). *Revue d'Intelligence Artificielle* 18, 1:15-40.

Roche, C., (2007). « Dire n'est pas concevoir ». *Actes IC 2007.* Toulouse: Cépaduès éditions.

Slodzian, M. (1999). Le sens retrouvé : avènement d'une terminologie textuelle. *Le sens en terminologie*. Lyon: Centre de Recherche en Terminologie et Traduction de l'Université Lyon 2.

Vygotski, L. (1985). *Pensée et langage*. Paris: Messidor / Ed. Sociales.

Watts, D. (2003). *Six degrees: the science of a connected Age*. New York: Norton.