



Construction d'une ontologie du domaine HSE

Sylvie Despres, Frederic Furst, Sylvie Szulman

► To cite this version:

Sylvie Despres, Frederic Furst, Sylvie Szulman. Construction d'une ontologie du domaine HSE. 18es Journées Francophones d'Ingénierie des Connaissances, Jul 2007, Grenoble, France. pp.133-144. hal-00509935

HAL Id: hal-00509935

<https://hal.science/hal-00509935>

Submitted on 17 Aug 2010

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.

Construction d'une ontologie du domaine HSE

Sylvie Després¹, Frédéric Fürst², Sylvie Szulman¹

¹ LIPN CNRS-UMR 7030, 99 avenue J.B. Clément, 93430 Villetaneuse
sd@math-info.univ-paris5.fr,
sylvie.szulman@lipn.univ-paris13.fr

² LARIA CNRS-FRE 2733, 33 rue Saint Leu, 80039 Amiens
frederic.furst@u-picardie.fr

Résumé : Dans cet article, nous présentons un travail de construction d'une ontologie du domaine de la réglementation Hygiène-Sécurité-Environnement (HSE) concernant les installations industrielles. Ce travail a été mené selon deux approches : l'utilisation de nomenclatures et d'interviews d'experts du domaine pour aider à la compréhension des notions du domaine et à leur conceptualisation et l'utilisation d'outils d'analyse de textes à travers la méthode Terminae, permettant de repérer les concepts dans les textes. La complémentarité des deux approches permet de construire une conceptualisation sémantiquement correcte et couvrant le corpus utilisé.

Mots-clés : Ontologies juridiques, Construction d'ontologie à partir de textes.

1 Introduction

Les ontologies trouvent depuis plusieurs années un champ d'application important dans le domaine juridique. Ces applications sont motivées par l'importance et l'augmentation régulière du volume des connaissances et textes juridiques, qui conduit de plus en plus les professionnels du droit à souhaiter bénéficier de systèmes performants d'aide à la gestion des textes de loi. De plus, la réalisation de ces applications est facilitée par le fait que les connaissances juridiques sont souvent organisées de façon très structurée, s'appuyant sur des définitions existantes et s'exprimant par des ensembles de règles (Lame, 2002).

Le travail présenté dans cet article concerne le développement d'une ontologie de la réglementation française du domaine Hygiène-Sécurité-Environnement (HSE) et est mené en collaboration avec l'entreprise Tennaxia¹, société de conseil et d'aide aux entreprises dans le domaine HSE. L'offre de Tennaxia, qui vise à permettre aux entreprises industrielles de garantir la conformité de leurs installations vis-à-vis de la loi,

¹Tennaxia - 6 rue Léonard de Vinci, 53001 Laval Cedex - <http://www.tennaxia.com/>

comprend à la fois l'assistance de consultants spécialistes de la réglementation HSE et un ensemble de logiciels de gestion de cette réglementation. Tennaxia a souhaité développer une ontologie couvrant les domaines réglementaires concernés par son activité pour permettre à la fois :

- une capitalisation des connaissances des consultants à des fins de préservation et de transmission ;
- une certaine normalisation des connaissances expertes de Tennaxia permettant éventuellement de normaliser les pratiques des experts et d'améliorer la cohérence et l'utilisabilité des applications Tennaxia ;
- une ressource de référence pour le travail des experts, en particulier le travail collaboratif de veille juridique et de gestion des textes réglementaires.

Cette ontologie doit couvrir, à terme, le domaine HSE et l'ensemble des textes réglementaires du domaine, mais, dans une première étape, expérimentale, il a été décidé de se restreindre à un corpus réduit de textes portant uniquement sur la protection de l'environnement et limité à 5 textes. En outre, il est vite apparu nécessaire pour exploiter ces textes de recourir à des interviews d'experts de la réglementation HSE et sur des documents annexes, en particulier des nomenclatures, ceci afin, d'une part, de mieux comprendre le sens des termes contenus dans les textes et, d'autre part, de tenir compte de la façon dont les experts conçoivent et manipulent les connaissances du domaine dans leurs activités.

La première partie de l'article présente le travail de conceptualisation du domaine mené à partir des nomenclatures et des interviews d'experts. Une deuxième partie est consacrée au travail de construction de l'ontologie à partir des textes, utilisant la méthode Terminae du LIPN.

2 L'exploitation des nomenclatures et des interviews

2.1 Les nomenclatures

Un des outils largement utilisé par les experts HSE pour rechercher des textes réglementaires est une nomenclature disponible en ligne sur le site de l'INERIS (Institut National de l'Environnement Industriel et des Risques). Le service gratuit AIDA² recense tous les textes réglementaires concernant les installations industrielles classées (IC), ces textes étant regroupés dans de grandes rubriques liées aux substances et aux types d'activités industrielles (voir le classement par substances et activités de la figure 1). De plus, cette nomenclature est organisée en rubriques et sous-rubriques sur 3 niveaux, ce qui constitue un embryon de structuration du domaine : un niveau *Substances/Activités*, un niveau détaillant les grands types de *Substances* et d'*Activités* et un dernier niveau où apparaissent les substances et les activités elles-mêmes (par exemple plomb, chlore, travail mécanique des métaux et alliages, affinage des fromages). Une

²http://aida.ineris.fr/sommaires_textes/sommaire_thematique/index.htm

classification transversale permet également de rechercher des textes par type de risque environnemental (voir le sommaire thématique de la figure 1).

Ces rubriques constituent une nomenclature générale permettant d'appréhender l'étendue du domaine et la façon dont les experts manipulent les textes. En effet, les entrées de la nomenclature (substances, activités, risques) constituent les concepts utilisés par les experts pour catégoriser les textes et structurer le domaine, et doivent donc apparaître comme des notions centrales dans l'ontologie. Les nomenclatures ont donc permis une première conceptualisation du domaine en identifiant ses principaux concepts. Cependant, ces nomenclatures ne font pas apparaître clairement et précisément les liens entre les concepts du domaine et la façon dont les experts naviguent dans la nomenclature pour retrouver les textes. Deux interviews de consultants de Tennaxia ont donc été réalisées, l'une centrée sur l'obtention d'explications et d'éclaircissements sur les textes, les nomenclatures et de manière générale les connaissances du domaine, l'autre constituée essentiellement par l'observation d'une recherche de textes menée par un consultant.

Sommaire Thématique	Classement par substances et activités
Généralités sur la réglementation des IC	1.Substances
Nomenclature des IC (consultation du fichier pdf)	1.0. Substances/ préparations
Risques industriels	1.1. Toxiques
Air	1.2. Comburantes
Eau	1.3. Explosible
Bruit	1.4. Inflammables
Déchets	1.5. Combustibles
Sites et sols pollués	1.6. Corrosives
Carrières	1.7. Radioactives
Substances radioactives - INB	1.8. Divers
Organismes génétiquement modifiés (OGM)	2. Activités
Effet de serre	2.1. Activités agricoles et animaux
Produits chimiques	2.2. Agro-alimentaire
Transport de marchandises dangereuses	2.3. Textiles, cuirs et peaux
Risques naturels	2.4. Bois, papier, carton, imprimerie
Protection de la nature	2.5. Matériaux, minerais et métaux
Autres textes généraux relatifs à l'environnement	2.6. Chimie, caoutchouc
Acteurs de l'environnement	2.7. Déchets
Produits antiparasitaires	2.9. Divers

FIG. 1: Structuration de la nomenclature HSE de l'INERIS, par risques à gauche et par substances et activités à droite.

2.2 Les interviews des experts

Les interviews des experts ont été dirigées par la modélisation car elles sont intervenues alors que la construction de l'ontologie avait déjà débutée, à partir des nomenclatures et des textes. Elles ont essentiellement servi à valider ou clarifier la structuration des connaissances métier (c'est-à-dire des connaissances liées aux installations et activités industrielles, et indépendantes des contraintes juridiques). Elles n'ont pas servi à

désambigüer des termes mais à introduire un niveau de ressources intermédiaires entre les ressources très techniques du corpus et des nomenclatures et les connaissances de sens commun. Elles ont conduit à modifier la structuration proposée pour l'ontologie et à introduire des concepts nouveaux permettant de factoriser des concepts issus des candidats-termes extraits des textes (voir section 3).

La première interview a permis de clarifier pour les intervenants universitaires les connaissances du domaine et leur structuration. Les explications de l'expert ont porté sur :

- les différents types de textes juridiques du domaine HSE et leur intérêt ou non pour l'activité de Tennaxia ;
- la structuration et le contenu des textes et l'analyse des textes par les experts ;
- les différents types d'installations concernées par ces textes, en particulier les Installations Classées pour la Protection de l'Environnement (ICPE) ;
- les différents types de produits et substances concernées par ces textes ;
- les différents acteurs juridiques : exploitants, institutions, ... ;
- la description générale de l'activité des consultants : veille juridique et analyse des nouveaux textes, audit ;
- la classification existante des textes sur le site de l'INERIS : classification par types de produits, par types d'activité mais aussi par types de risques ;
- la notion d'exigence propre à Tennaxia : une exigence est une portion de texte homogène portant sur un contexte industriel donné et sur un type d'obligation ou d'interdiction donné. Pour chaque nouveau texte qui paraît, un expert y repère les exigences et les annoté à l'aide de catégories de la nomenclature.

La deuxième interview a comporté, d'une part, l'observation d'un expert dans l'activité de veille juridique consistant à analyser un nouveau texte pour en identifier les exigences et à l'intégrer dans le système d'annotation et de classification de Tennaxia et, d'autre part, une présentation générale de l'activité d'audit d'une installation. Cette interview a permis d'identifier les points d'entrée des consultants lors de la recherche d'exigences concernant une installation : les substances utilisées dans l'installation tout d'abord, ensuite le type d'activité industrielle menée dans l'installation (production, transformation, ...) et finalement les types de risque environnemental et de pollution liés à ces substances et cette activité.

2.3 Conceptualisation à partir des nomenclatures et des interviews

Les nomenclatures et les interviews ont permis de dégager les concepts centraux du domaine et les liens conceptuels qu'ils entretiennent. Le premier constat tiré de l'étude du travail des experts est que les connaissances mobilisées dans le cadre de leur activité sont pour la plus grande part des connaissances métier, liées aux installations et activités industrielles, et de façon moins importante des connaissances juridiques concernant la

nature des textes réglementaires et la façon dont ils s'appliquent. L'ontologie construite est donc essentiellement une ontologie métier et non une ontologie juridique à proprement parler.

Mais les structurations des concepts proposées dans les nomenclatures ne sont pas fondées sémantiquement et il a été nécessaire de définir les liens hiérarchiques entre concepts. Pour cela, nous nous sommes appuyés sur l'ontologie fondationnelle DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) (Masolo *et al.*, 2003), qui comporte des concepts abstraits ayant vocation à généraliser l'ensemble des concepts que l'on peut rencontrer dans les différents domaines de connaissances. DOLCE offre une conceptualisation philosophiquement fondée et dotée d'une axiomatisation riche, qui a été utilisée comme ressource pour clarifier la nature conceptuelle des objets du domaine, et comme structure d'accueil pour les concepts du domaine, ce qui a facilité leur positionnement les uns par rapport aux autres.

L'utilisation des nomenclatures et interviews a donc servi essentiellement à définir la structure générale de l'ontologie en mettant en évidence les concepts centraux du domaine et les liens entre eux et en offrant une vue générale du domaine facilitant la conceptualisation. La construction de l'ontologie a cependant été basée sur les textes du corpus fourni par Tennaxia.

3 Construction de l'ontologie à partir des textes

Comme il a déjà été dit, nous cherchons à créer une ontologie permettant de capitaliser le savoir-faire des experts qui à partir des caractéristiques d'une entreprise permet de retrouver la législation qu'elle doit respecter pour être en conformité avec la loi. Nous utilisons la méthode semi-automatique de construction d'ontologies à partir de textes Terminae, enrichie par la réutilisation de ressources termino-ontologiques (RTO) existantes (Després & Szulman, 2005). Cette méthode de construction d'ontologie à partir de textes s'applique plus particulièrement à un domaine constitué de plusieurs points de vue tels que le point de vue juridique et le point de vue métier. A partir des candidats termes extraits par un extracteur de termes, nous identifions les termes associés aux différents points de vue en s'aidant des différentes RTO disponibles. Ces termes peuvent appartenir à une ou plusieurs RTO. Dans ce dernier cas, ces termes sont polysémiques, et donnent naissance à plusieurs concepts dans l'ontologie. L'association d'un terme à un ou des concepts est exprimée dans une fiche terminologique.

Une fiche terminologique contient pour chaque terme les occurrences de ce terme dans le corpus. À partir des occurrences du terme dans le corpus, nous modélisons chaque sens identifié, par un concept. Le sens d'un terme est décrit par une définition en langue naturelle établie à partir des caractéristiques de l'utilisation du terme dans le corpus. Chaque définition est ensuite traduite par un concept dans le langage de description de l'ontologie. Cette traduction peut se faire à l'aide de la méthode OntoSpec (Kassel, 2002). Cette insertion se fait par un rattachement à un concept déjà existant dans l'ontologie. Un tel concept peut être un concept déjà construit par le processus précédent ou un concept existant dans les *core-ontologies* ou *top-ontologies*³. Dans

³Une *top-ontologie* ou ontologie de haut-niveau regroupe des notions très générales telles que les concepts

ce dernier cas, le concept de rattachement est appelé **concept d’ancrage** (Breuker & Winckels, 2003).

3.1 Description du corpus

Le corpus est constitué de cinq textes relatifs à l’environnement, sélectionnés par les experts du domaine et choisis pour leur diversité et leur représentativité relativement à l’ensemble des textes réglementaires HSE disponibles :

- le Décret n° 77- 1133 du 21 septembre 1977 pris pour l’application de la loi n° 76-663 du 19 juillet 1976 relative aux Installations Classées pour la Protection de l’Environnement
- le Décret n° 99-374 du 12 mai 1999 relatif à la mise sur le marché des piles et accumulateurs et à leur élimination
- l’Arrêté du 29 juin 2004 relatif au bilan de fonctionnement prévu par le décret n° 77-1133 du 21 septembre 1977 modifié
- l’Arrêté type - Rubrique n° 2910 : Combustion
- le Décret n° 2005-635 du 30 mai 2005 relatif au contrôle des circuits de traitement des déchets

Ces textes réglementaires concernent le domaine de la protection de l’environnement et contiennent des termes relevant à la fois des métiers, de la protection de l’environnement et de la protection des employés. Chaque texte comporte un titre et est organisé en articles. Le premier article donne la limite du texte et le dernier donne le moment de l’application de l’arrêté ou du décret.

3.2 Outil d’extraction de termes

L’outil d’extraction YaTeA (Aubin & Hamon, 2006) (Yet another Term extrActor) identifie et extrait dans un corpus des groupes nominaux pouvant être des termes candidats. Un terme candidat est un syntagme nominal susceptible d’être un terme du domaine considéré. Chaque terme candidat est analysé syntaxiquement pour faire apparaître sa structure sous la forme de têtes et de modifieurs. Les ressources linguistiques nécessaires à l’identification et l’analyse des termes candidats sont fournies pour le français et l’anglais. Ces ressources peuvent être modifiées par l’utilisateur. De nouvelles ressources peuvent également être créées pour une langue ou une langue de spécialité. Plusieurs formats de sortie sont proposés parmi lesquels :

- une liste de termes candidats à valider (interface de validation en cours de développement)

de temps, d’espace, d’action etc. Une *core-ontologie* ou ontologie noyau regroupe des notions générales sur un domaine, plus spécifiques que les notions des top-ontologies mais néanmoins indépendantes des applications, par exemple ici les notions de substance chimique, d’installation industrielle ou d’activité.

- une liste de termes candidats au format XML, compatible avec le format d'entrée de SynoTerm T. Hamon & Gros (1998) qui est un logiciel de détection de relations de synonymie.

La figure 2 présente un extrait de la liste des termes candidats fournie par YaTea.

```
<TERM_CANDIDATE>
  <ID>terme4692</ID>
  <FORM>élimination de déchets</FORM>
  <LEMMA>élimination de déchet</LEMMA>
  <MORPHOSYNTACTIC_FEATURES>
    <SYNTACTIC_CATEGORY>NOM de NOM</SYNTACTIC_CATEGORY>
  </MORPHOSYNTACTIC_FEATURES>
  <HEAD>terme267</HEAD>
  <NUMBER_OCCURRENCES>1</NUMBER_OCCURRENCES>
  <LIST_OCCURRENCES>
    <OCCURRENCE>
      <ID>occl</ID>
      <MNP>1</MNP>
      <DOC>1</DOC>
      <SENTENCE>232</SENTENCE>
      <START_POSITION>27</START_POSITION>
      <END_POSITION>49</END_POSITION>
    </OCCURRENCE>
  </LIST_OCCURRENCES>
  <TERM_CONFIDENCE>1</TERM_CONFIDENCE>
  <LOG_INFORMATION>extracteur</LOG_INFORMATION>
  <SYNTACTIC_ANALYSIS>
    <HEAD>
      terme267
    </HEAD>
    <MODIFIER POSITION="AFTER">
      terme268
    </MODIFIER>
  </SYNTACTIC_ANALYSIS>
</TERM_CANDIDATE>
```

FIG. 2: Extrait de la liste des termes candidats construite par YaTeA.

3.3 Mise en oeuvre de la méthode

Nous avons identifié trois types de concepts : métier, environnement et juridique. Or nous travaillons essentiellement à partir de textes, ce qui nous conduit à rechercher dans les textes, les termes métiers, environnements et juridiques afin de définir les concepts associés. Nous appliquons les différentes phases de la méthode précédemment décrites :

- identifier les termes juridiques en utilisant la RTO du droit français ;
- identifier les termes métiers en utilisant la nomenclature et les informations recueillies lors de l'interview d'un expert ;
- des fiches terminologiques sont alors construites pour les termes jugés pertinents pour l'application ;

- en suivant la méthode Terminae, un modèle du concept décrivant les relations qui le lie à d'autres concepts, est élaboré pour chaque signification du terme. Chaque concept est ensuite inséré dans l'ontologie suivant le principe décrit précédemment en recherchant le concept de rattachement.

L'intersection entre la liste des termes candidats extraits du corpus par YATEA, et la RTO du droit français a permis d'identifier les termes qui relèvent du domaine juridique. Ces termes appartiennent à un ensemble de codes dont le code civil à partir desquels cette RTO a été construite. Le contexte du travail a conduit à retenir les termes communs aux codes du travail et de l'environnement. Parmi les termes simples extraits par YATEA on compte 702 termes appartenant au code du travail et 643 termes appartenant au code de l'environnement, dont 575 communs aux deux tel que le terme *bruit*. Nous avons procédé de manière identique pour définir les termes métiers à partir de la nomenclature.

Pour illustrer notre méthode, nous considérons le terme *bruit*. En effet, il appartient au code du travail, au code de l'environnement, à la nomenclature. La figure 3 présente la fiche terminologique de *bruit*. Trois concepts terminologiques (BRUITM, BRUITJ, BRUITE) sont construits à partir du terme *bruit*. Du point de vue de l'environnement, le bruit est *sorte de risques et pollutions* qui a une mesure avec un seuil. Il est lié à un terme métier *émergence* qui nécessite la définition de deux sous-concepts (BRUIT AMBIANT, BRUIT RÉSIDUEL) du concept BRUIT. Le terme *pollution* est représenté dans l'ontologie par deux concepts POLLUTIONE et POLLUTIONJ. Dans l'ontologie, POLLUTIONE a pour père le concept PROCESSUS_AVEC_PATIENT_NON_HUMAIN qui est ancré dans Dolce sur le concept DOLCE-LITE_PROCESS.

Du point de vue juridique, le bruit est une pollution qui est un risque. Un risque est un dommage qui demande des réparations et conduit au concept de RESPONSABILITÉ présent dans la core ontologie CLO (core legal ontology). Ce terme *bruit* constitue un exemple de l'articulation possible entre les objets métiers et les exigences juridiques.

Une première ontologie a été créée en utilisant les interviews des experts et la nomenclature. Nous avons cherché à retrouver dans les textes les termes et leurs contextes, correspondants aux concepts de l'ontologie. Nous avons ainsi créé 57 fiches terminologiques donnant pour chaque terme dénotant un concept de l'ontologie, les occurrences de ce terme dans le corpus.

4 L'ontologie du domaine HSE

À partir des textes, des interviews et des nomenclatures, la structure générale de l'ontologie a pu être fixée et enrichie des concepts apparaissant dans le corpus. L'ontologie a été développée sous Protégé, un éditeur d'ontologie open-source développé à Stanford (Gennari *et al.*, 2003). Les 4 principaux concepts apparaissant dans l'ontologie sont les objets d'activité (entités produites ou consommées par une activité), les activités, les installations et équipements et les processus dommageables.

Entre ces concepts existent des relations qui expriment la sémantique du domaine. Les relations portent des cardinalités qui expriment le nombre potentiel de relations pouvant exister entre instances des concepts liés (ainsi, une activité donnée produit au

Fiche terminologique : bruit

Fichier Terme Concept Traçabilité

Date création 6 octobre 2006
Auteur sd_ss

Terme: bruit

Validation:
☒ En cours
☐ Terminé

Informations lexicales

Rubrique	valeur
forme	bruits
catégorie syntaxique	NOM

Concepts

bruitE
bruitI
bruitM

Définition LN

ou genante, tout phénomène acoustique produisant cette sensation, tout son ayant un caractère aléatoire qui n'a pas de composante définie

bruitM: Le bruit a une mesure. Le niveau de bruit a une valeur limite.
Un bruit a plusieurs origines possibles (aérien ou solidaire).
Peut avoir des conséquences sur la santé la sécurité du voisinage.
bruitI: liés aux commodités de voisinage (contrôle des nuisances acoustiques) et exposition aux risques dus au bruit

Liste des occurrences

pour la période de nuit, sauf si le bruit résiduel (hors fonctionnement de l'installation) dépasse ces limites.

2: Une mesure du niveau de **bruit** et de l'émergence doit être effectuée au moins tous les trois ans.

3: - émergence : la différence entre les niveaux de pression continus équivalents pondérés A, notés LAeq, T, du **bruit** ambiant (installation en fonctionnement) et du bruit résiduel (installation à l'arrêt) ;

4: - émergence : la différence entre les niveaux de pression continus équivalents pondérés A, notés LAeq, T, du **bruit** ambiant (installation en fonctionnement) et du bruit résiduel (installation à l'arrêt) ;

5: 8.1 - Valeurs limites de **bruit**.

Synonyme

Voir aussi

FIG. 3: Fiche terminologique du concept de bruit.

minimum un objet d'activité mais peut ne pas rejeter d'objet d'activité, c'est-à-dire de déchet). Ces relations peuvent être présentées schématiquement ainsi :

- un objet d'activité, une installation ou un équipement peuvent provoquer un processus dommageable ;
- une activité produit, utilise et/ou rejette un ou plusieurs objets d'activité ;
- une installation réalise une activité et un équipement participe à cette réalisation. D'autre part, un équipement est rattaché à une installation. Dans les équipements, on distingue les équipements de sécurité qui sont ceux qui participent à la prévention ou à la lutte contre un processus dommageable.

Une grande partie des substances apparaissant dans les textes sont des substances susceptibles de provoquer différents types de dommages aux installations, à l'environnement ou aux personnes. Ces substances font l'objet d'une classification en fonction des risques qu'elles présentent. Bien entendu, ces substances peuvent être considérées du point de vue de leurs propriétés chimiques générales, ou des éléments qui les composent, ou encore des activités où elles sont utilisées. Le point de vue adopté ici est celui du domaine HSE qui considère essentiellement une substance suivant les risques qu'elle présente (cf. Fig. 4). Cela n'interdit évidemment pas de lier une substance à une activité, mais ce point de vue contraint la structuration des concepts de substances.

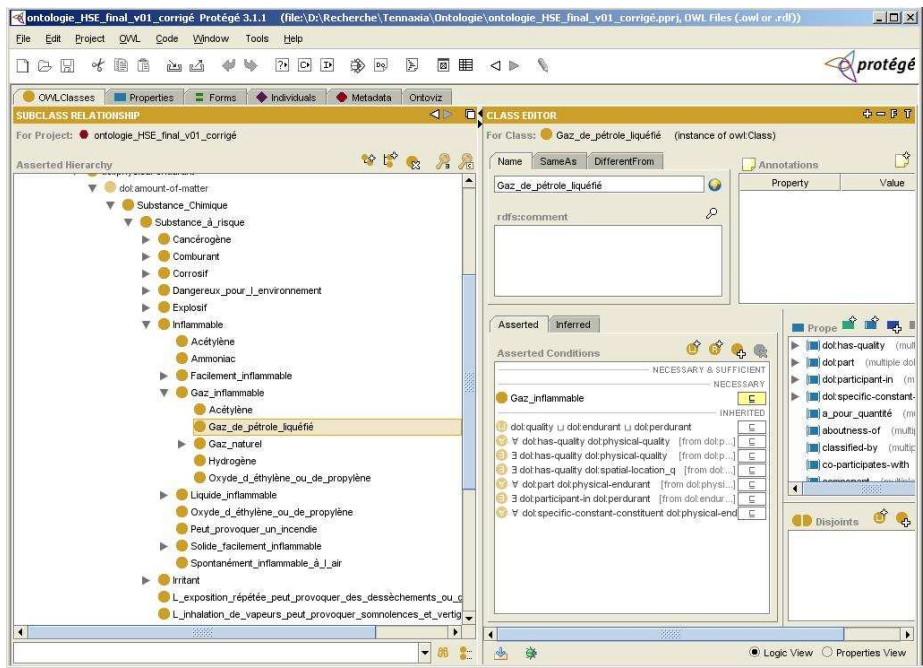


FIG. 4: Extraits de la hiérarchie des substances à risque.

Les risques présentés par les substances étant des concepts en eux-mêmes, ils sont donc modélisés en tant que tels et considérés comme des *Processus Dommageables*. À la hiérarchie des substances, organisée par types de processus dommageables, correspond donc une hiérarchie de processus dommageables. Bien entendu, il existe des processus dommageables qui n'apparaissent pas dans la nomenclature des substances dangereuses, mais font partie du domaine HSE, par exemple les nuisances sonores, ce qui justifie d'autant plus le choix de modéliser les processus dommageables comme des concepts propres.

Les substances dangereuses sont subsumées par la notion d'*Objet d'Activité*, introduite pour regrouper toutes les entités intervenant dans les activités industrielles comme ressources, produits ou déchets. Ces objets d'activité sont produits, rejetés ou utilisés par les activités, en fonction des types d'activité. L'analyse des textes, des interviews et des nomenclatures a permis de dresser la typologie suivante des activités concernées par la réglementation HSE :

- *production* : une activité de production a pour but de fabriquer un type d'objets (y compris des substances) en utilisant d'autres objets ou substances et en rejetant éventuellement des objets qui sont des déchets. La production inclut la transformation d'objets (conditionnement, traitement).
- *logistique* : une activité logistique a pour but de modifier la localisation d'un objet ou d'un ensemble d'objets.

- *contrôle* : une activité de contrôle a pour but de produire des mesures et des rapports sur une autre activité.
- *élimination* : une activité d'élimination a pour but de détruire un objet ou d'assurer son stockage définitif.
- *valorisation* : une activité de valorisation a pour but de transformer un déchet en un objet pouvant servir à une activité.
- *commerce* : une activité de commerce modifie le propriétaire d'un objet.

Ces grands types d'activités sont spécialisés par les activités citées dans les textes, qui peuvent préciser le type des objets concernés par l'activité (par exemple production de vin) et/ou préciser le type d'activité (par exemple stockage par immersion). Les activités se déroulent dans des *installations*, qui sont le véritable objet de la réglementation. Les installations et équipements industriels et agricoles sont de façon générale tous les objets physiques qui concourent à la réalisation d'une activité. Une installation est l'ensemble des objets participant à une activité donnée. À l'intérieur d'une installation, différents équipements concourent à la réalisation de l'activité, chaque équipement ne pouvant pas réaliser l'activité à lui tout seul. Par exemple, dans une installation de production de produits chimiques, une cuve pourra jouer un rôle mais ne suffira pas à elle seule à assurer la production. Les installations sont structurées selon les produits qu'elles utilisent ou produisent et selon les activités.

5 Conclusion

Plusieurs enseignements peuvent être tirés de ce travail de construction d'une ontologie du domaine HSE. Tout d'abord, la ressource de base du domaine, les textes juridiques, ne sont pas suffisants pour appréhender correctement les connaissances du domaine. L'apport de nomenclatures, ébauches de structuration de ces connaissances et surtout le recours à des experts permettent de bien comprendre les notions importantes et les liens qu'elles entretiennent. Mais seuls les textes et leur analyse permettent d'avoir accès à tous les concepts du domaine et de développer une ontologie qui permette de décrire complètement les textes HSE.

Cette expérience montre donc la complémentarité des deux approches de construction d'ontologie, l'une s'appuyant principalement sur les interviews d'experts et une nomenclature et l'autre sur l'étude des textes et la réutilisation de RTO. Cette complémentarité ne garantit cependant pas que l'ontologie modélise correctement et complètement les connaissances du domaine, et une validation de l'ontologie par les experts, et par l'expérience, reste nécessaire. La validation de l'ontologie construite par les experts de Tennaxia a d'ailleurs permis de corriger certains points de la conceptualisation.

Références

AUBIN S. & HAMON T. (2006). Improving term extraction with terminological resources. In S. P. T. P. TAPIO SALAKOSKI, FILIP GINTER, Ed., *Advances in Natural*

- Language Processing (5th International Conference on NLP, FinTAL 2006)*, number 4139 in LNAI, p. 380–387: Springer.
- BREUKER B. & WINCKELS R. (2003). Use and reuse of legal ontologies in knowledge engineering and information management. In *ICAIL 2003 Workshop on Legal Ontologies & Web based Legal Information Management*.
- DESPRÉS S. & SZULMAN S. (2005). Construction d'une ontologie du droit communautaire. In *IC'2005*, p. 85–96.
- GENNARI J., MUSEN M., FERGERSON R., GROSSO W., CRUBEZY M., ERIKSSON H., NOY N. & TU S. (2003). The evolution of protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, **58**, 89–123.
- KASSEL G. (2002). Une méthode de spécification semi-informelle d'ontologies. In *Actes des 13ièmes journées francophones d'Ingénierie des Connaissances (IC 2002)*, p. 75–87.
- LAME G. (2002). *Construction d'ontologie à partir de textes. Une ontologie du droit dédiée à la recherche d'information sur le Web*. Thèse de doctorat, École des Mines de Paris.
- MASOLO C., BORGO S., GANGEMI A., GUARINO N., OLTRAMARI A. & SCHNEIDER L. (2003). *The WonderWeb Library of Foundational Ontologies and the DOLCE ontology*. final report vr. 1.0, 31-12-2003, WonderWeb Deliverable D18.
- T. HAMON A. N. & GROS C. (1998). A step towards the detection of semantic variants of terms in technical documents. In *Proc. of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*, p. 498–504: Morgan Kaufmann.