



Réseau terminologique versus Ontologie

Sylvie Després, Sylvie Szulman

► To cite this version:

Sylvie Després, Sylvie Szulman. Réseau terminologique versus Ontologie. Toth 2008, Jun 2008, France. pp.17-34. hal-00423525

HAL Id: hal-00423525

<https://hal.science/hal-00423525>

Submitted on 11 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Réseau terminologique *versus* Ontologie

Sylvie Despres - Sylvie Szulman

LIPN – Université Paris 13

99 av J.B. Clément

93430 Villetaneuse

{prenom.nom}@lipn.univ-paris13.fr

http://www-lipn.univ-paris13.fr/~{nom}

Résumé : Les méthodes de construction d'ontologies à partir de textes comportent une phase de conceptualisation au cours de laquelle s'effectue le passage du terme au concept. Dans ce papier, nous montrons comment se construit ce passage du terme au concept *via* des étapes se situant sur le plan linguistique, termino-ontologique et ontologique permettant ainsi l'articulation entre l'expression des connaissances en corpus *via* la langue et leur expression formelle *via* un langage de représentation des connaissances. La spécificité des ressources obtenues à chaque pas de la conceptualisation permet de distinguer clairement leurs usages et leur complémentarité.

Mots-clés : réseau terminologique, ontologie, conceptualisation

1. Introduction

Les méthodes de construction d'ontologies à partir de textes comportent une phase de conceptualisation [Cimiano, 2006; Aussenac et al., 2008] au cours de laquelle s'effectue le passage du terme au concept.

Cette phase de conceptualisation nécessite des traitements se situant à la fois aux plans linguistique et ontologique. Pour faciliter la présentation, nous distinguons trois grandes étapes : l'étude linguistique du corpus qui aboutit à la construction d'un réseau terminologique, l'étape termino-ontologique qui permet de construire un réseau termino-ontologique et l'étape ontologique dont la finalité est l'élaboration de l'ontologie.

Réseau terminologique *versus* Ontologie

L'étude linguistique permet d'organiser les unités linguistiques *via* un ensemble de réseaux terminologiques partiels qui traduisent la structure lexicale des textes. Chacun de ces réseaux terminologiques est centré sur l'étude d'un terme représentatif du domaine. Cette étude est réalisée à partir des résultats des traitements obtenus par des outils de traitement automatique des langues (TAL).

L'exploitation des résultats issus de l'étude linguistique en vue de la construction de l'ontologie est pour une grande partie réalisée manuellement.

Un réseau termino-ontologique est construit à partir de l'interprétation, dans un cadre applicatif, des unités linguistiques (termes et relations lexicales les liant) constituant le réseau terminologique. Il est constitué d'unités dites termino-ontologiques (concepts terminologiques et relations sémantiques les liant).

Enfin les concepts de l'ontologie et les relations conceptuelles les associant sont construits à partir des unités termino-ontologiques figurant dans les réseaux termino-ontologiques. Ces concepts ontologiques sont décrits dans un langage formel, organisés dans une structure hiérarchique, liés par des relations conceptuelles et contraints par des règles et des axiomes.

L'articulation entre ces deux plans linguistique et ontologique constitue une tâche difficile puisqu'il s'agit de mettre en correspondance d'une part, des unités linguistiques caractérisées par leur polysémie, la diversité de leur emploi et d'autre part des concepts qui sont univoques et normés par le contexte de l'application. Par conséquent, il ne peut y avoir d'isomorphisme entre les deux structures. Le réseau termino-ontologique permet la transition entre les plans linguistique et ontologique.

Cette contribution insiste sur le fait, que la phase de conceptualisation ne relève pas de l'extraction de connaissances, mais consiste en une construction assurant le passage du terme au concept et permettant l'articulation entre l'expression des connaissances en corpus *via* la langue et leur expression formelle *via* un langage de représentation des connaissances.

Après avoir précisé le sens des notions utilisées, les différences entre les notions de réseau terminologique et ontologie sont clarifiées. Enfin, le passage du plan linguistique au plan ontologique réalisé dans la phase de conceptualisation est explicité. Un exemple dans le domaine juridique est utilisé pour illustrer notre propos. Enfin, l'intérêt de distinguer ces différents plans est analysé.

2. La phase de conceptualisation

Dans un schéma faisant maintenant référence (cf. figure 1), Cimiano [2006] fait apparaître les entités (termes, synonymes, concepts, hiérarchies de concepts, relations entre concepts autres que hiérarchiques, les axiomes et les règles relatifs aux concepts représentés) nécessaires à la construction d'une ontologie à partir de textes.

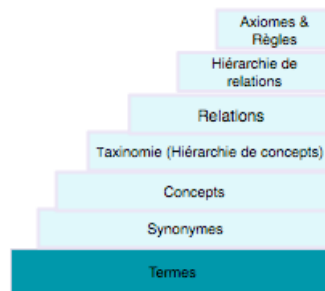


Fig. 1 : « Ontology learning layer cake » d'après Cimiano

A la lecture de ce schéma, il apparaît clairement que la nature des éléments utiles pour construire l'ontologie ne se situent pas sur le même plan. Construire l'ontologie revient à tirer parti de l'ensemble des résultats obtenus à partir des traitements réalisés sur les termes et les relations lexicales qui les lient dans une phase dite de conceptualisation.

La phase de conceptualisation est celle de la construction des concepts ontologiques à partir des résultats issus de l'analyse automatique des textes.

Réseau terminologique *versus* Ontologie

Elle est structurée en quatre étapes : - extraction des entités linguistiques et construction d'un réseau terminologique ; - élaboration de concepts termino-ontologiques et d'un réseau terminologique partiel ; - élaboration des concepts ontologiques et d'un réseau ontologique.

Ces différentes étapes permettent la transition du plan du discours au plan ontologique. Ce passage est réalisé en réalisant des traitements qui relèvent de la terminologie puis de la modélisation des connaissances et enfin de la représentation des connaissances. Nous avons figuré les différents plans dans lequel ce processus intervient dans le schéma extrait de [Mondary et *al.*, 2008] figurant figure 2.

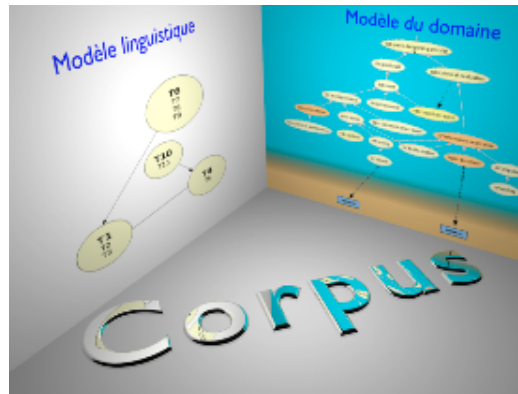


Fig. 2 : Articulation des plans textuel, linguistique et conceptuel

Passer des textes à l'ontologie, c'est donc passer du discours pris comme réalisation linguistique au modèle ontologique en passant par un plan termino-ontologique exprimé *via* un modèle conceptuel du domaine. Le modèle linguistique joue un rôle charnière entre le texte et le modèle conceptuel tandis que le modèle conceptuel assure le même rôle entre le modèle linguistique et le modèle ontologique. L'articulation de ces différents modèles requiert la construction de différents types de ressources qui sont caractérisées par la finalité du plan où elles sont situées.

3. Les notions utilisées

Afin de fixer l'usage des notions utilisées dans ce papier, nous rappelons les définitions que nous leur assignons et nous les illustrons sur un exemple extrait d'un travail dont la finalité était la construction d'une ontologie du domaine pour faire de la recherche d'information [Després *et al.*, 2007].

Les définitions proposées dans ce paragraphe sont associées aux notions qui interviennent dans les grandes trois étapes du processus de conceptualisation. Elles sont certainement discutables, mais elles nous permettent toutefois de mieux dégager la nature des notions sur lesquelles nous travaillons.

3.1. Une présentation succincte situant l'exemple

Le domaine évoqué est celui de l'organisation de la gestion de l'hygiène, la sécurité et l'environnement (HSE) des entreprises. Le corpus de référence est constitué d'un ensemble de textes réglementaires concernant les installations industrielles classées. Les connaissances du domaine font référence à trois contextes distincts : - le contexte métier, chaque installation relève d'un métier (chaudronnerie, peinture, ...) ; - le contexte environnement, chaque entreprise a un impact différent sur l'environnement (nuisances au niveau du bruit, de l'air, ...) ; - le contexte juridique, le cadre juridique à appliquer dépend du niveau de dangerosité de l'entreprise (risque SEVESO). Les notions servant à l'exemple sont celles de bruit et installation.

L'ontologie à construire devait aider à capitaliser les savoir-faire des experts chargés de réaliser les audits auprès de leurs entreprises clientes. Dans ce contexte, le rôle de ces experts est de déterminer, à partir des caractéristiques d'une entreprise, les éléments qui lui permettent d'être en conformité avec la législation.

3.2. Les notions intervenant dans l'étude linguistique

L'étude linguistique permet d'établir une analyse terminologique du texte guidée par l'objectif de construction d'un modèle du domaine étudié.

Réseau terminologique *versus* Ontologie

Définitions

Un **candidat terme** est un mot ou une séquence de mots susceptibles d'être retenus comme terme par un terminologue [Bourigault et *al.*, 2000] ou par un analyste et de fournir les étiquettes des concepts [Bourigault et *al.*, 2003].

Une **relation lexicale** est une relation entre termes (hyponymie, hyperonymie, holonymie, méronymie, synonymie, antonymie, causalité, etc.) [Cruse, 1986].

Un **réseau terminologique** est un ensemble de candidats termes reliés entre eux par des relations lexicales [Bourigault et *al.*, 2005].

Exemple

La figure 3 présente un réseau terminologique partiel établi à partir des termes *bruit* et *installation* qui désignent des notions centrales du domaine de la gestion HSE.

Dans l'extrait de texte sélectionné dans la figure 1, les candidats termes retenus sont ceux relatifs à la notion de *bruit* {bruit ambiant, bruit résiduel, bruit aérien, bruit solidien, niveau de bruit, niveau de bruit ambiant, niveau de bruit global} et à la notion d'*installation* {installation, bruit de l'installation, installation en fonctionnement, hors fonctionnement de l'installation, installation à l'arrêt}.

Des relations d'hyponymie lient les candidats termes {bruit ambiant, bruit résiduel, bruit aérien, bruit solidien} au terme bruit. En outre, le texte permet d'établir une relation d'antonymie entre les candidats termes {bruit ambiant, bruit résiduel} et une relation de synonymie entre les termes {mesure de bruit et mesure de niveau de bruit}. Ces deux derniers termes suggèrent une relation fonctionnelle entre les notions de bruit et niveau de bruit.

Le terme {émergence} est également sélectionné car il est décrit comme une fonction du bruit ambiant et du bruit résiduel.

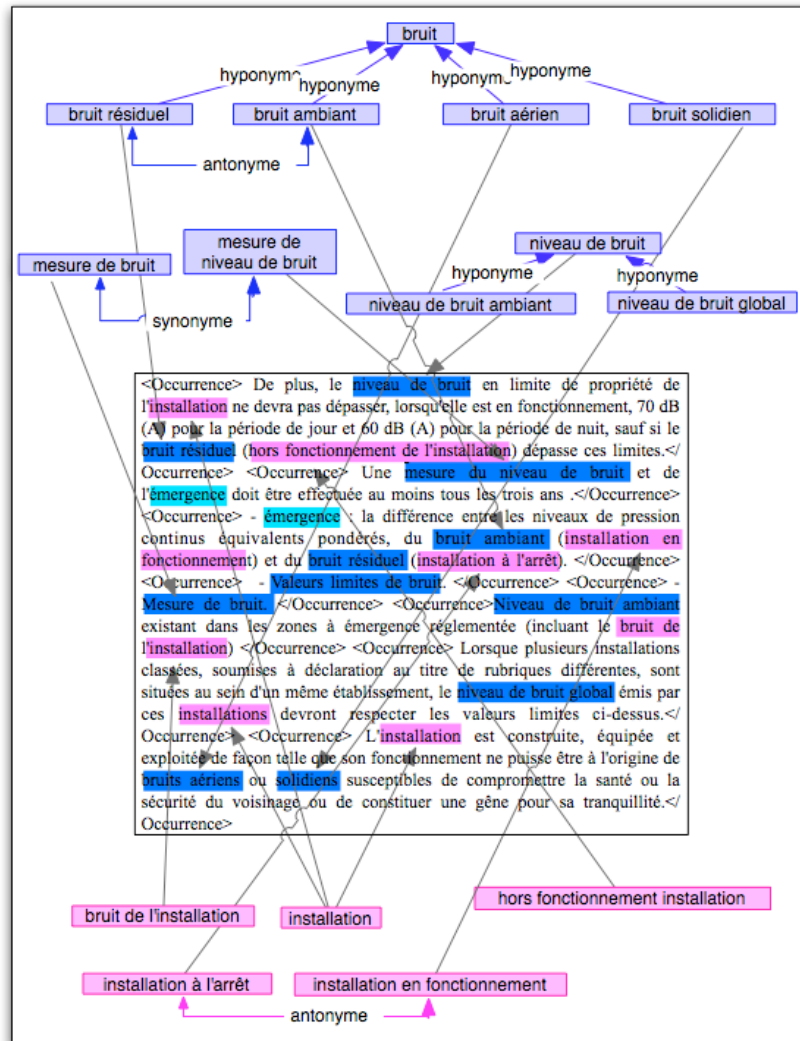


Fig. 3 : réseau terminologique partiel associé au terme bruit

3.3. Les notions associées à l'étape termino-ontologique

L'étape termino-ontologique assure une transition vers le plan ontologique et consiste à construire un modèle conceptuel à partir du réseau terminologique. Les notions de concept terminologique, relation et réseau termino-ontologique utilisées sont introduites *infra*.

Définitions

Un **concept terminologique** est un terme désambiguïsé c'est-à-dire un terme ayant un sens unique dans un contexte d'interprétation [Biebow et *al.*, 1999]. Il est décrit par des propriétés structurelles et définitionnelles [Després et *al.*, 2005].

Une relation **termino-ontologique** est une relation entre concept terminologique (inclusion, identité, disjonction, partie-tout, etc.).

Un réseau **termino-ontologique** est un ensemble de concepts terminologiques reliés entre eux par des relations sémantiques.

Dans cette étape, une transition s'opère. Les termes candidats jugés pertinents servent de point de départ à la construction du modèle. Les relations lexicales qui les lient sont interprétées. Des choix sont à effectuer pour décider ce qui sera concept terminologique ou relation termino-ontologique.

Exemple

La figure 4 présente l'interprétation des unités linguistiques (candidats termes et relations lexicales) décrites dans la figure 3 dans le contexte de la gestion HSE des entreprises.

Les candidats termes « bruit », « niveau de bruit », « mesure de bruit », ... et « installation » sont organisés dans une représentation mettant à jour les relations qui les lient.

Les relations d'hyponymie reliant les candidats termes (tête, expansion) y sont traduites par des relations de subsomption (lien d'inclusion).

Par exemple, « bruit ambiant » est_une_sorte_de « bruit », « bruit résiduel » est_une_sorte_de « bruit ». La différence entre ces deux notions est due à l'état de l'installation (en fonctionnement ou hors fonctionnement). Entre outre, les notions désignées par ces deux termes servent au calcul de l'émergence. Ils sont de bons candidats au statut de concept terminologique.

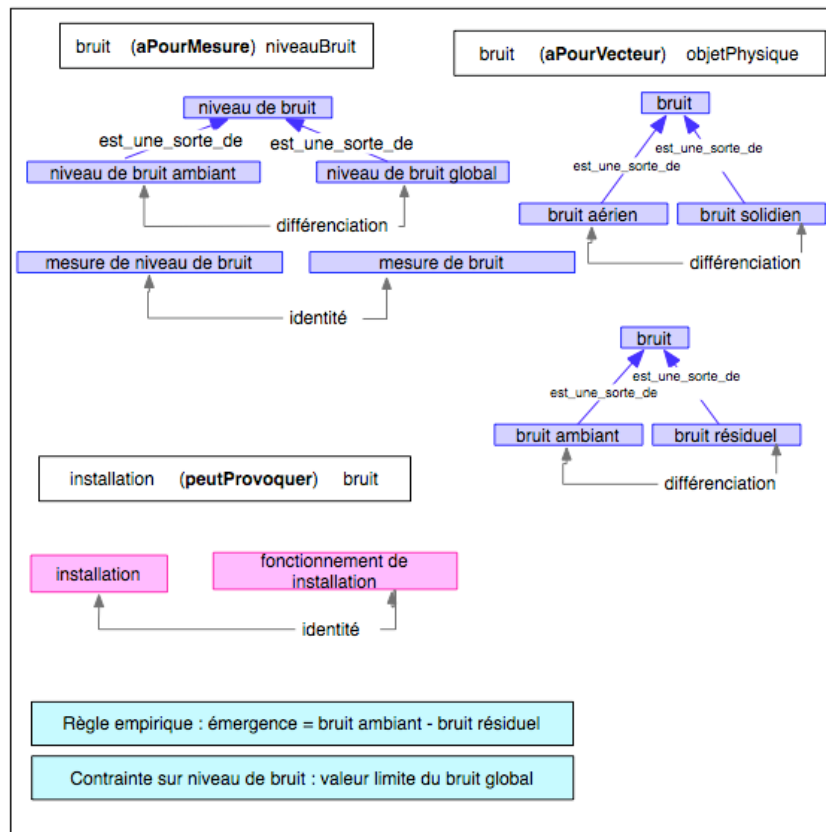


Fig. 4 : liens entre les concepts terminologiques

Réseau terminologique *versus* Ontologie

La différence entre les notions désignées par les deux candidats termes « bruit aérien » et « bruit solidien » tient à l'origine du bruit (air ou solide). Une relation afférente au bruit peut être définie. La relation fonctionnelle « aPourVecteur » liant « bruit » à « objetPhysique » introduit un concept qui n'est pas issu du texte mais qui est utile à la définition de la relation. Cette relation peut ensuite être spécialisée afin de décrire le concept terminologique « bruit aérien » (respectivement « bruit solidien »). Un « bruit aérien » « aPourVecteur » « air » (respectivement « solide »).

Les relations de synonymie sont interprétées comme des relations d'identité. Il reste alors à décider du terme qui désignera le concept terminologique associé.

Le candidat terme « mesure de niveau de bruit » synonyme de « mesure de bruit » est décrit par une relation fonctionnelle « aPourMesure » liant les concepts terminologiques « bruit » et « niveauBruit ».

Dans le contexte de notre application, la notion de bruit y est envisagée selon le point de vue de la réglementation. Le bruit est un phénomène acoustique qui peut provoquer une nuisance. Une nuisance constitue un processus dommageable qui est réglementé par la loi. La notion de bruit dans ce contexte de réglementation est défini par une mesure qui est calculée à partir du bruit en tant que grandeur physique. Le candidat terme « bruit » est polysémique. Par conséquent, deux concepts terminologiques bruitGrandeurPhysique et bruitJuridique sont définis à partir du candidat terme « bruit » afin de construire un modèle du domaine HSE. Une définition en langue naturelle est associée à chacun de ces deux concepts. Les termes les dénotant n'apparaissent pas dans le corpus étudié. En revanche, le terme « installation » est utilisé pour désigner le concept terminologique installation.

La figure 3 présente une représentation partielle du réseau termino-ontologique associé aux concepts terminologiques « bruitGrandeurPhysique », « bruitJuridique » et « installation ».

Le concept terminologique bruitGrandeurPhysique possède un unique sens déterminé par une définition donnée en langue naturelle, des

propriétés intrinsèques et des relations le liant aux concepts terminologique bruitJuridique et installation.

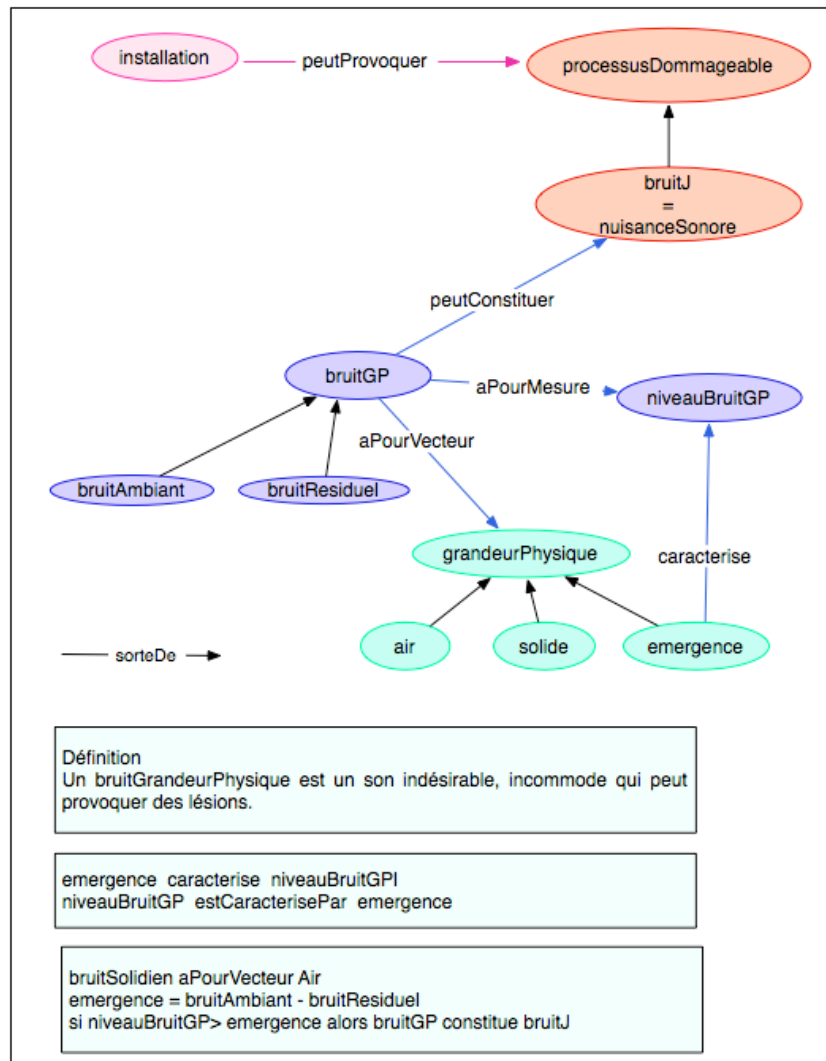


Fig. 5 : réseau termino-ontologique partiel

Réseau terminologique *versus* Ontologie

Définition

Un **concept terminologique** est défini de façon univoque par

- un terme vedette qui peut ne pas figurer dans le corpus ;
- des occurrences du terme qu'il désigne dans le corpus ;
- une définition en langue naturelle, fournie dans le corpus ou élaborée par le cognéticien ;
- un ensemble de synonymes ;
- des propriétés intrinsèques ;
- des relations hiérarchiques ou fonctionnelles les liant aux concepts terminologiques constituant le réseau.

Une **relation termino-ontologique**

- possède une cardinalité ;
- peut être caractérisée par des propriétés (reflexivité, symétrie, transitivité, ...);
- peut avoir une relation inverse.

3.4. Les notions ontologiques

L'étape ontologique consiste à traduire le modèle conceptuel obtenu à la fin de l'étape termino-ontologique dans un langage formel qui permet à la fois de s'affranchir des problèmes liés à la langue naturelle et au contexte, et d'effectuer des calculs et des raisonnements sur les entités représentées.

Définitions

Une **ontologie informatique** est constituée de concepts, appelés ontologique afin d'éviter toute confusion, organisés dans une hiérarchie avec héritage des propriétés, liés par des relations ontologiques existant entre eux et contraints par des règles et des axiomes. Elle permet en particulier d'implanter des mécanismes de raisonnement déductif, de classification automatique ou de recherche d'information.

Un **concept ontologique** est décrit dans un langage formel de représentation des connaissances. Il possède des propriétés.

Une **relation ontologique** (hiérarchique, descriptive) est une relation liant des concepts ontologiques, construite à partir de relations termino-ontologiques et décrite dans un langage formel.

Une **relation hiérarchique** exprime un héritage des propriétés du concept. Une **relation descriptive** est une relation binaire qui exprime un lien entre deux concepts.

Les entités à représenter sont :

1. celles constituant la base de l'ontologie : (a) les classes représentant un ensemble d'individus partageant les mêmes propriétés. Elles sont assimilées aux concepts ontologiques.

Réseau terminologique *versus* Ontologie

(b) les propriétés associant des paires attributs/valeurs à des individus ou restreignant des classes ; (c) les individus représentant une instance d'une classe.

2. les relations sémantiques qui comportent celles : (a) liant les entités constituant l'ontologie ; (b) décrivant les relations.

La représentation de ces différentes entités requiert de respecter les contraintes liées au formalisme adopté. Cette traduction n'est pas toujours possible et exige alors une adaptation du modèle conceptuel et des choix de représentation.

Différents formalismes de représentation des connaissances peuvent être utilisés pour décrire une ontologie. Parmi ceux existants, tous ne possèdent pas les mêmes niveaux d'expressivité et ne garantissent pas le même niveau de calculabilité.

Dans ce papier, nous adoptons la représentation dans le langage OWL DL (<http://www.w3.org/2004/OWL/>) qui repose sur les logiques de description [Baader et al., 2003].

Exemple

Un concept ontologique est décrit par une expression logique exprimant des relations du concept avec d'autres concepts de l'ontologie. La figure 5 présente une expression partielle de bruit en langage OWL utilisant la syntaxe XML.

```

bruitAmbiant est une sous classe de bruitGP
<SubClassOf>
  <OWLClass URI="&Ontology1222676008394;bruitAmbiant"/>
  <OWLClass URI="&Ontology1222676008394;bruitGP"/>
</SubClassOf>

bruitAmbiant et bruitResiduel sont des classes disjointes
<DisjointClasses>
  <OWLClass URI="&Ontology1222676008394;bruitAmbiant"/>
  <OWLClass URI="&Ontology1222676008394;bruitResiduel"/>
</DisjointClasses>

bruitGP aPourOrigine {air ou son }
<ObjectPropertyRange>
  <ObjectProperty
    URI="&Ontology1222676008394;aPourOrigine"/>
  <ObjectUnionOf>
    <OWLClass URI="&Ontology1222676008394;air"/>
    <OWLClass URI="&Ontology1222676008394;solide"/>
  </ObjectUnionOf>

```

Fig. 4 : expression partielle du concept ontologique bruitGP

La figure 6 présente une expression partielle des différentes relations liant les concepts ontologiques bruit GP, bruitJ et la hiérarchie des relations.

```

relation entre bruitGP et bruit] : peutProvoquerNuisanceSonore
relation entre bruitGP et air ou solide : aPourOrigine
<ObjectPropertyRange>
  <ObjectProperty
    URI="&Ontology1222676008394;peutProvoquerNuisanceSonore"/>
  <OWLClass URI="&Ontology1222676008394;bruit]"/>
</ObjectPropertyRange>

<ObjectPropertyRange>
  <ObjectProperty
    URI="&Ontology1222676008394;aPourOrigine"/>
  <ObjectUnionOf>
    <OWLClass URI="&Ontology1222676008394;air"/>
    <OWLClass URI="&Ontology1222676008394;solide"/>
  </ObjectUnionOf>
</ObjectPropertyRange>

hiérarchie de relations entre : peutProvoquer et
peutProvoquerNuisanceSonore
<Declaration>
  <ObjectProperty
    URI="&Ontology1222676008394;peutProvoquerNuisanceSonore"/>
</Declaration>

<ObjectPropertyRange>
  <ObjectProperty
    URI="&Ontology1222676008394;peutProvoquer"/>
  <OWLClass URI="&Ontology1222676008394;nuisance"/>
</ObjectPropertyRange>

```

Fig. 5 : expression partielle des relations ontologiques

A partir de cette représentation, il devient possible de raisonner. Ainsi, il est possible de déduire l'émergence à partir des concepts mesureDeBruitAmbient et mesureDeBruitResiduel et de déterminer le seuil pour lequel l'émergence est considérée comme correspondant à une nuisanceSonore qui est un concept équivalent de bruit].

4. Conclusion

La finalité de ce papier était de montrer comment le passage du terme au concept se construit.

Notre position est que la transition s'effectue graduellement en passant par trois plans bien identifiés : linguistique, termino-ontologique et ontologique. En cela, nous proposons une étape supplémentaire à l'approche ontoterminologique proposée par Roche [Roche, 2007]. En effet, la réalisation du modèle conceptuel est dissociée de la phase ontologique.

Le concept terminologique joue le rôle de passeur entre les plans linguistique et ontologique. Il peut-être envisagé selon deux facettes. Il a d'une part, une expression en langue, mais restreinte par une expression dans un langage semi-formel [Kassel, 2002] d'autre part, il est défini par des propriétés structurelles et fonctionnelles. Il en est de même pour les relations termino-ontologiques qui constituent une traduction des relations lexicales repérées dans le texte.

Les ressources associées à chacun de ces plans sont caractérisées par des propriétés qui les distinguent mais qui les rendent complémentaires. Ainsi, l'ontologie permet d'effectuer des tâches de classification et de raisonnement qui sont logiquement fondées tandis que le réseau terminologique permet de représenter les notions exprimées dans les textes selon le contexte de l'application visée. Le modèle conceptuel dit réseau termino-ontologique permet de changer de point de vue et d'assurer la transition entre le modèle linguistique et ontologique.

A l'issue de ce papier, il devient possible de répondre positivement à la question « Est-il possible de construire une ontologie à partir de textes ? ».

5. Remerciements

Que soient ici remercier Jérôme Nobecourt membre du LIM&BIO et Haïfa Zargayouna membre de l'équipe RCLN pour les discussions fructueuses occasionnées par l'écriture de ce papier.

Bibliographie

Aussenac-Gilles N., Després S., Szulman S. *The TERMINAE Method and Platform for Ontology Engineering from Texts. Bridging the Gap between Text and Knowledge – Selected Contributions to Ontology Learning and Population from Text*. Eds Buitelaar P., Cimiano, P. pp. 199-223. IOS Press, 2008.

Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P. F. *The Description Logic Handbook: Theory, Implementation, Applications*. Cambridge University Press, Cambridge, UK, 2003.

Biebow B., Szulman S. *TERMINAE : a method and a tool to build a domain ontology*, In Proceedings of International Workshop on Ontological Engineering on the Global Information Infrastructure, eds Benjamins V. R. and Fensel D. and Gomez Perez A., pp. 25-30, 1999.

Bourigault D., Charlet J. *Construction d'un index thématique de l'ingénierie des connaissances*, Ingénierie des connaissances. De Régine Teulier, Jean Charlet Publié par L'Harmattan, pp.29-47, 2005.

Bourigault D. , Aussenac-Gilles N. *Construction d'ontologies à partir de textes*, in TALN, Batz sur Mer, 2003.

Bourigault D., Jacquemin C. *Construction de ressources terminologiques*, in J.-M. Pierrel (éd.), Industrie des langues, Hermès, Paris, pp. 215-233, 2000.

Cimiano P., Volker J., Studer, R. *Ontologies on Demand? - A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text*. In Information Wissenschaft and Praxis, 57, pp.315-320, 2006.

Cimiano P. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer. November 2006.

Cruse, D. A. 1986. *Lexical semantics*. Cambridge, England: University Press.

Després S., Szulman S. — *Construction d'une ontologie formelle à partir d'un texte de droit*, pp. 261-281, numéro 11(2) de International journal of theoretical and applied issues in specialized communication. John Benjamins Publishing Company, 2005.

Després S., Furst F., Szulman S. *Construction d'une ontologie du domaine HSE*. In actes de la conférence Ingénierie des Connaissances IC'2007, pp. 133-144, 2007.

Kassel G. *Une méthode de spécification semi-informelle d'ontologies*. In Actes des 13 èmes journées francophones d'Ingénierie des Connaissances IC 2002 pp. 75-87, 2002.

Mondary T., Després S., Nazarenko A., Szulman, S. *Construction d'ontologies à partir de textes : la phase de conceptualisation*. In actes de la conférence Ingénierie des Connaissances IC '2008, pp. 87-98, 2008.

Roche C. *Le terme et le concept : fondements d'une ontoterminologie*. In acte Conférence Terminologie & Ontologie : Théories et Applications Toth 2007.