



HAL
open science

Ontologies pour le Web sémantique

Jean Charlet, Bruno Bachimont, Raphaël Troncy

► **To cite this version:**

Jean Charlet, Bruno Bachimont, Raphaël Troncy. Ontologies pour le Web sémantique. CNRS, réseau thématique pluridisciplinaire Documents publié dans le numéro spécial Web sémantique de la revue I3 (https://www.irit.fr/journal-i3/hors_serie/annee2004/index_fr.php). 2004. hal-03854420

HAL Id: hal-03854420

<https://hal.science/hal-03854420v1>

Submitted on 15 Nov 2022

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.

Ontologies pour le Web sémantique

Jean Charlet¹, Bruno Bachimont^{2,3}, Raphaël Troncy⁴

¹Mission de recherche STIM, AP-HP & INSERM ERM 0202

Jean.Charlet@spim.jussieu.fr

²Institut National de l'Audiovisuel

bbachimont@ina.fr

³Université Technologique de Compiègne

bruno.bachimont@utc.fr

⁴ISTI-CNR

raphael.troncy@isti.cnr.it

Résumé

Les ontologies sont un des concepts de base du Web sémantique. Dans cette partie, nous voulons revenir sur l'origine des travaux sur les ontologies au sein de la communauté Ingénierie des connaissances pour montrer les acquis sur lesquels les chercheurs se sont appuyés pour développer le concept au sein du Web sémantique. Dans un souci de clarté et de précision, nous revenons sur la définition des ontologies pour clarifier les limites de l'approche ontologique et les contraintes qui s'imposent à elle. Cela nous permet d'aborder les méthodologies de construction d'ontologies et la question de la généricité.

Ceci posé, nous pouvons passer en revue les apports du Web sémantique en termes de méthodologies, d'outils, d'éditeurs en essayant de préciser les problématiques particulières dans chaque domaine. Enfin, dans une dernière section, nous tentons de faire le point sur les problèmes que rencontre le Web sémantique par rapport aux ontologies et les axes de réflexion ou de recherche qui semblent prometteurs en ce domaine.

Mots-clés : *Ontologies, Ingénierie des connaissances, thésaurus, corpus.*

Abstract

Ontologies play a key role in the Semantic Web. In this paper, we would like to come back on the origins of the works on ontologies inside the knowledge engineering community in order to show the assets on which the researchers have developed this concept for the Semantic Web. First, we give some definitions of ontologies in order to clarify the limits of such approaches and to remind the constraints that they have to respect. We can therefore present the current state of the art on the methodologies of ontology construction and question the genericity of the ontologies.

Within this framework, we review the results already achieved about the methodologies and the tools for building ontologies while pointing out some problems raised by the Semantic Web. Finally, we try to state the critical issues that the Semantic Web has to face to with respect to ontologies, and to draw the future research directions in order to give a hint of the promising results we can expect in this domain.

Keywords : *Ontologies, Knowledge Engineering, thesaurus, corpora.*

1. PRÉSENTATION ET IMPORTANCE DE LA PROBLÉMATIQUE DU POINT DE VUE DES USAGES

Permettre un traitement symbolique des connaissances, le but premier de l'Intelligence Artificielle, rejoint celui du Web sémantique dans les attendus de ses « créateurs », à savoir, se donner les moyens de faire faire des traitements automatiques à des modules logiciels au sein du Web, que ce soit pour faire interagir et interopérer des machines entre elles ou faire interagir des machines avec des humains. En effet, ces traitements rejoignent directement la question de la représentation des connaissances et des inférences en Intelligence Artificielle.

Tout programme informatique manipule, à travers des symboles, les objets du domaine modélisé. L'ensemble de ces objets correspond à ce qui est appelé *référentiel* dans le domaine des systèmes d'informations. Pour un domaine mettant en œuvre des connaissances complexes sur lesquelles on veut effectuer des traitements intelligents¹, le programme

¹ Nous n'entrons pas dans le débat sur ce qu'est l'intelligence. Nous nous bornons à noter que l'Intelligence artificielle intervient dans des domaines et pour des problèmes où l'on ne peut construire des programmes avec des solutions algorithmiques simples et pour lesquels les connaissances à utiliser sont difficilement formalisables.

élaboré est un système à base de connaissances qui manipule une base de connaissances. Cette base répertorie, entre autres, et de la même façon que dans les systèmes d'information, les concepts du domaine hiérarchiquement organisés dans une « ontologie », nommée ainsi en faisant référence à ARISTOTE².

Les ontologies sont alors centrales pour le Web sémantique qui, d'une part, cherche à s'appuyer sur des modélisations de ressources du Web à partir de représentations conceptuelles des domaines concernés et, d'autre part, a pour objectif de permettre à des programmes de faire des inférences dessus. Les recherches à leur sujet sont donc indispensables. Une fois construite et acceptée par une communauté particulière, une ontologie doit en effet traduire un consensus explicite et un certain niveau de partage, deux aspects essentiels pour permettre l'exploitation des ressources du Web par différentes applications ou agents logiciels. D'autre part, la formalisation, autre facette des ontologies, est nécessaire pour qu'il soit possible de raisonner automatiquement dessus afin de décharger les utilisateurs d'une partie de leur tâche d'exploitation et de combinaison des ressources du Web. Les ontologies servent alors (1) pour le vocabulaire, la structuration et l'exploitation des méta-données (*cf.* chap. 3, ce volume), (2) comme représentation pivot pour l'intégration de sources de données hétérogènes (*cf.* chap. 5, ce volume) (3) pour décrire les services Web et, en général, partout où il va être nécessaire d'appuyer des modules logiciels sur des représentations sémantiques nécessitant un certain consensus.

2. MÉTHODES, TECHNIQUES, OUTILS EXISTANTS SUR LESQUELS ON PEUT S'APPUYER

² ARISTOTE a défini l'Ontologie comme la science de l'Être. Définition que l'on retrouve dans le Petit Robert avec « **Ontologie** » : *la partie de la métaphysique qui s'intéresse à l'Être en tant qu'Être* ». Mais l'Ontologie est habituellement davantage comprise comme une science des étants que comme une science de l'Être en tant qu'Être, c'est-à-dire qu'elle s'intéresse davantage à ce qui existe (les étants ou existants) qu'aux principes de ce qui existe (l'Être). Cette science, l'Ontologie, produit des ensembles, les ontologies. Pour être complet, notons que le mot lui-même date du 17^e siècle, avec des dates précises discordantes selon les sources. Enfin, le mot ontologie, utilisé dans le contexte de l'Ingénierie des connaissances ou du Web sémantique, semble d'un usage abusif à certains. Nous n'entrerons pas dans ce débat de dénomination en nous bornant ici à éclairer la filiation de réflexion.

2.1. Définitions

2.1.1. Les ontologies en Ingénierie des connaissances

Les ontologies sont apparues au début des années 90 dans la communauté Ingénierie des connaissances, dans le cadre des démarches d'acquisition des connaissances pour les systèmes à base de connaissances (SBC). Faisant suite aux systèmes experts qui séparaient une base de connaissances « déclarative » et un moteur d'inférence « procédural », les SBC proposaient alors de spécifier, d'un côté, des connaissances du domaine modélisé et, de l'autre, des connaissances de raisonnement décrivant les règles heuristiques d'utilisation de ces connaissances du domaine. L'idée de cette séparation modulaire était de construire mieux et plus rapidement des SBC en réutilisant le plus possible des composants génériques, que ce soit au niveau du raisonnement ou des connaissances du domaine. Ces dernières précisant tout ce qui a trait au domaine. Dans ce contexte, les chercheurs ont proposé de fonder ces connaissances sur la spécification d'une *ontologie*, ensemble structuré par différentes relations, principalement l'hyperonymie³ des objets du domaine dont on note déjà que son élaboration relève de choix du modélisateur.

Ontologie (déf. 1) : *Ensemble des objets reconnus comme existant dans le domaine. Construire une ontologie c'est aussi décider de la manière d'être et d'exister des objets.*

Dans cette définition, les objets ne sont pas pris dans un sens informatique mais comme objets du monde réel que le système modélise. Pour poursuivre vers une définition de l'ontologie, il nous semble indispensable de rappeler que les travaux sur *les ontologies sont développés dans un contexte informatique – que ce soit celui de l'Ingénierie des connaissances, de l'Intelligence artificielle ou de la gestion et des systèmes d'information ou plus spécifiquement ici le contexte du Web sémantique – où le but final est de spécifier un artefact informatique.* Dans ce contexte, l'ontologie devient alors un modèle des objets existants qui y fait référence à travers des concepts, les concepts du domaine. Cette ontologie et les connaissances du domaine

³ Pour ne pas dévier de notre propos, nous ne développons pas l'historique de la classification en Intelligence artificielle mais un certain nombre de travaux s'étaient déjà posés des questions au sujet de la classification des objets du monde. Parmi ceux-ci, on peut citer les travaux autour des réseaux sémantiques, comme Brachman [13, 14] s'interrogeant sur la nature du lien is-a ou Woods [52] s'attachant à des questions de représentation et de niveau de description des connaissances.

qui s'appuient dessus correspondent, à la suite des travaux de Newell [34], à une modélisation au « niveau des connaissances » (le *Knowledge Level* de Newell). Ce niveau de modélisation de l'ontologie nous situe au niveau du sens, au niveau de l'*intension*. Cette ontologie devra ensuite être *opérationnalisée*, c'est-à-dire codée dans un langage opérationnel, exécutable.

Pour avancer vers une définition complète, on peut revenir à un travail de N. Guarino et P. Giaretta [26] qui ont essayé de trouver des définitions de l'ontologie durant son processus d'élaboration⁴, à savoir l'ontologie comme un *système conceptuel informel*, puis l'ontologie comme *la représentation d'un système conceptuel via une théorie logique et son vocabulaire*. On peut alors proposer la seconde définition de ce qu'est une ontologie, en reprenant les spécifications de Gruber [23] et M. Uschold *et al.* [49], l'équipe de ce dernier ayant appliqué les résultats des réflexions du premier dans le contexte applicatif d'une ontologie d'entreprise, *the enterprise ontology* :

Ontologie (déf. 2) : *Une ontologie implique ou comprend une certaine vue du monde par rapport à un domaine donné. Cette vue est souvent conçue comme un ensemble de concepts – e.g. entités, attributs, processus –, leurs définitions et leurs interrelations. On appelle cela une conceptualisation.*

[...]

Une ontologie peut prendre différentes formes mais elle inclura nécessairement un vocabulaire⁵ de termes et une spécification de leur signification.

[...]

Une ontologie est une spécification rendant partiellement compte d'une conceptualisation.

Cette seconde définition propose un autre point de vue que la première, cohérent avec elle mais plus précis, en termes de spécification et par rapport à une application informatique. Elle nous permet de

⁴ Les développements faits ici sont une libre interprétation des attendus des travaux de N. Guarino et P. Giaretta [26]. Ils ont pour but de progresser vers une définition rendant compte d'un processus de construction évolutif. Nous renvoyons le lecteur qui veut approfondir la question aux articles de ces auteurs.

⁵ « Vocabulaire » est utilisé ici tel qu'il apparaît dans le texte. Il doit être compris dans un sens logique et être vu comme le vocabulaire des expressions manipulées par une théorie logique. Ceci dit, ce vocable prête à confusion dans le domaine des ontologies où la question de la langue et l'opposition *termes versus concepts* est fondamentale (*cf.* § 2.2.2).

préciser les contraintes qui s'imposent successivement au concepteur d'ontologies :

- une ontologie est bien une conceptualisation, entendons par là que l'on y définit des concepts ;
- devant être par la suite utilisée dans un artefact informatique dont on veut spécifier le comportement, l'ontologie devra également être une théorie logique pour laquelle on précisera le vocabulaire manipulé ;
- enfin, la conceptualisation étant spécifiée parfois de manière très précise, une théorie logique ne peut pas toujours en rendre compte de façon exacte : elle ne peut assumer la richesse interprétative du domaine conceptualisé dans une ontologie et ne le fait donc que partiellement⁶.

2.1.2. Que représente-t-on dans une ontologie ?

À partir des définitions proposées jusque-là pour les ontologies⁷, quatre grands types de caractéristiques nous permettent de préciser ce qui peut être représenté dans une ontologie ainsi que le processus de modélisation :

Le type d'ontologie. Les méthodes en Ingénierie des connaissances ont répertorié plusieurs types d'ontologie liés à l'ensemble des objets conceptualisés et manipulés au sein d'un SBC. Nous allons en citer quelques-unes : *a*) l'ontologie du domaine (*cf.* § 1⁸), *b*) l'ontologie générique, ou qui se veut comme telle, qui repère et organise les concepts les plus abstraits du domaine (*cf.* § 2.2.4), *c*) l'ontologie d'une méthode de résolution de problème où le rôle joué par chaque concept dans le raisonnement est rendu explicite (*p. ex. signe* ou *syndrome* dans le cadre du raisonnement médical), *d*) l'ontologie d'application qui se veut une double spécialisation : d'une ontologie du domaine et d'une ontologie de méthode, enfin *e*) l'ontologie de représentation qui repère et organise les primitives de la théorie

⁶ Cet écart est principalement dû des problèmes de calculabilité des théories qui pourraient rendre compte de la spécification et sont donc inadéquates. Cet écart entre la conceptualisation et la spécification formelle est décrit par N. Guarino comme l'*engagement ontologique* que le concepteur doit assumer au passage de l'une à l'autre [26].

⁷ Nous ne développerons pas ici l'historique de la classification, de Aristote aux Sciences naturelles. Nous renvoyons le lecteur intéressé à [15] et rappelons qu'un des apports de cet historique est de montrer toute la variabilité de définition des objets du monde en fonction des buts du modélisateur.

⁸ La numérotation des références de paragraphes correspond aux différentes parties de cet article.

logique permettant de représenter l'ontologie (*p. ex.* la *frame ontology* d'ONTOLINGUA [23]) ou l'ontologie de « propriétés » de Guarino & Welty [27].

Les propriétés. Une ontologie est non seulement le repérage et la classification des concepts mais c'est aussi des caractéristiques qui leur sont attachées et qu'on appelle ici des propriétés⁹. Ces propriétés peuvent être évaluées. En s'intéressant aux taxinomies en sciences naturelles, les vertébrés ont un tégument (la peau) comportant des poils – *p. ex.* pour les mammifères – ou des plumes – *p. ex.* pour les oiseaux. Dans une ontologie sur le monde animal, on pourra ainsi avoir les concepts de « mammifère » ou « d'oiseau » pour lesquels est précisé le type de tégument, respectivement à poil et à plume. En pratique, un attribut « tégument » pourra être attaché aux concepts et sa valeur variera suivant le concept auquel on fait référence.

La relation « is-a ». La relation de subsomption *is-a* qui définit un lien de généralisation – *i.e. hyperonymie* – est utilisée pour structurer les ontologies. Cette relation qui permet formellement l'héritage de propriétés est un choix qui s'impose depuis ARISTOTE. Elle doit être complétée par d'autres relations pour exprimer la sémantique du domaine.

Les autres relations. Les relations unissent les concepts ensemble pour construire des représentations conceptuelles complexes qui vont être autant de connaissances nécessaires au SBC que l'on construit. Si la connaissance construite correspond à un concept dans le monde modélisé, celui-ci est dit *défini*, à l'opposé des concepts insérés dans l'arborescence de l'ontologie qui sont dits *primitifs*. Par exemple, si l'on définit l'*appendicite* comme une *inflammation localisée-sur l'appendice*, c'est un concept dit défini. Dans l'exemple précédent, *localisée-sur* est une relation binaire qui se définit par les concepts qu'elle relie et par le fait qu'elle est, comme les concepts, insérée dans une hiérarchie, ici de relations.

La relation *is-a* qui structure l'ontologie est une relation du même type que les autres. Elle a cela de spécifique que c'est elle qu'on a justement choisi comme relation de structuration de l'arborescence ontologique. Elle est donc implicite dans cette ontologie. Au niveau des choix, il faut aussi remarquer que les concepts et relations de l'ontologie sont duals l'un par rapport à l'autre. Un concept primitif pourrait être un concept défini, une relation pourrait se retrouver implicitement définie au sein

⁹ Des attributs dans le contexte des langages à objets.

d'un concept primitif. Ce sont les choix assumés du concepteur de l'ontologie qui auront permis de décider de ce qui est essentiel – et donc primitif – ou non (*cf.* § 2.2.2). Ainsi, on peut décider que le fait, pour un être humain, d'être un étudiant est temporaire donc non définitoire. On caractérise alors les êtres humains avec une relation de rôle social qui permettra de préciser une fonction d'étudiant ou de professeur.

Un autre choix de conception qui doit être fait durant l'élaboration d'une ontologie est de décider si une connaissance doit être modélisée dans une propriété ou à l'aide d'une relation pointant sur un autre concept. Un critère peut être de dire que c'est une propriété dès lors que les valeurs possibles sont d'un type dit primitif (entier, chaîne de caractères), et c'est une relation dès lors que les valeurs possibles sont d'un type dit complexe c'est-à-dire un autre concept de l'ontologie. Mais cette frontière peut aussi être remise en question.

Enfin, dans certains cas, il peut être nécessaire de compléter la structuration de l'ontologie par la relation is-a avec une relation de partie-tout ou méronymie. Ce type de relation est, par exemple, indispensable en anatomie médicale où il est nécessaire de décrire des organes ou des systèmes et ce qui les compose. Cette relation n'est pas sans poser des problèmes de modélisation dans la mesure où elle est, selon les situations, transitive ou intransitive.

Les réflexions sur les ontologies dans le contexte du Web sémantique s'appuient sur ces différents acquis en notant que dans le cas où l'ontologie est utilisée comme repérage et structuration de méta-données, le fait qu'elle permette de faire des inférences est moins mis en avant et donc moins recherché que dans le cas où l'ontologie est utilisée au sein d'un module logiciel type SBC, nécessitant justement d'effectuer des inférences.

2.2. Quelles méthodes pour construire des ontologies ?

2.2.1. À la recherche d'une méthodologie constructive

Les méthodologies de construction d'ontologies ne sont pas légion. Nous entendons par là, la donnée argumentée de procédures de travail, d'étapes, qui décrivent le pourquoi et le comment de la conceptualisation puis de l'artefact construit. Sur le fond, on trouve des réflexions linguistiques et épistémologiques [3] et des réflexions sur les ontologies formelles [26]. Elles sont à l'origine de méthodes de construction

d'ontologies à partir de corpus que nous allons aborder dans la section suivante. Ensuite, on trouve des travaux qui proposent de bonnes pratiques méthodologiques de construction d'ontologies comme ceux de Th. R. Gruber [23] ou de M. Fernández *et al.* [19]. Nous les présenterons plus rapidement dans la section 3.2.1.

2.2.2. Acquérir des ontologies à partir de corpus

La première méthode que nous présentons ici a été mise au point à l'occasion de la construction de l'ontologie du projet MENELAS¹⁰ [53]. Elle est fondée sur les travaux de B. Bachimont [3] enrichis de considérations sur les corpus textuels développées au sein du groupe TIA¹¹. Nous choisissons de la décrire ici car l'ensemble de ses étapes est paradigmatique de ces approches et nous permettra d'aborder, dans la suite de cette étude, les problèmes liés à la construction des ontologies. Les 4 étapes de la méthode sont :

1^{re} étape : la primauté du corpus et son analyse.

Nous sommes, par hypothèse, dans des domaines où les données et les connaissances s'expriment en langue. Il est alors aisé¹² de trouver des textes explicitant les connaissances du domaine, dans des buts de pratique ou de transmission des connaissances.

On peut ainsi construire un corpus textuel qui sera la source privilégiée permettant de caractériser les notions utiles à la modélisation d'une ontologie et le contenu sémantique qui lui correspond. Pour ce faire, on utilise une « démarche de corpus » et des outils terminologiques pour commencer à modéliser le domaine. Ces outils, pour la plupart, reposent sur la recherche de formes syntaxiques particulières manifestant les notions recherchées comme des syntagmes nominaux pour des candidats termes, des relations syntaxiques marqueurs de relations sémantiques, ou des proximités d'usage – *e.g. contextes partagés* – pour des regroupements de notions [9, 10].

2^e étape : la normalisation sémantique.

¹⁰ <http://www.biomath.jussieu.fr/projets/Menelas/ontologie>

¹¹ <http://www.biomath.jussieu.fr/TIA/>

¹² Dans le sens où ces textes existent bien et en quantité. Nous n'abordons pas la question, parfois complexe, de la disponibilité réelle de ces textes sous forme numérique. La question des corpus est développée plus spécifiquement dans le cadre de l'Action spécifique du CNRS «Construction de ressources terminologiques à partir de corpus» – ASSTICCOT <<http://www.irit.fr/ASSTICCOT/>>.

L'étape précédente fournit des candidats termes dont les libellés ont un sens pour le lecteur, souvent spécialiste du domaine. Mais rien n'assure que ce sens soit unique : au contraire, nous sommes dans un fonctionnement linguistique où les significations sont ambiguës, les définitions circulaires et dépendent en particulier du contexte interprétatif des locuteurs. Or, dans la modélisation ontologique, on cherche à construire des primitives dont le sens ne dépend pas des autres primitives et est surtout non contextuel. Il est nécessaire, pour prendre le chemin du formel, de *normaliser* les significations des termes pour ne retenir, pour chacun d'eux, qu'une seule signification, qu'une seule interprétation possible par un être humain. C'est ce que permet l'utilisation de la *sémantique différentielle*, proposée par B. Bachimont et travaillant les proximités d'usage produites à l'étape précédente. À la fin de cette étape, nous avons un arbre de primitives conceptuelles valable dans la seule région du monde modélisée où les concepts retenus correspondent bien à ceux de l'ontologie, par définition décontextualisée. Nous avons une *ontologie régionale*.

Corollaire de la méthodologie, le sens d'un concept est principalement défini en fonction de ce qu'il est (héritage par rapport au père) et de ce qu'il n'est pas (différence par rapport au père ou aux frères).

3^e étape : l'engagement ontologique.

L'engagement ontologique correspond à l'évolution de l'ontologie régionale vers une ontologie formelle. La sémantique formelle ne considère plus des notions sémantiques mais des extensions, c'est-à-dire l'ensemble des objets qui vérifient des propriétés définies en intension dans l'étape précédente, propriétés ayant une définition formelle à ce niveau. La structure de l'ontologie formelle est alors un treillis [3].

Au sein de cette ontologie formelle, le treillis des concepts doit être compris comme la possibilité de créer des concepts dits définis en combinant les concepts primitifs comme, par exemple, une *personne* qui a pour rôle social d'être un *médecin*. Cet « objet » défini en extension hérite des caractéristiques des personnes et des médecins alors qu'au niveau précédent, l'intension des personnes et des médecins était irréductible.

4^e étape : l'opérationnalisation.

Dernière étape de la méthodologie (et de beaucoup d'autres), l'opérationnalisation consiste en la représentation de l'ontologie

dans un langage de représentation des connaissances permettant de surcroît des services inférentiels de type classification des concepts ou généralisation, etc. Selon les langages considérés, les calculs possibles et donc les services inférentiels ne sont pas identiques et, à ce niveau aussi, il y a un engagement qui est pris avec de nouvelles contraintes et possibilités, justifiant l'existence d'une *ontologie computationnelle*.

Il y a plusieurs possibilités de choix de langage d'opérationnalisation, comme les graphes conceptuels et les logiques de description, permettant, l'un comme l'autre, d'effectuer un certain nombre d'opérations sur des ontologies : inférences propres aux structures de graphes comme la jointure ou la projection pour les graphes conceptuels, classifications dans des structures arborescentes pour les logiques de description. Les logiques de description sont le type de langage retenu par le Web sémantique pour représenter les ontologies et faire des inférences dessus. Les différents axiomes qui accompagnent chaque possibilité ne sont pas anodins en termes de connaissances représentables et de calculabilité. Ce problème est approfondi au chapitre 2.

Cette méthodologie a été utilisée dans plusieurs domaines comme l'audiovisuel [46] ou la médecine [31, 51].

Une 2^e méthodologie, réifiée dans le système TERMINAE développé par N. Aussenac-Gilles *et al.* [2], est dans un paradigme identique à la précédente mais *a*) le dépouillement des corpus et l'étude linguistique y sont encore plus mis en avant, en particulier le repérage des relations et *b*) la question de l'opérationnalisation dans une logique de description et son influence sur la modélisation a été spécifiquement étudiée. Ainsi, aux outils de repérage de candidats termes et de regroupement de contextes – SYNTAX et UPERY [9] – est rajouté un outil de repérage de relations, YAKWA [38], proposant une meilleure aide au repérage de celles-ci.

Enfin, des approches cherchant à plus automatiser la structuration de l'ontologie via des algorithmes d'apprentissage ont été développées comme celle de Maedche et Staab [32] ou ASIUM développé par Faure et Nedellec [18] :

- La première approche cherche à minimiser les interactions et construit automatiquement une taxinomie de concepts à partir d'un dictionnaire du domaine. Le résultat obtenu, comportant beaucoup de concepts « feuilles » (terminaux) doit être complété par une extraction de candidats termes placés manuellement pour finir de structurer l'ontologie.

- La seconde approche, ASIUM, utilise des techniques d'apprentissage pour construire interactivement une ontologie. Plus précisément, ASIUM apprend des cadres de sous-catégorisation de verbes et des ontologies à partir de l'analyse syntaxique de corpus de textes techniques. Cette technique est à rapprocher de la construction de notions par les contextes partagés (*cf. supra*).

Les différentes expériences menées jusque-là avec Les méthodologies décrites ici montrent que l'intervention humaine est indispensable et même primordiale dans les choix de conceptualisation à faire. Il est alors plus important d'outiller cette intervention que de chercher à la minimiser.

2.2.3. Ontologie *versus* thesaurus

Des modélisations conceptuelles ou terminologiques existent depuis longtemps dans le domaine de la recherche d'information au sein des bibliothèques et dans le domaine de la terminologie, par exemple en médecine où il existe, entre autres, des thesaurus de spécialités répertoriant l'ensemble des termes médicaux à utiliser pour décrire l'activité médicale – *i.e.* un vocabulaire contrôlé. En reprenant un vocable de Bourigault *et al.* [10], les différentes ressources terminologiques ou ontologiques (RTO) élaborées dans différents domaines doivent être précisées et *conceptuellement* caractérisées pour bien comprendre leur signification par rapport à une modélisation conceptuelle et formelle et si elles peuvent être ou servir d'ontologies et à quelles conditions. Pour cela, nous allons reprendre ici, rapidement, trois définitions par rapport à des produits terminologiques existant parallèlement aux ontologies, les *thesaurus*, *classifications* et *terminologies* :

Un thesaurus est un ensemble de termes normalisés fondé sur une structuration hiérarchisée. Les termes y sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques. Organisé alphabétiquement, il forme un répertoire alphabétique de termes normalisés pour l'analyse de contenu, le classement et donc l'indexation de documents d'information (dans de nombreux cas, les thesaurus proposent aussi une définition des termes utilisés).

Une classification est l'action de distribuer par classes par catégories (rien n'est dit sur le type d'objets classifiés). C'est aussi le résultat de cette action.

Une terminologie est un ensemble des termes particuliers à une science, à un art, à un domaine. Les termes y sont également définis par un texte en langue naturelle et caractérisés par différentes propriétés linguistiques ou grammaticales suivant l'usage prévu de cette terminologie. Avec leur mise sur support informatique, les terminologies ont beaucoup évolué et sont parfois enrichies de relations entre termes, formant ainsi un réseau terminologique.

À partir de là, on peut s'intéresser à quelques RTO, par exemple WORDNET ou le MeSH, pour vérifier quelle est leur nature exacte, ce qu'on peut en faire et pourquoi :

WORDNET est une base de données lexicales. Les termes y sont organisés sous formes d'ensembles de synonymes, les *synsets*. Chaque *synset* est un concept lexicalisé (en reprenant les termes de M. Slodzian [41]). Ces concepts lexicalisés sont reliés par des relations conceptuelles (*is-a*, *has-a*). Les concepteurs de WORDNET affirment ainsi construire une ontologie linguistique. Si nous pensons qu'une ontologie a pour but de conceptualiser un domaine et nous le maintenons, alors le processus qui mène à cette ontologie doit clairement en prendre la direction. Ce n'est pas le cas de WORDNET qui jongle allègrement entre relations conceptuelles et termes¹³. Précisons bien la motivation de ces remarques : WORDNET est un énorme dictionnaire hypermédia de l'anglais-américain (plus de 100 000 *synsets*) et sa richesse et sa facilité d'accès en font un intéressant outil pour la recherche d'information ou d'autres tâches comme le traitement du langage naturel mais ce n'est pas une ontologie, cela n'en prend pas le chemin et essayer de l'utiliser tel quel ou avec un minimum de modification dans un système formel est voué à l'échec.

MeSH (*Medical Subject Heading*) est un thesaurus médical. C'est le thesaurus d'indexation de la base bibliographique MEDLINE¹⁴. Il est traduit en français par l'INSERM et sert aussi de thesaurus au site CISMef (*cf.* chap. 8, ce volume). Le MeSH offre une organisation hiérarchique et associative et comprend jusqu'à neuf niveaux de profondeur. Sans revenir sur sa description précise qui montre que l'on est en face d'un thesaurus développé pour l'indexation et non pour les inférences [15], on peut citer les concepteurs qui écrivent :

¹³ M. Slodzian [41, 42] développe très longuement des arguments épistémologiques et linguistiques à ce sujet. Nous renvoyons le lecteur intéressé à ces articles.

¹⁴ Le MeSH est accessible à <http://www.nlm.nih.gov/mesh/meshome.html>

Many individuals have tried to use MeSH as a concept representation language with only modest success. That the relationships in the MeSH tree structure were designed with a different view, and with a different (an not formal) meaning of "broader-than", has frustrated their efforts. The MeSH hierarchical structure was designed to reflect a view of the literature for a user.[...] The trees thus indicate what appears to be a useful set of relationships, based on the perceived needs of searchers [33].

À la suite de ces quelques exemples¹⁵ on peut constater que ces ressources terminologiques ne sont pas des ontologies. Ce n'est pas étonnant : dans de nombreux domaines, des thésaurus ont été développés d'abord pour des besoins de normalisation des termes employés et à un moment où une utilisation informatique¹⁶ n'était pas envisagée. Les ontologies n'étant pas sans rapport avec les terminologies, on peut trouver dans ces thésaurus des ressources pour amorcer une ontologie. Mais il faut prendre garde au fait que ce ne sont que des « ressources pour » et pas des « embryons de ».

2.2.4 La réutilisation et les ontologies génériques

La première et originelle utilité d'une ontologie était liée à une volonté de réutilisation (*cf.* § 2.1.1). Cette propriété étant attendue mais ni démontrée, ni définitoire, elle est restée comme pétition de principe alors que la nature de l'ontologie se précisait au fur et à mesure que des travaux utilisaient le concept. Ainsi, réutilisable ou pas, l'ontologie devait servir de représentation des connaissances du domaine pour un SBC. Plus précisément, on peut dire maintenant qu'elle sert de squelette à la représentation des connaissances du domaine dans la mesure où elle décrit les concepts, leurs propriétés et la façon dont ils peuvent se combiner pour constituer des connaissances du domaine complètes. Cette caractéristique est valable pour tout SBC, qu'il soit isolé ou doive servir à faire des inférences dans le contexte du Web sémantique.

Des questions épistémologiques sur la réutilisabilité et incidemment la nature des ontologies ont alimenté de longs débats [16, 24, 50]. La communauté française est particulièrement active dans ce domaine, en

¹⁵ Voir [15] pour un panorama plus large en notant que nous nous sommes intéressés sciemment à des ressources au statut problématique. Il y a évidemment d'autres ressources dont le statut ontologique est incontestable comme la GeneOntology <<http://www.geneontology.org>>.

¹⁶ Utilisation, au sens fort, c'est-à-dire pour faire des inférences.

particulier au niveau du groupe TIA, au sein duquel ontologies, terminologies, bases de connaissances terminologiques, entre autres « produits » de la recherche, sont caractérisés par rapport aux besoins de la Terminologie et de l'Ingénierie des connaissances. Des nombreuses expériences menées au sein du groupe, il ressort que la réutilisation simple des ontologies espérée au début des années 90 a disparu et que des réflexions et outils nouveaux sont apparus nécessaires pour aborder la question de la construction des ontologies et de leur réutilisation (*cf.* § 2.2.2). L'autre conséquence est la nécessité d'assumer une certaine non réutilisabilité des ontologies : en effet, les ontologies sont des artefacts construits en fonction d'une tâche précise et ne peuvent être réutilisées, en tant qu'objet formel, pour une autre tâche.

Ce constat fait, il en découle un certain nombre de recherches sur la réutilisabilité du « haut » des ontologies dont l'argumentaire est le suivant : puisqu'il est difficile, voire impossible, de réutiliser directement des ontologies, trop proches de vues détaillées qu'on peut avoir sur un domaine, intéressons-nous au « haut » de l'ontologie qui répertorie et organise de grandes catégories de la pensée ou de la société humaine qui devraient pouvoir être réutilisables dans de très nombreuses applications et être alors « génériques ». C'est l'objectif du groupe SUO¹⁷ (*Standard Upper Ontology*) qui réfléchit à la constitution d'un haut d'ontologie, *the SUMO (Suggested Upper Merged Ontology)*, qui se voudrait universel pour les grandes catégories d'objets et de pensées. Les discussions intenses au sein de ce groupe de réflexion semblent montrer que le but ne sera pas atteint tout de suite. Mais sera-t-il atteint un jour ? Avec les mêmes arguments épistémologiques que précédemment sur la non réutilisabilité des ontologie, on peut penser qu'on ne sera pas capable de construire ce type d'ontologie universelle.

Ajouté au fait qu'il est difficile de mettre d'accord de nombreuses personnes ou institutions sur des conceptualisations partagées, une conséquence des réflexions précédentes et qui semble inéluctable pour le Web sémantique est que les différentes communautés de travail ou de pratiques vont développer et mettre à disposition des ontologies portant sur des domaines restreints avec des ontologies différentes portant sur des champs de conceptualisation identiques. Les questions autour principalement de la comparaison, de la fusion, etc. des ontologies vont alors devenir cruciales (*cf.* § 3.4).

¹⁷ <http://suo.ieee.org>

3. TRAVAUX ET RÉSULTATS EXISTANTS DU WEB SÉMANTIQUE

3.1 Le Web sémantique au sein des recherches en RC et IC

Les travaux et recherches décrits jusque-là ont commencé antérieurement à la diffusion des recherches sur le Web sémantique. C'est la raison pour laquelle ils sont situés dans les acquis même si des recherches sont encore nécessaires. Ceci étant, les recherches sur le Web sémantique ont pris leur essor avec l'implication de chercheurs en Représentation des connaissances, autour des langages et systèmes d'inférences et avec les chercheurs de l'Ingénierie des connaissances sur la question des méthodologies d'élaboration des ontologies. C'est pourquoi, à partir de là, nous ne chercherons pas à préciser l'origine de ces recherches et considérerons que ce sont des acquis du Web sémantique.

Un des résultats forts du Web sémantique sur les ontologies est la normalisation de leur expression. Ce point, indispensable si l'on veut que les ontologies puissent être partagées, semble justement trouver une solution dans le contexte du Web sémantique : la définition du langage OWL (*Ontologies Web Language*) à différents niveaux de complexité (capacité de complexité des descriptions *versus* calculabilité) en est le meilleur exemple. Cette question et d'autres sur les langages sont développées au chapitre 2 (ce volume).

Après avoir abordé, à la section précédente, des problématiques que nous considérons fondamentales et qui préexistaient aux recherches sur le Web sémantique, nous allons poursuivre cette réflexion méthodologique avec des considérations plus particulièrement issues de ces recherches.

3.2. Des méthodologies de construction d'ontologies

3.2.1. De bons principes méthodologiques

Passées les questions de méthodologies fondamentales, un certain nombre de travaux proposent des principes de construction d'ontologies. Nous allons nous attarder sur deux d'entre eux, paradigmatiques, les travaux de Th. R. Gruber [23] et ceux de M. Fernández *et al.* [19].

Th. R. Gruber propose ainsi un certain nombre de principes à respecter pour construire une ontologie :

Clarté. Les ambiguïtés doivent être réduites. Quand une définition peut être axiomatisée, elle doit l'être. Dans tous les cas, des définitions en langage naturel doivent être fournies.

Cohérence. Une ontologie doit être cohérente. Les axiomes doivent être consistants. La cohérence des définitions en langage naturel doit être vérifiée autant que faire se peut.

Extensibilité. L'ontologie doit être construite de telle manière que l'on puisse l'étendre facilement, sans remettre en cause ce qui a déjà été fait.

Biais d'encodage minimal. L'ontologie doit être conceptualisée indépendamment de tout langage d'implémentation. Le but est de permettre le partage des connaissances (de l'ontologie) entre différentes applications utilisant des langages de représentation différents.

Engagement ontologique minimal. Une ontologie doit faire un minimum d'hypothèses sur le monde : elle doit contenir un vocabulaire partagé mais ne doit pas être une base de connaissances comportant des connaissances supplémentaires sur le monde à modéliser.

D'autres principes du même type sont proposés par d'autres auteurs (*cf.* article de A. Gómez-Pérez [22]). Il est facile de voir que la méthodologie que nous avons décrite (*cf.* § 2.2) fournit des moyens d'appliquer ces principes, au niveau sémantique ou au niveau formel selon les cas.

M. Fernández *et al.* proposent, dans la méthodologie METHONTOLOGY, de construire une ontologie en respectant des activités de gestion de projet (planification, assurance qualité), de développement (spécification, conceptualisation, formalisation, implémentation, maintenance) et des activités de support (intégration, évaluation, documentation). On retrouve là des problématiques de génie logiciel et de gestion de projet informatique qu'on a tout intérêt à voir s'appliquer à la construction de grandes ontologies, si on a une méthodologie réelle de construction, évidemment. Cette méthodologie rejoint pour partie celle décrite par F. Gandon [20] et développée au sein de l'équipe ACACIA de l'INRIA.

Enfin, nous renvoyons à l'article de A. Gómez-Pérez [22], pour une description plus complète d'ontologies et de méthodologies centrées sur des problématiques de cycle de vie des ontologies.

3.2.2. Acquérir une ontologie à partir de DTD

Dans le cadre du Web sémantique, les systèmes médiateurs permettent d'établir une connexion entre différentes sources d'information sur un même domaine. Ils utilisent une ontologie qui sert de schéma global intermédiaire dans lequel sont exprimées les requêtes qui peuvent être traduites dans la représentation de chaque source d'information. Si ces sources d'information sont été développées avant le médiateur, ce qui est le cas des organisations relatives à l'industrie du voyage, on peut disposer des ontologies des sources (souvent sous forme de DTD XML) pour essayer de construire l'ontologie du médiateur. C'est ce type de travail qui a été expérimenté dans le projet PICSEL et qui est abordé au chapitre 5, section 3 [21].

Il est important de noter que le Web sémantique ne se construira que si les personnes et les institutions au sens large partagent une même sémantique d'un domaine. Dans un domaine comme le voyage, décrit ici, il y a un existant dont il faut tenir compte et on ne peut créer une ontologie ex-nihilo. Des approches à partir de DTD, si elles n'aboutissent pas encore à des méthodologies stables permettent de tenir compte de l'existant et sont ainsi indispensables à l'avènement et au fonctionnement du Web sémantique (*cf.* § 3.4).

3.3. Des éditeurs d'ontologies

3.3.1. Introduction

De nombreux outils permettent aujourd'hui d'éditer des ontologies. Parmi ceux-ci, quelques uns essaient de guider leur utilisateur dans l'élaboration de l'ontologie en suivant une méthodologie de conception plus ou moins complète, que ce soit en respectant des principes de cycle de vie et validation logiciels (*cf.* § 3.2.1), d'un côté, ou, de l'autre côté, en outillant une réflexion épistémologique (*cf.* § 2.2.2). Dans tous les cas, force est de constater qu'aucun de ces outils n'a réussi à s'imposer¹⁸ et la

¹⁸ A l'exception de *Protégé*, si l'on considère ce dernier non pas comme un simple environnement de construction d'ontologies basé sur les *frames*, mais comme un *framework* logiciel ouvert capable d'intégrer de multiples *plugins* implémentant les nombreuses recherches effectuées autour des ontologies (*cf.* § 3.3.2).

réflexion sur l'outillage de la construction des ontologies reste donc ouverte. Les outils proposés peuvent se regrouper grossièrement en deux catégories. Dans la première, on trouve les plus anciens historiquement, qui permettent de spécifier les ontologies au niveau symbolique (voir, par exemple, le serveur ONTOLINGUA¹⁹ [17]) : une grande partie des définitions des objets se fait directement dans un langage de représentation de connaissances donné (pour ONTOLINGUA, il s'agit de KIF), auquel le créateur et l'utilisateur de l'ontologie doivent se plier. Dans la seconde catégorie, les outils prennent mieux en compte l'importance du niveau des connaissances : ils proposent à leur utilisateur de créer l'ontologie de manière relativement indépendante de tout langage implémenté et prennent ensuite automatiquement en charge l'opérationnalisation de l'ontologie, en la transposant dans divers langages. Cette évolution tend à rapprocher les ontologies de leur but original : il semble en effet naturel de chercher à s'abstraire – dans un premier temps – du niveau symbolique si on veut obtenir une ontologie permettant un réel partage d'une compréhension. Cette dernière catégorie regroupe les outils principalement utilisés aujourd'hui.

3.3.2. PROTÉGÉ-2000

PROTEGE-2000²⁰ [35] est un environnement graphique de développement d'ontologies développé par le SMI de Stanford. Dans le modèle des connaissances de PROTEGE, les ontologies consistent en une hiérarchie de *classes* qui ont des attributs (*slots*), qui peuvent eux-mêmes avoir certaines propriétés (*facets*). L'édition des listes de ces trois types d'objets se fait par l'intermédiaire de l'interface graphique, sans avoir besoin d'exprimer ce que l'on a à spécifier dans un langage formel : il suffit juste de remplir les différents *formulaires* correspondant à ce que l'on veut spécifier. Ce modèle autorise d'ailleurs une liberté de conception assez importante puisque le contenu des formulaires à remplir peut être modifié suivant les besoins *via* un système de *métaclasses*, qui constituent des sortes de « patrons » de connaissance. L'interface, très bien conçue, et l'architecture logicielle permettant l'insertion de *plugins* pouvant apporter de nouvelles fonctionnalités ont participé au succès de PROTÉGÉ-2000. Aujourd'hui, il regroupe une large communauté d'utilisateurs et bénéficie des toutes dernières avancées en matière de recherche ontologique : compatibilité OWL de référence, services inférentiels, gestion de bases de connaissances, visualisation d'ontologies, alignement et fusion, etc.

¹⁹ <http://www-ksl-svc.stanford.edu:5915/>

²⁰ <http://protege.stanford.edu/index.shtml>

3.3.3. OIEd

OIEd²¹ [7], développé sous la responsabilité de l'université de Manchester, a été conçu pour éditer des ontologies dans le langage de représentation OIL, un des précurseurs du langage OWL²² (*Ontology Web Language*) qui est aujourd'hui en voie d'être une recommandation W3C (*cf.* chap. 2, ce volume). Officiellement, il n'a pas d'autre ambition que de construire des exemples montrant les vertus du langage pour lequel il a été créé. A ce titre, OIEd est souvent considéré comme une simple interface de la logique de description SHIQ. Néanmoins, il offre la plus grande partie de ce que l'on peut attendre d'un éditeur d'ontologies. On peut créer des hiérarchies de classes et spécialiser les rôles, et utiliser avec l'interface les types d'axiomes les plus courants. Cet éditeur offre également les services d'un raisonneur, FaCT, qui permet de tester la satisfaisabilité des définitions de classes et de découvrir des subsomptions restées implicites dans l'ontologie.

3.3.4. OntoEdit

Contrairement aux deux outils précédents, ONTOEDIT [43] n'est pas disponible gratuitement dans sa version complète²³. Il présente les fonctionnalités essentielles communes aux autres éditeurs (hiérarchie de concepts, expression d'axiomes, export de l'ontologie dans des langages divers) et a le mérite de s'appuyer sur une réflexion méthodologique significative. La modélisation des axiomes a fait l'attention de soins particuliers pour pouvoir être effectuée – en tout cas pour les types les plus répandus – indépendamment d'un formalisme privilégié et cela pour faciliter la traduction d'un langage de représentation à un autre. Il propose également une gestion originale des *questionnaires de compétences*. Des questions pour les réponses desquelles l'ontologie doit fournir le matériel conceptuel, on peut extraire les termes appelés à intégrer l'ontologie. Un petit outil fait une comparaison lexicale des termes extraits des différentes questions pour en déduire automatiquement d'éventuelles subsomptions. Le procédé semble cependant loin d'être fiable puisqu'il repose sur l'hypothèse que le nom d'un concept se retrouve parfois dans le nom de ses spécialisations.

3.3.5. WebODE

²¹ <http://oiled.man.ac.uk/>

²² <http://www.w3.org/2001/sw/WebOnt/>

²³ Une version de démonstration est disponible sur le site d'Ontoprise, la société qui le développe en collaboration avec l'AIFB de Karlsruhe

WebODE²⁴ [1], développé par le LAI de Madrid, est une plate-forme de conception d'ontologies fonctionnant en ligne. D'un point de vue méthodologique, l'outil fait suite à ODE, un éditeur qui assurait fidèlement le support de la méthodologie maison METHONTOLOGY (cf. § 3.2.1). Il illustre bien l'évolution des outils de construction d'ontologies, puisque les nombreuses tables de son prédécesseur ont été remplacées par une interface très travaillée, réalisant un pas supplémentaire vers une conception au niveau des connaissances. On peut cependant regretter que cette évolution se soit faite au détriment de l'application des contraintes méthodologiques : les *représentations intermédiaires* utilisées dans le processus de conception sont désormais moins mises en avant, à tel point que le guide de l'utilisateur ne les signale que pour « assurer la compatibilité conceptuelle avec ODE ». L'accent a plus été mis sur la possibilité d'un travail collaboratif ou sur la mise à disposition d'outils complémentaires, comme un moteur d'inférences.

3.3.6. DOE

Le dernier outil présenté ici est DOE pour *Differential Ontology Editor*²⁵ [5, 47]. Cet outil n'a pas pour ambition de concurrencer les grands environnements existants, mais plutôt de fournir un début d'implémentation à la méthodologie de structuration différentielle proposée par B. Bachimont (cf. § 2.2.2). A l'instar des autres éditeurs, il offre une représentation graphique des arbres de concepts et des relations de l'ontologie et permet d'interagir avec les hiérarchies. L'outil assiste également la saisie des principes différentiels issus de la méthodologie en automatisant partiellement cette tâche. Le modèle de représentation de l'ontologie est finalement proche de celui du langage RDFS, à ceci près qu'il autorise la modélisation de relations n-aires. Au niveau formel, l'éditeur est capable de faire quelques inférences en vérifiant la consistance de l'ontologie (propagation de l'arité le long de la hiérarchie des relations et héritage des domaines par exemple).

3.3.7. Des outils à parfaire

Tous ces outils sont capables aujourd'hui de produire des ontologies dans les langages OWL et RDFS standardisés par le W3C. Mais comme le montrent les expérimentations effectuées dans le cadre des Workshops EON [44, 45], les ontologies générées sont loin d'être interoperables

²⁴ <http://delicias.dia.fi.upm.es/webODE/>

²⁵ L'outil est disponible gratuitement à <http://opales.ina.fr/public/>

malgré un langage commun. Le support de ces outils vis-à-vis du langage OWL est ainsi très hétérogène puisque les principaux APIs et parsers (e.g. Jena ou OWL-API) ne semblent pas manipuler le langage de la même façon. Il en va de même pour les langages n3, RDFS ou DAML+OIL. On peut toutefois espérer que ces produits vont s'harmoniser très prochainement. Enfin, et c'est l'objet de la prochaine section, ils n'ont pas toutes les fonctionnalités, de plus en plus nombreuses au fur et à mesure que les recherches se précisent, que réclame le développement distribué du Web sémantique.

3.4. Comparaison et versions des ontologies

Nous avons vu précédemment les problèmes fondamentaux liés à la construction d'ontologies réutilisables ou universelles (*cf.* § 2.2.4). La conséquence de cette situation est de solliciter des réflexions sur la façon de comparer et fusionner des ontologies. En effet, même si on admet que les ontologies pour le Web sémantique vont être partagées, au moins dans un premier temps, par des communautés d'intérêts restreintes, les ontologies déjà développées pour un même domaine sont multiples, par exemple dans le cas des voyages (*cf.* § 3.2.2) ou de la médecine. De plus, on peut très facilement imaginer que les ontologies vont être développées dans des environnements distribués et décentralisés qui vont requérir des fonctionnalités spécifiques [29].

Les travaux autour de la question de la comparaison/fusion visent donc à prendre en charge d'abord la tâche de comparaison de concepts au sein d'une ontologie : c'est, en effet, via cette comparaison que l'on pourra décider de la fusion. L'apparition d'environnements décentralisés connectés nous interrogent eux sur la gestion des versions des ontologies au sens large. Sur ces questions, un rapport du projet KnowledgeWEB²⁶ fait le point sur le fonctionnement et l'intégration d'un certain nombre d'outils et nous a servi de point d'entrée pour cette analyse [37] mais en plus d'y renvoyer le lecteur nous voulons discuter les aspects sémantiques et techniques importants de ces questions :

Comparaison des concepts dans la structure ontologique globale.

La comparaison des ontologies pour partage ou fusion n'est possible que si les langages de représentation sont identiques. Leur puissance d'expression permet alors des comparaisons logiques, sur les propriétés, etc. très fructueuses mais le point crucial reste le sens réel des concepts : 2 concepts A et B , respectivement dans 2

²⁶ <http://knowledgeweb.semanticweb.org>

ontologies O_1 et O_2 , qui semblent semblables correspondent-ils réellement au même concept ? Le nom du concept, son label dans l'ontologie ne suffit pas, *surtout pas*, à répondre : en effet, ce n'est pas le label du concept qui fait sa signification dans une ontologie, c'est sa place dans l'arborescence (*cf.* § 2.2.2). Si ce label nous aide à intuiter la signification du concept, il n'est porteur de rien vis-à-vis de la représentation et peut même laisser croire à des identités qui n'en sont pas [14]. Dans ce contexte, certains travaux comme ceux de Noy et Musen [36] sont intéressants : ils partent de paires de concepts qui semblent proches (découverts de façon automatique ou proposés manuellement) et calculent leur similarité « hors contexte » en étudiant les chemins qui relient les paires de concepts. Le « hors contexte » vient de ce que les paires de concepts étudiées (les *ancres* de la recherche) peuvent être éloignées. Ces réflexions sont implémentées dans Anchor-PROMPT et, à notre avis, prometteuses pour une comparaison sémantique des ontologies. Très récemment, une compétition²⁷ visant à évaluer les différents algorithmes d'alignement d'ontologies existants a d'ailleurs confirmé les bonnes performances de cet outil. Pour un état de l'art complet des différentes techniques d'alignement et de fusion d'ontologie, nous renvoyons le lecteur à [30].

Gestion des versions. La gestion des versions est une tâche indispensable de l'ingénierie ontologique au sein des environnements distribués (*cf. supra*). La plupart des éditeurs d'ontologies (*cf.* § 3.3) ont une fonction de *log* qui offre une trace des changements effectués sur l'ontologie éditée mais elle est insuffisante dans de tels environnements : en effet, il ne suffit pas d'avoir une simple trace des changements effectués dans une ontologie, il faut pouvoir prendre en charge un certain nombre de fonctionnalités répondant au questionnement suivant [29] :

Mise à jour d'ontologies. Les évolutions d'une ontologie sont-elles conceptuelles ou sont-elles des évolutions de la spécification, c'est-à-dire de la théorie logique sous-jacente ? Il est clair que ces 2 caractéristiques peuvent évoluer différemment et que la spécification (le formel) ne suffit pas à décrire une ontologie. Quel est alors le bon niveau de description d'une ontologie en vue de gérer ses évolutions [28] ?

²⁷ The *EON Ontology Alignment Contest*,
<http://co4.inrialpes.fr/align/Contest/>

Mise à jour d'ontologies distribuées. Si on adapte une ontologie distante à un usage local et que cette ontologie distante évolue, il faut pouvoir (si nécessaire) faire évoluer l'ontologie locale. À la suite de la question précédente, peut-on construire une grammaire des changements à prendre en compte ? Peut-on construire une ontologie des changements [28] ?

Vérification et validation d'une ontologie. Si plusieurs personnes ou équipes interviennent pour faire évoluer des ontologies, il faut pouvoir valider ou pas les changements et faire marche arrière si nécessaire. Cela pose des questions du même type que précédemment.

Consistance. Si une ontologie évolue, quel est le statut de cette évolution, formellement parlant ? La théorie logique sous-jacente peut évoluer. Les axiomes valables dans l'ancienne version, sont-ils toujours valables dans la nouvelle ? Les données interprétées dans l'ancienne version sont-elles toujours correctement interprétées dans la nouvelle ?

Ces questions sont abordées et obtiennent des propositions de réponses dans les références citées ici et d'autres au sein de la communauté Web sémantique.

4. RECHERCHES FUTURES POUR LE WEB SÉMANTIQUE

4.1. Vers une méthodologie et des outils intégrés

Les propositions méthodologiques faites dans la section 2.2.2 n'abordent pas toutes les conséquences logiques de l'*engagement ontologique*, pourtant nécessaire à toute formalisation. Si l'on veut développer une méthodologie qui prenne en compte toutes les étapes de la construction d'ontologie, il est alors nécessaire d'étudier les liens qu'entretiennent la normalisation sémantique et l'*engagement ontologique* de N. Guarino²⁸.

²⁸ La question a été abordée dans [4] qui constate que la formalisation est une spécification formelle donc extensionnelle de l'ontologie ainsi définie et que le sens des concepts est alors dans les objets définis en extension. Ainsi, N. Guarino a abordé des problèmes analogues à la normalisation sémantique, en proposant, au niveau formel [25], une méthode à base de métapropriétés (*identité, rigidité, anti-rigidité, dépendance*) qui contraignent l'organisation

Par ailleurs, il ne se dégage pas de consensus sur les méthodologies de construction d'ontologies. Cela est dû en partie à la diversité des motivations de construction des ontologies et à la complexité des domaines. Une ontologie d'indexation d'équipes de recherche comme exemplifiée dans le projet (KA)² [8] est beaucoup plus simple et appréhendable qu'une ontologie de la réanimation chirurgicale destinée à rendre des services terminologiques [31]. Ce sont pourtant deux facettes des problèmes d'indexation dans le contexte du Web sémantique. La conséquence d'une telle disparité se retrouve au sein des nombreux outils, les éditeurs d'ontologies, développés. Ces outils ont besoin d'être utilisés dans de nombreux projets pour tester leur utilisabilité et réfléchir sur les fonctionnalités nécessaires à l'édition d'ontologies du début à la fin du processus. Nous verrons, dans les cas les plus complexes, que d'autres solutions doivent être envisagées que l'élaboration et l'utilisation d'une « simple » ontologie (*cf.* § 4.3).

4.2. Comparaison et fusion des ontologies

Si les travaux sur les ontologies ont une dizaine d'années, ceux sur la fusion ou la gestion des versions sont évidemment encore plus récents (*cf.* § 3.4) et leur poursuite est indispensable pour que les ontologies servent le Web sémantique : que ce soit dans la dimension syntaxique, où les travaux du Web sémantique sont bien avancés même s'ils soulèvent de nombreuses questions de représentation (*cf.* chap. 2, ce volume), ou dans la dimension logique où l'utilisation d'une ontologie doit pouvoir être caractérisée et assurée durant l'évolution de celle-ci. Plus important encore, la dimension sémantique des ontologies doit être prise en compte dans les comparaisons/fusions, c'est-à-dire que les ontologies doivent pouvoir être comparées pour la signification des concepts et pas seulement, même si c'est nécessaire, pour leur caractéristique logique. Cette prise en compte est ainsi cruciale dans de nombreux points du Web sémantique, en particulier, dans le contexte des systèmes de médiations centralisés (*cf.* chap. 5, § 4.1, ce volume). Nous renvoyons le lecteur à ce dernier chapitre pour approfondir les nécessités autour des ontologies en ce domaine.

4.3. Cacher la complexité

d'une ontologie formelle, en particulier la relation de subsomption, en fonction des propriétés des nœuds qui la composent. Cette convergence de vues peut être intéressante à exploiter dans la recherche d'une méthodologie intégrée.

Un des aspects importants, en vue de l'adoption du Web sémantique, est la capacité d'aider des utilisateurs non spécialistes à créer et à exploiter des ressources exploitables dans le cadre des infrastructures et des outils proposés par le Web sémantique. En d'autres termes, comment cacher la complexité des technologies du Web sémantique ?

Avant de répondre à cette dernière question, il est nécessaire de prendre conscience que le niveau de conceptualisation des ontologies n'est pas toujours accessible aux utilisateurs, même spécialistes du domaine. Ainsi, les ontologies en médecine sont inaccessibles à d'autres que leurs créateurs : d'abord parce qu'elles représentent un niveau de conceptualisation utile pour la tâche qui est demandé au SBC mais inutile à l'expression des concepts médicaux dans la pratique médicale, ensuite parce qu'en médecine comme dans d'autres domaines, l'expression des connaissances se fait en langue naturelle et que cette expression est normée par des thesaurus développés et utilisés depuis de très nombreuses années (*cf.* § 2.2.3). Ainsi, l'accès aux ontologies par des utilisateurs, mêmes professionnels, nécessite de gérer le lien entre les concepts des ontologies et les termes du langage naturel, que ce soit pour une simple compréhension ou pour l'indexation et la construction de requêtes destinées à des tâches de recherche d'information. Les solutions mises en œuvre à ce jour passent par deux types de propositions :

- des méthodologies séparant explicitement les termes et les concepts d'un domaine et c'est ce qui est expérimenté en médecine dans des serveurs de terminologie [39] ou dans d'autres domaines dans ce que les chercheurs appellent des thesaurus sémantiques [40] ;
- des méthodologies s'écartant de la formalisation des ontologies et recherchant des proximités conceptuelles dans les termes d'un domaine permettant d'en appréhender intuitivement la complexité [12] ;

Cette dernière approche interroge directement les chercheurs du Web sémantique : saurons-nous mettre en place des ressources formelles, des ontologies, pour le Web sémantique ou devons-nous en passer par des ressources moins formelles mais appréhendables par les utilisateurs ? Ces deux possibilités n'étant d'ailleurs pas exclusives.

5. RÉFÉRENCES

- [1] ARPÍREZ J., CORCHO O., FERNÁNDEZ-LÓPEZ M. & GÓMEZ-PÉREZ A. (2001). WebODE : a Workbench for Ontological Engineering. In *First international*

Conference on Knowledge Capture (K-CAP'01), p. 6–13, Victoria, Canada: ACM.

- [2] AUSSENAC-GILLES N., BIEBOW B. & SZULMAN S. (2003). D'une méthode à un guide pratique de modélisation des connaissances à partir de textes. In *Actes des 5^{es} journées Terminologie et Intelligence Artificielle*, p. 41–53, Strasbourg.
- [3] BACHIMONT B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In J. CHARLET, M. ZACKLAD, G. KASSEL & D. BOURIGAULT, Eds., *Ingénierie des connaissances : évolutions récentes et nouveaux défi*, chapter 19. Paris: Eyrolles.
- [4] BACHIMONT B. (2001). Modélisation linguistique et modélisation logique des ontologies : l'apport de l'ontologie formelle. In J. CHARLET, Ed., *Actes des 5^{es} Journées Ingénierie des Connaissances*, p. 349–68, Grenoble, France.
- [5] BACHIMONT B., ISAAC A. & TRONCY R. (2002). Semantic Commitment for Designing Ontologies: A Proposal. In A. GOMEZ-PÉREZ & V. BENJAMINS, Eds., *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, volume (2473) of *Lecture Notes in Artificial Intelligence*, p. 114–121, Sigüenza, Espagne: Springer Verlag.
- [6] BACHIMONT B. (2004). *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Habilitation à diriger des recherches, Université de Technologie de Compiègne.
- [7] BECHHOFFER S., HORROCKS I., GOBLE C. & STEVENS R. (2001). OilEd: a Reason-able Ontology Editor for the Semantic Web. In *Joint German/Austrian conference on Artificial Intelligence (KI'01)*, volume (2174) of *Lecture Notes in Artificial Intelligence*, p. 396–408, Vienne, Autriche: Springer Verlag.
- [8] BENJAMINS V. R. ., FENSEL D. P. A. (1998). Knowledge management through ontologies. In *Proceedings of the 2nd Conf. On Practical Aspects of Knowledge Management*, Basel, Switzerland.
- [9] BOURIGAULT D. (2002). Analyse distributionnelle étendue. In *Actes de la 9^e conférence sur le traitement automatique des langues*, Nancy.
- [10] BOURIGAULT D., AUSSENAC-GILLES N. & 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* **18**(1), numéro spécial *Techniques informatiques et structuration de terminologies*.
- [11] BOURIGAULT D. & FABRE C. (2000). Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de Grammaires*, (25), 131–51. numéro spécial « sémantique et corpus ».
- [12] BOURIGAULT D. & LAME G. (2002). Analyse distributionnelle et structuration de terminologie. application à la construction d'une ontologie documentaire du droit. *Traitement automatique des langues*, **43**(1).

- [13] BRACHMAN R. (1983). What IS-A Is and Isn't: An analysis of taxonomic links in semantic networks. *IEEE Computer*, **16**(10), 30–6.
- [14] BRACHMAN R. J., MCGUINNESS D. L., PATEL-SCHNEIDER P. F., RESNIK L. A. & BORGIDA A. (1991). Living with Classic: When and how to use a KL-ONE-like language. In J. F. SOWA, Ed., *Principles of Semantic Networks*, chapter 14, p. 401–456. San Mateo, Ca. Morgan Kaufmann Publishers.
- [15] CHARLET J. (2002). *L'Ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales*. Habilitation à diriger des recherches, Université Paris 6.
- [16] CHARLET J., BACHIMONT B., BOUAUD J. & ZWEIGENBAUM P. (1996). Ontologie et réutilisabilité : expérience et discussion. In N. AUSSENAC-GILLES, P. LAUBLET & C. REYNAUD, Eds., *Acquisition et ingénierie des connaissances : tendances actuelles*, chapter 4, p. 69–87. Cepaduès-éditions.
- [17] FARQUAR A., FIKES R., PRATT W. & RICE J. (1995). *Collaborative Ontology Construction for Information Integration*. Rapport de recherche KSL-95-63, Knowledge Systems Laboratory, Department of Computer Science.
- [18] FAURE D. & NEDELLEC C. (1999). Knowledge acquisition of predicate argument structures from technical texts using machine learning: The system ASIUM. In *Proc. of the 11th European Workshop, Knowledge Acquisition, Modelling and Management*, number 1937 in LNAI, p. 329–334, Juan-les-Pins, France: Springer-Verlag.
- [19] FERNÁNDEZ M., GÓMEZ-PÉREZ A., PAZOS J. & PAZOS A. (1999). Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent System and their Applications*, **14**(1), 37–45.
- [20] GANDON F. (2002). *Ontology Engineering : a Survey and a Return on Experience*. Rapport interne 4396, INRIA. 181 p., ISSN 0249-6399.
- [21] GIRALDO G. & REYNAUD C. (2002). Construction semi-automatique d'ontologies à partir de DTDs relatives à un même domaine. In B. BACHIMONT, Ed., *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 53–61, Rouen, France.
- [22] GOMEZ-PEREZ A. (2000). Développements récents en matière de conception, de maintenance et d'utilisation d'ontologies. *Terminologies Nouvelles*, (19), 9–20. Traduit de l'anglais par S. Descotte.
- [23] GRUBER T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, **5**, 199–220.
- [24] GUARINO N. (1997). Understanding, building, and using ontologies. *International Journal of HumanComputer Studies*, **45**(2/3), 293–310.
- [25] GUARINO N. (1999). The role of identity conditions in ontology design. In V. BENJAMINS, B. CHANDRASEKARAN, A. GOMEZ-PEREZ, N. GUARINO & M. USCHOLD, Eds., *Proc. of the IJCA'99 Workshop on Ontologies and Problem-Solving Methods*, p. 2/1–2/7, Sweden.

- [26] GUARINO N. & GIARETTA P. (1995). Ontologies and knowledge bases. In *Towards Very Large Knowledge Bases*. Amsterdam: IOS Press.
- [27] GUARINO N. & WELTY C. (2000). A Formal Ontology of Properties. In R. DIENG & O. CORBY, Eds., *12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'00)*, volume (1937) of *Lecture Notes in Computer Science*, p. 97-112, Juan-les-Pins, France: Springer Verlag.
- [28] KLEIN M., KIRYAKOV A., OGNYANOFF D. & FENSEL D. (2002). Finding and specifying relations between ontology versions. In *Proceedings of the workshop on Ontologies and Semantic Interoperability at the 15th ECAI*, Lyon, France. Accessible à <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-64/>.
- [29] KLEIN M. & NOY N. F. (2003). A component-based framework for ontology evolution. In *Proceedings of the workshop on Ontologies and Distributed Systems at the 9th IJCAI*, Acapulco, Mexico.
- [30] KNOWLEDGEWEB C. (2004). *State of the Art on Ontology Alignment*. Deliverable 2.2.3, FP6-507482.
- [31] LE MOIGNO S., CHARLET J., BOURIGAUULT D. & JAULENT M.-C. (2002). Construction d'une ontologie à partir de corpus : expérimentation et validation dans le domaine de la réanimation chirurgicale. In B. BACHIMONT, Ed., *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 229-38, Rouen, France.
- [32] MAEDCHE A. & STAAB S. (2000). Mining ontologies from text. In R. DIENG & O. CORBY, Eds., *Proc. of the 12th European Workshop, Knowledge Engineering and Knowledge management: methods, models and Tools*.
- [33] NELSON S. J., JOHNSTON D. & HUMPHREYS B. L. (2001). Relationships in medical subject headings. In C. A. BEAN & R. GREEN, Eds., *Relationships in the organization of knowledge*, New York: Kluwer Academic Publishers.
- [34] NEWELL A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87-127.
- [35] NOY N., FERGERSON R. & MUSEN M. (2000). The knowledge model of Protégé2000 : Combining interoperability and flexibility. In R. DIENG & O. CORBY, Eds., *12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'00)*, volume (1937) of *Lecture Notes in Computer Science*, p. 17-32, Juan-les-Pins, France: Springer Verlag.
- [36] NOY N. F. & MUSEN M. A. (2001). Using non-local context for semantic matching. In *Proceedings of the workshop on Ontologies and Information Sharing at the 17th IJCAI*, Seattle.
- [37] ONTOWEB C. (2002). *Ontology-based information exchange for knowledge management and electronic commerce*. Deliverable 1.3, IST Project IST-2000-29243 OntoWeb.

- [38] REBEYROLLES J. (2000). Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires. *Cahiers de Grammaires*, (25), 153–74. numéro spécial «sémantique et corpus ».
- [39] RECTOR A. L. (1998). Thesauri and formal classifications: Terminologies for people and machines. *Methods of Information in Medicine*, **37**(4–5), 501–509.
- [40] ROUSSEY C., CALABRETTO S. & PINON J.-M. (2002). Le thésaurus sémantique : contribution à l'ingénierie des connaissances documentaires. In B. BACHIMONT, Ed., *Actes des 6^{es} Journées Ingénierie des Connaissances*, p. 209–20, Rouen, France.
- [41] SLODZIAN M. (1999). WordNet et EuroWordNet : questions impertinentes sur leur pertinence linguistique. *Sémiotiques*, (17), 51–70. Numéro spécial *Dépasser les sens iniques dans l'accès automatisé aux textes*, coordonné par B. Habert.
- [42] SLODZIAN M. (2000). Wordnet: what about its linguistic relevancy? In R. DIENG, Ed., *Proc. of the EKAW conference*, Juan-les-Pins, France.
- [43] SURE Y., ERDMANN M., ANGELE J., STAAB S., STUDER R. & WENKE D. (2002). OntoEdit: Collaborative Ontology Engineering for the Semantic Web. In I. HORROCKS & J. HENDLER, Eds., *First International Semantic Web Conference (ISWC'02)*, volume (2342) of *Lecture Notes in Computer Science*, p. 221–235, Chia, Sardaigne, Italie: Springer Verlag.
- [44] Y. SURE & O. Corcho, Eds. (2003). *Second International Workshop Evaluation of Ontology-based Tools (EON'03)*, volume (87) of *CEUR-WS*, Sanibel Island, Floride. <http://CEUR-WS.org/Vol-87/>.
- [45] Y. SURE, O. CORCHO, J. EUZENAT & T. HUGHES, Eds. (2004). *Third International Workshop Evaluation of Ontology-based Tools (EON'04)*, Hiroshima, Japon.
- [46] TRONCY R. (2003). Le raisonnement dans les descriptions documentaires : l'apport de la représentation des connaissances. In R. DIENG-KUNTZ, Ed., *Actes des 7^{es} Journées Ingénierie des Connaissances*, Laval, France: Presses universitaires de Grenoble.
- [47] TRONCY R. & ISAAC A. (2002). DOE : une mise en œuvre d'une méthode de structuration différentielle pour les ontologies. In *13^{es} Journées Francophones d'Ingénierie des Connaissances (IC'02)*, p. 63–74, Rouen, France.
- [48] TRONCY R. (2004). *Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies : application à l'audiovisuel*. Thèse de doctorat, Université Joseph Fourier, Grenoble.
- [49] USCHOLD M. & GRUNINGER M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review*.

- [50] VA N HEIJST G., SCHREIBER A. T. & WIELINGA B. J. (1997). Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, **45**(2/3), 183–292.
- [51] WEIS J.-C. & CHARLET J. (2003). Construction d'ontologie à partir de textes : application à un réseau de périnatalité. In R. DIENG-K UNTZ, Ed., *Actes des 7^{es} Journées Ingénierie des Connaissances*, Laval, France: Presses universitaires de Grenoble.
- [52] WOODS W. A. (1991). Understanding subsumption and taxonomy: A framework for progress. In J. F. SOWA, Ed., *Principles of Semantic Networks*, chapter 1, p. 45–94. San Mateo, Ca: Morgan Kaufmann Publishers.
- [53] ZWEIGENBAUM P., BACHIMONT B., BOUAUD J., CHARLET J. & BOISVIEUX J.-F. (1995). Issues in the structuring and acquisition of an ontology for medical language understanding. *Methods of Information in Medicine*, **34**(1/2).