



UFR SHA

**Mention Information-Communication**  
*Spécialité Documentation*

Année universitaire 2015-2016

Des ontologies pour les humanités :  
Les ontologies de domaine  
à l'épreuve des humanités numériques

Mémoire pour l'obtention du Master esDOC

Présenté par

**Eric Fourcassier**

Le 29 septembre 2016

Sous la direction de

Monsieur David Guillemin  
Université de Poitiers







UFR SHA

**Mention Information-Communication**  
*Spécialité Documentation*

Année universitaire 2015-2016

Des ontologies pour les humanités :  
Les ontologies de domaine  
à l'épreuve des humanités numériques

Mémoire pour l'obtention du Master esDOC

Présenté par

Eric Fourcassier

Le 29 septembre 2016

Sous la direction de

Monsieur David Guillemin  
Université de Poitiers



## Remerciements

Je tiens à d'abord à remercier toute l'équipe enseignante du master esDOC pour sa confiance et tout particulièrement mon directeur de mémoire David Guillemin, pour sa patience et ses précieux conseils.

Je remercie également les professionnels qui, sur mon lieu de stage ont fait preuve d'une grande gentillesse et disponibilité, en particulier :

Olivier Marlet, mon maître de stage, pour la gentillesse de son accueil et le plaisir de notre collaboration,

Béatrice Markhoff, pour sa disponibilité, ses conseils techniques décisifs et ses réflexions sur les ontologies de domaine,

Thomas Francart, pour le temps qu'il a bien voulu m'accorder et ses suggestions techniques indispensables,

Anne-Violaine Szabados pour avoir partagé son expérience et ses réflexions sur les ontologies,

Stéphane Curet pour ses suggestions avisées,

Je remercie enfin mes proches et en particulier ceux que j'ai ennuyé pendant les quelques semaines consacrées à la rédaction de ce travail...

## Sommaire

Partie 1. Les ontologies info-documentaires, architectures du savoir.....	7
1. Ontologie, ontologies.....	8
2. Les humanités numériques : une nouvelle donne pour les ontologies ? .....	31
3. Des ontologies pour les humanités : un contresens ? .....	39
Partie 2. Des ontologies pour les humanités numériques.....	44
1. Emergence du thème des ontologies en humanités numériques dans la littérature spécialisée en ligne : données statistiques et analyse .....	45
2. Quelques ontologies du domaine des humanités : examen et appréciation.....	59
Partie 3. Vers Une interprétation fonctionnaliste des ontologies de domaine ?.....	76
1. Introduction .....	76
2. ....	78
3.1. Une ontologie doit-elle être mimétique ? [réf. À note p.25 article Charlet et Declerck).....	81
3.2. La carte et le territoire.....	83
3.3. Décrire des connaissances pour prescrire des usages : la portée régulatrice des ontologies.....	84
4. Orientations méthodologiques pour la construction d'ontologies dans le champ des humanités numériques.....	85
Bibliographie des ouvrages et documents cités et consultés.....	90

## Introduction

« to on legetai pollachôs »

« l'être se dit en plusieurs sens »

Aristote, *Métaphysique*, Gamma.

Annoncée à grands renforts de promesses radieuses pour le traitement du document et de la donnée, l'avènement du web sémantique dans le monde des bibliothèques et de la documentation interroge tout autant qu'il enthousiasme. Tandis que ses partisans s'empressent d'y voir, à la suite de son principal initiateur et promoteur Tim Berners-Lee, rien moins que l'intelligence artificielle du web, certains éprouvent plus de réserve à l'endroit de ce qu'ils estiment être, comme ... outre atlantique ou Manuel Zacklad en France, un outil ambigu, hanté par la vieille illusion techniciste du sens, qui voudrait qu'on pût automatiser les savoirs et opérations cognitives de l'homme au moyen de syntaxes purement formelles. Dans ce paysage, les ontologies de domaine semblent assumer un rôle majeur. Précédées par l'aura philosophique plus ou moins mystifiante qui émane de leur curieuse appellation, les « ontologies » de domaine réactivent à bien des égards l'idéal rationaliste d'une connaissance ultime et surplombante se déployant dans l'élément d'une *lingua universale* conjurant par un ingénieux artifice, la malédiction de Babel. Car tel est bel et bien hier comme aujourd'hui l'enjeu que les ontologies, philosophique et informatique, s'efforcent de relever : surmonter la diversité des langues naturelles en exhumant sous la variété lexicale, une architecture conceptuelle commune, voire une raison universelle. En capturant cette « conceptualisation partagée » (Gruber) par une « formalisation » implémentable, et donc lisible par des machines, les ontologies info-documentaires offrent par surcroît la possibilité de dominer la masse croissante des données en les liant à l'aide de langages, schémas et modèles qui en assurent la cohérence.

Toutefois, ce qu'il est convenu d'appeler l'ingénierie ontologique (nous aurons l'occasion de voir

que le registre technico-scientifique de l'expression n'est pas indifférent) ne semble avoir su coloniser que les seuls domaines des sciences et des techniques, laissant relativement inexploré le territoire des « humanités ». Pourtant, l'essor récent du mouvement des *digital humanities* (humanités numériques) semble annoncer une nouvelle donne pour les technologies sémantiques et singulièrement pour les ontologies de domaine. L'usage de technologies numériques visant à ouvrir un meilleur accès aux savoirs humanistes, correspond à une dynamique qui tend aujourd'hui à se généraliser, sous l'impulsion d'initiatives soutenues par des acteurs institutionnels majeurs. Cependant, l'orientation des humanités numériques demeure obscure, aussi longtemps que l'on ne perçoit pas la signification exacte de cette étrange notion. Car que peut signifier faire des humanités numériques ? S'agit-il de faire entrer les désuètes humanités dans l'horizon numérique du temps présent, ou bien penser un environnement numérique pour les humanités ? Ce mémoire s'efforce de faire droit à la seconde hypothèse, tout en prenant au sérieux le projet des *digital humanities*, dont l'objet, tenterons-nous de montrer, ne consiste pas tant à faire du numérique avec les humanités qu'à faire numériquement des humanités, conformément à la perspective d'augmentation cognitive propre au web sémantique et prioritairement aux ontologies de domaine.

En quel sens est-il possible, légitime et souhaitable, de construire des ontologies de domaine pour les savoirs humanistes dans le contexte de leur devenir numérique et connecté ? Faut-il en promouvoir le développement, dans le cadre d'une philosophie de l'augmentation des capacités sémantiques des machines, des individus et des communautés, ou en contester, au contraire, le réductionnisme dénaturant que ces ontologies imposent à des savoirs qui, en raison de leur spécificité, se déroberont largement au formalisme logique ?

Pour examiner cette question, nous proposons dans un premier chapitre de retracer la genèse des ontologies afin d'en préciser la nature, les finalités ainsi que les composantes essentielles. Puis, éclairé quant à leur essence, nous chercherons à en en cerner l'existence concrète. Nous ouvrirons alors un deuxième chapitre consacré au recueil d'éléments factuels sur les ontologies de domaine pour les humanités numériques, ainsi qu'à l'examen de quelques exemples d'ontologies que nous avons jugé significatifs et instructifs. Enfin, nous tenterons de reprendre ces deux faces, théorique et pratique, de notre objet d'étude dans un troisième chapitre synthétique, conjuguant à la nécessaire reconnaissance de la spécificité des savoirs humanistes, la possibilité de modélisations ontologiques fécondes qui les prolongent sans les dénaturer.

## Partie 1. Les ontologies info-documentaires, architectures du savoir

La tâche est de rendre explicite ce qui a été laissé tacite et de rendre précis ce qui a été laissé vague ; la tâche est d'exposer et de résoudre les paradoxes, de raboter les aspérités, de faire disparaître les vestiges des périodes transitoires de croissance, de nettoyer les bidonvilles ontologiques.

W. V. O. Quine,  
*Le mot et la chose* (1960),  
Paris, Flammarion, 1999, p. 377-378.

## 1. Ontologie, ontologies...

### 1.1. Qu'appelle-t-on ontologie ? un mot de philosophie

#### 1.1.1. Un sens classificatoire : les êtres

Pour chacun d'entre nous, l'être désigne d'abord l'ensemble infiniment différencié *des* êtres qui nous entourent. A la question « qu'existe-t-il ? », qui résume l'objet de ce que l'on appelle l'ontologie, nous répondrons spontanément « toutes sortes de choses singulières, innombrables et variées ». De sorte qu'entreprendre d'en produire un discours rationnel, signifie ici s'efforcer d'explorer la variété des êtres, dans l'espoir de saisir les liens qui articulent cette multiplicité (le « logos » est « legein », et l'intelligence des choses désigne l'opération de les lier entre elles -« interlegere »). Ce geste intellectuel est bien représenté par l'approche du classificateur, qui travaille à l'inventaire systématique des réalités du monde ou d'une partie du monde.

Il arrive en ce sens que l'on parle d'ontologie pour qualifier la représentation plus ou moins systématique (rationnellement structurée) qu'un individu ou une communauté se fait de tout ou partie du monde. On parlera ainsi de l'ontologie du mathématicien, faite de nombres, du physicien, composée d'atomes et d'énergie, du botaniste, foisonnant de plantes en tous genres, ou encore d'un Lewis Carroll, peuplée de créatures et de syllogismes subtiles et loufoques.

#### 1.1.2. Un sens métaphysique : l'Être

Mais face à ce réel « mosaïque », multiple et changeant, les philosophes ont soulevé une interrogation : de tous les êtres dont on constate l'existence, nous disons qu'ils *sont*. Un passant, un chapeau, une bourrasque, un pécheur, une canne-à-pêche, une chaussure, une perruque, tous ces êtres ont en commun, par-delà leurs différences, le fait d'être. Ce qu'il y a de commun à tous les êtres, c'est l'être. Qu'entend-on par un terme si général ? Quel est le sens de « être » ? L'éclaircissement de cette question appelle une investigation d'une autre nature que celle de

l'observation physique des choses. Car l'être qui donne l'existence, ne peut désigner *un* être parmi les existants. On franchit alors un pas supplémentaire dans le degré d'abstraction et, quittant la physique (qui traite *des* êtres), on aborde une région du savoir que la tradition nommera, un peu par hasard<sup>1</sup>, la « métaphysique ».

On qualifiera alors une ontologie de « métaphysique » pour exprimer qu'elle vise à élucider rationnellement le sens de « l'être » pris grammaticalement au singulier, et philosophiquement au général, désignant le principe unique et unitaire qui domine la variété des êtres. L'ontologie en ce sens tente de répondre à la question qu'« est-ce que » ?

### 1.1.3. L'ontologie informatique et documentaire : un emprunt partiel à la philosophie

En dépit d'une distance à certains égards considérable séparant l'ontologie des philosophes et les ontologies des informaticiens et documentalistes, nous voudrions soutenir qu'il existe sous l'emprunt contingent d'un *terme* la permanence réelle d'un *thème* qui relie l'ingénierie ontologique à une racine philosophique essentielle transmettant l'héritage du réalisme d'Aristote aux ingénieurs de la connaissance en passant par le positivisme logique.<sup>2</sup>

Toutefois, eu égard aux deux significations de l'ontologie esquissées plus haut, il semble que les ontologies informatiques et documentaires s'inscrivent pour l'essentiel dans le prolongement de l'ontologie au sens classificatoire du terme. D'apparition récente, ces langages contrôlés articulés autour de relations logiques plus riches que celles proposées par les thésaurus et taxonomies<sup>3</sup>,

1 L'historiographie atteste (cf Diogène Laërce, *Vies, doctrines et sentences de philosophes illustres*) que le nom de métaphysique est l'invention du bibliothécaire Andronicos de Rhodes. Cherchant une place aux ouvrages d'Aristote traitant de l'être en tant qu'être, il conçut de les placer *après* (« meta ») ses traités de physique, les baptisant alors du nom de « méta-physique ». Le préfixe « meta » (après, au-delà) désigne alors autant la nature transcendante de l'objet de cette science, qu'un ordre de succession dans le classement proposé par un bibliothécaire. Cette anecdote nous enseigne donc ici au passage que la représentation des connaissances a un impact potentiellement considérable sur la façon de produire le savoir. Les options de classement choisies par le bibliothécaire ou le documentaliste véhiculent une certaine interprétation de l'objet qu'il classe.

2 Pour cette hypothèse, voir par exemple Nigel Shadbod, « Philosophical engineering », *Words and Intelligence II*, 2007, pp 195-207, Aurélien Bénel ou Alexandre Monnin. Au delà de son intérêt historique, cette généalogie nous permet de mettre d'invalidier une vision techniciste des ontologies qui tend à les réduire à leur fonction, en perdant de vue qu'elles véhiculent un certain type de rapport aux savoirs qui n'a rien d'évident.

3 En plus des relations méréologiques (tout/partie), les ontologies intègrent dans leur principe toutes les relations formalisables en logique du premier ordre.

suscitent, depuis leur émergence dans les années 1990, un engouement qui n'est sans doute pas étranger aux représentations plus ou moins imaginaires véhiculées par les technologies de l'intelligence artificielle. Ces outils informatiques et documentaires laissent en effet espérer une optimisation de la recherche d'information *via* l'automatisation de l'indexation, la sémantisation des données et l'interopérabilité résultant des technologies du web sémantique. Il devient ainsi possible d'effectuer un traitement « intelligent »<sup>4</sup> des données d'un domaine de connaissance déterminé : l'ontologie permet de former à un raisonneur automatique de former des inférences valides (c'est-à-dire cohérentes), univoques (non ambiguës) et sémantiquement pertinentes, au regard des attentes exprimées par les requêtes utilisateurs.

## 1.2. Panorama historique : de l'ontologie philosophique aux ontologies info-documentaires

### 1.2.1. L'ontologie avant l' « ontologie »

Cette architecture des concepts d'un domaine, organisée autour de dichotomies et relations logiques précises n'est pas sans rappeler la démarche classificatoire héritée d'Aristote et transmise, *via* la scolastique médiévale et les théories de la « *scala naturae* » (l'échelle naturelle des êtres) aux savants qui développèrent dans « l'histoire naturelle » d'abord, puis dans les sciences du vivant proprement dites, les grands systèmes de classification des espèces vivantes (les arbres dichotomiques des naturalistes, la classification binominale des végétaux -selon la différence spécifique et le genre prochain- chez Linné, entre autres exemples).

Toutefois, le *terme* « ontologie » est très postérieur à l'apparition du *thème* ontologique lui-même, dans la philosophie de son fondateur, Aristote. La question de la représentation des rapports entre des concepts primitifs soulevée par l'ingénierie ontologique constitue en effet un lointain héritage des développements du chapitre II des *Catégories* d'Aristote<sup>5</sup>.

4 On entendra tout au long de ce mémoire, à défaut d'autre explicitation, qu'un traitement automatique des données est « intelligent », dès lors qu'une machine peut opérer des inférences valides à partir de données préalablement liées entre elles (du latin *inter-legere*) selon des lois logiques. Le web est sémantique parce qu'il est un web des données liées.

5 Contrairement à un usage répandu dans la littérature sur les ontologies de domaine, ce n'est pas dans le texte de la *Métaphysique* d'Aristote qu'on trouvera une préfiguration significative de ces outils de représentation des connaissances, mais dans ses traités de logique réunis dans l'*Organon*, en particulier les *Catégories* et *De l'Interprétation*.

Suivant le programme qu'il assigne à la « philosophie première », l'auteur des *Catégories* s'emploie dans ce traité à dégager les éléments fondamentaux de l'être. L'ontologie que propose Aristote comprend alors des entités (ou catégories) et relations primitives permettant de guider la prédication dans les jugements de connaissance et jeter les bases de la logique que les logiciens de la fin du XIX<sup>e</sup> et du début du XX<sup>e</sup> siècle formaliseront sous le nom de calcul des prédicats.

Par souci de simplification du propos, nous nous contenterons d'en reprendre l'exposition scolastique sous la forme traditionnelle du carré suivant :

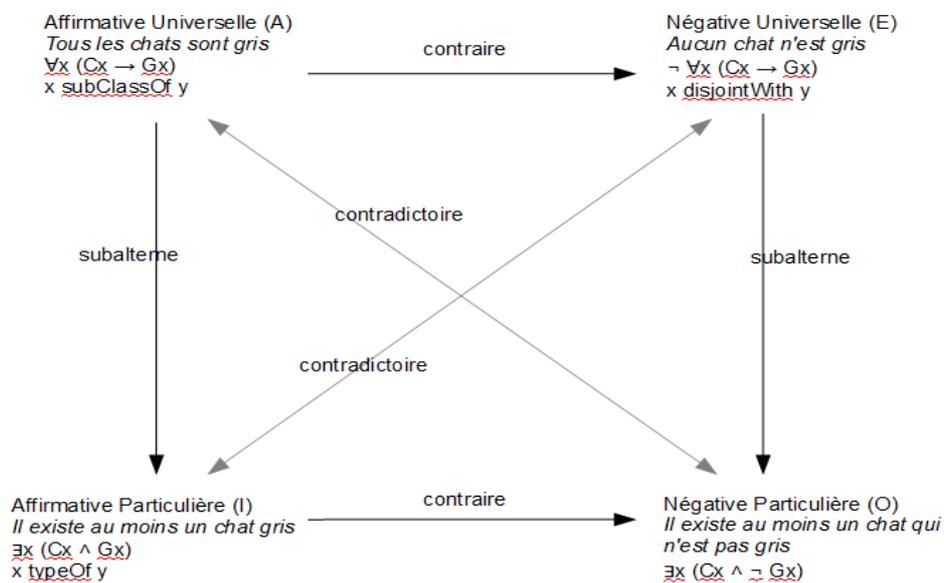


Figure 1. Le « Carré d'Aristote », appelé aussi « carré logique », assorti de notations correspondantes en logique des prédicats et langage OWL.

Cette préfiguration aristotélicienne de la notion d'ontologie développée en ingénierie de la connaissance se précise encore sous la plume d'un des plus célèbres commentateurs des *Catégories* : Porphyre de Tyr. Dans son *Introduction aux Catégories d'Aristote* (*l'Isagogè*) ce philosophe néoplatonicien du 3<sup>e</sup> siècle propose, dans l'esprit platonicien des dichotomies et rassemblements, une présentation des catégories sous forme d'arbre dichotomique des substances, connu depuis Boèce sous le nom d'« arbre de Porphyre »<sup>6</sup>.

<sup>6</sup> Porphyre, *Isagogè*. Texte grec, traduction française en vis-à-vis, texte latin, introduction et notes par A. de Libera. Vrin, « Sic et Non » (244 p.), ISBN 978-2-7116-1344-1. Plus proche de nous, Barry Smith définit les ontologies comme une « famille d'arbres » (in *The Basic Tools of Formal Ontology. Formal Ontology in Information Systems*, Amsterdam, Oxford, Tokyo, Washington, DC : IOS Press, 1998).

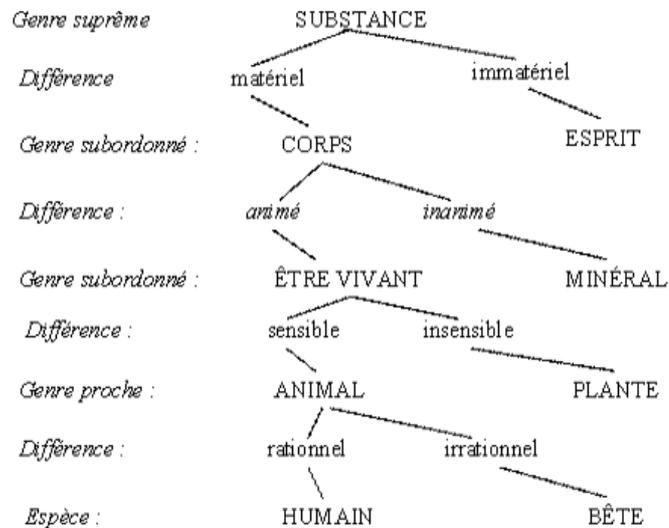


Figure 2 : l'arbre de Porphyre. Un lointain ancêtre de nos graphes conceptuels. source : [Liris.cnrs.fr/amille/enseignements/DEA-ECD/ontologie/notion\\_ontologie](http://Liris.cnrs.fr/amille/enseignements/DEA-ECD/ontologie/notion_ontologie)

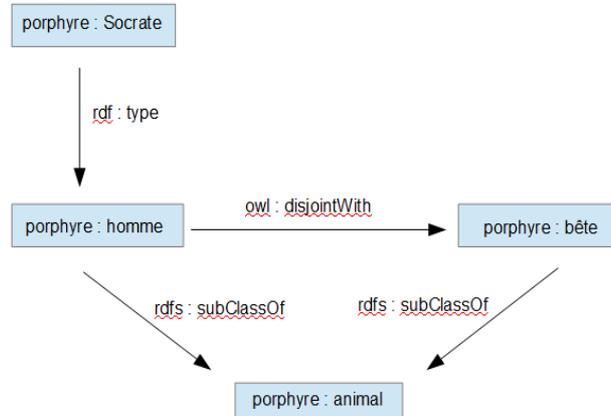


Figure 3. Ebauche d'ontologie pour le web inspirée de Porphyre. Source : Aurélien Bénel, *Ontologies du web : histoire refoulée et perspectives paradoxales*, in *Intellectica*, 2014/1, 61, 123-141.

L'ontologie des *Catégories* s'y trouve ainsi représentée sous cette forme diagrammatique qu'on lui connaît aujourd'hui encore : l'arborescence étend une multitude de branches représentant l'ensemble des subdivisions dichotomiques des êtres selon la différence spécifique et le genre prochain. Cette ontologie (ou échelle des êtres) ainsi constituée, trace un schéma où le général se

trouve progressivement particularisé, par une suite de différenciations successives, pour aboutir aux unités singulières (les « individus », littéralement « indivisibles ») vers lesquelles pointent les ramifications dernières. La hiérarchie des substances s'effectue ainsi selon les rapports logiques guidant toute prédication qui se voudra valide quant à sa forme.

Cette articulation logique des genres et des espèces a par ailleurs une dimension proprement sémantique, car elle offre un modèle rigoureux pour la construction de définitions : une espèce est strictement coextensive avec le genre et la différence spécifique qui en forment la définition. La définition de l'homme se construit par articulation du genre le plus proche (animal) et de sa différence spécifique (rationnel). La relation est coextensive : tout homme est animal rationnel et tout animal rationnel est homme.

L'ontologie prise ici au sens d'un système des catégories (ou concepts les plus généraux) ne consiste pas en une simple classification d'entités, mais constitue un ensemble structuré de relations logiques de prédication, c'est-à-dire d'inférences<sup>7</sup>. La structure en triplet du modèle RDF, Sujet-Prédicat-Objet (ou Entité-Propriété-Instance de propriété), brique élémentaire du sens dans le web sémantique, emprunte directement à l'analyse logique des énoncés chez Aristote et ses épigones, de Porphyre (3<sup>e</sup> siècle) à Frege (19<sup>e</sup> siècle).

### 1.2.2. Le moment lexical de l'ontologie

A ce moment de constitution de l'ontologie comme *thème* philosophique, succède un moment « lexical » (Frédéric Nef<sup>8</sup>) qui voit l'apparition et l'adoption définitive du *terme* « ontologie ». On atteste l'usage explicite du substantif « *ontologia* » dans la littérature encyclopédique aux alentours de 1620 (Alsted, *Encyclopedia*). Souvent cité à ce propos, l'article de Goclénus témoigne

---

<sup>7</sup> Le substantif *κατηγορία* est formé sur le verbe *κατηγορεῖν* : « attribuer ». Chaque catégorie correspond littéralement à une façon particulière d'attribuer un prédicat à un sujet. Ceci signifie que le langage n'est pas un répertoire de dénominations, mais un support de prédication ; idée que l'on retrouve dans les ontologies informatiques et documentaires.

<sup>8</sup> *L'Ontologie au miroir de la Terminologie*, in *Terminologie et ontologie : Théories et Applications, Actes de la conférence TOTH*, 2010, p.9.

d'une filiation qui demeure aristotélicienne : l'auteur y conçoit l'ontologie comme l'étude de la structure conceptuelle du monde et la tentative de sa représentation sous forme de tableau dichotomique accueillant, chacune à sa place, les réalités du monde. Et il faut remarquer à cet égard que la démarche classificatoire et logique possède ici, comme chez Aristote, une portée métaphysique : l'ontologie est la science abstraite des termes transcendants qui « coiffent » la hiérarchie des êtres. Celle-ci est « abstraite » car elle se constitue par soustraction (ou défalcation) des caractères concrets qui particularisent les êtres du monde, afin de n'en retenir que les constituant les plus fondamentaux. On passe ainsi des objets concrets de perception, aux objets abstraits des sciences particulières, puis aux catégories qui en forment la sémantique générale (ontologie au sens classificatoire), et enfin aux termes transcendants (ontologie au sens métaphysique, comme étude de l'être en tant qu'être). Science du Principe, l'antique Protè Philosophia (ou Métaphysique) prend ainsi le nom d'*Ontologia*, recherche de l'être premier, source transcendante de la variété des êtres, comme l'atteste jusqu'au XVIII<sup>e</sup> siècle la philosophie de Christian Wolff<sup>9</sup>, avant la grande rupture kantienne.

### 1.2.3. Le moment logiciste de l'ontologie, « clarification logique de la pensée »<sup>10</sup>

Avec les travaux des philosophes et logiciens du Cercle de Vienne (dans le courant des années 1930), le projet de l'ontologie philosophique essuie d'abord une critique virulente dans le cadre d'une réfutation générale de la métaphysique par l'analyse logique des énoncés (Carnap, Wittgenstein). La notion d'ontologie ne ressurgit alors qu'à la faveur d'une assez profonde réforme de sa signification dans l'œuvre de Quine.

Prenant acte des postulats du positivisme logique, Quine propose d'entendre par ontologie, un réseau de déclarations ou « tissu de croyances » (a « web of beliefs ») au sujet de ce qui existe, basé sur les sciences de la nature. Chaque science secrète en effet une ontologie en ceci qu'elle a besoin de poser certaines entités sur l'existence desquelles elle s'engage. Abandonnant toute approche surplombante, l'ontologie se contente dans cette perspective de rendre manifestes les engagements ontologiques effectifs d'une science, au moyen de la logique du premier ordre,

9 WOLFF, Christian. *Philosophia prima sive ontologia. Francofurti et Lipsiae : prostate in officina libraria Rengeriana*, 1730, 696 p. La philosophie première (« Philosophia prima ») est synonyme (« sive ») d'ontologie.

10 Selon l'expression de Wittgenstein (*Tractatus*, § 4.112) résumant le programme du positivisme logique du « Cercle de Vienne ».

garante d'une clarté et d'une distinction supérieure à toute langue naturelle. Aussi, pour déterminer quels sont les engagements ontologiques propres à une science particulière, il faut disposer d'un critère d'identification. Fidèle à son inspiration logiciste, Quine pose alors ce qu'il appelle le « critère d'engagement ontologique » (« criterion of ontological commitment ») qu'il énonce par la formule suivante : « être, c'est être la valeur d'une variable liée » (« To be is the value of a bound variable »). En d'autres termes, pour identifier les engagements ontologiques d'une science (*i-e* les entités que son discours engage à tenir pour existantes), il faut déterminer les valeurs des variables quantifiées utilisées dans ses formulations canoniques. L'existence s'en trouve de ce fait réduite à l'extension logique des concepts : elle est l'ensemble de tous les individus ou occurrences d'objets auxquels un concept réfère, par opposition à son « intension » (sa définition en termes de propriétés ou d'attributs). Avec Quine, le jugement d'existence (« ceci existe ») ne porte plus sur les propriétés ou attributs qui entrent dans la définition d'un concept, mais uniquement sur l'individu ou la classe des individus auxquels les termes de la langue se réfèrent. On parle alors de « valeur » pour désigner le nombre ou la classe de ces individus et de « variable liée » pour signifier les termes ou concepts qui s'y réfèrent.

Cette conception logicienne de la notion d'existence permet de mettre au jour l'ontologie d'une science particulière en procédant à la réduction de son langage à une formulation en logique.

Les travaux de Quine ont inspiré nombre de pionniers de l'intelligence artificielle, dont en premier lieu John McCarthy qui, en référence à la définition quinienne de l'ontologie, déclare que la construction de systèmes intelligents basés sur la logique doit commencer par dresser la liste de ce qui existe en construisant une ontologie de notre monde. C'est cette approche de l'ontologie qui débouchera sur la création en 1999 de la Gene Ontology.

Dès lors, sans entrer plus avant dans les éléments de fond sur la théorie quinienne de l'ontologie, cet aperçu de ce que nous appelons ici le moment logiciste de l'ontologie nous permet d'identifier le point où un passage de relais s'effectue de la philosophie vers l'ingénierie informatique. De ce moment logiciste, nous retiendrons que le projet de l'ontologie est sorti profondément révisé : dégagé de toute prétention métaphysique, l'ontologie ne désigne plus qu'un « mythe » auquel nous nous « engageons » à croire et dont la vérité (ou la fausseté) ne dépend pas de l'expérience. L'ontologie se voit de la sorte réduite à « l'univers du discours » formel (Aurélien Bénéel).

### 1.3. Les ontologies de domaine en ingénierie des connaissances

#### 1.3.1. Emergence

Si la notion apparaît dès 1980 sous la plume de John McCarthy<sup>11</sup>, s'inspirant de Quine, c'est au début des années 1990 que les ontologies connaissent leurs premiers développements dans le cadre des recherches en ingénierie des connaissances où elles tendent à remplacer progressivement les systèmes experts, logiciels capables d'effectuer un certain nombre d'inférences à partir de règles prédéfinies. A ces moteurs d'inférences capables de reproduire le raisonnement humain (les démonstrations de la logique formelle), les systèmes à base de connaissances (ou « SBC ») ont apporté une amélioration en séparant, d'un côté, les connaissances du domaine envisagé et, de l'autre, les règles gouvernant les raisonnements à l'œuvre dans ce même domaine de connaissance. Cette double spécification du raisonnement et des connaissances d'un domaine fonde les spécifications des ontologies développées par les chercheurs en ingénierie de la connaissance.

Mais c'est avec l'émergence du web sémantique que les ontologies connaissent leur véritable essor. Les travaux sur la représentation des connaissances en réseaux sémantiques et graphes conceptuels, sont en effet directement débiteurs des avancées du web des données tel que Tim Berners-Lee en fixa en 2001 les grands principes fondateurs<sup>12</sup>. Ce nouvel âge du Web se caractérise en effet par le développement d'outils permettant de rendre accessible et exploitable le contenu des ressources du web, par des agents logiciels, au moyen des métadonnées structurées essentiellement à l'aide des langages définis par le W3C. Cette sémantisation des données permet aux ordinateurs de construire des relations douées de sens et, ainsi, d'exploiter les informations du web de façon pertinentes au regard des attentes d'un utilisateur humain.

C'est dans le cadre de ce projet de sémantisation des données qu'apparaît un nouvel usage

---

11 McCarthy, J., *Circumscription – A form of non-monotonic reasoning*, in *Artificial Intelligence*, 1980, pp. 27-39.

12 Berners-Lee, Tim *et al.* « The Semantic Web », *Scientific American*, 2001.

informatique et documentaire de la notion d' « ontologie ». Davantage qu'un simple emprunt terminologique, ce choix traduit une certaine conception de la sémantique des données et enveloppe à ce titre une théorie du sens, à la fois simple et opérationnelle.

Qu'est-ce, en effet, qui décide du sens d'un énoncé ? Rien d'autre, répond l' « ontologue », que l'ensemble des liens qui unissent cet énoncé aux représentations qu'un individu ou une communauté se fait du monde. En d'autres termes, le sens d'un énoncé se découpe toujours sur le fond d'un certain paysage mental et linguistique, un certain « tissu de croyances » (pour reprendre l'expression de Quine<sup>13</sup>, « Web of Beliefs »), à partir duquel nous sommes capables de l'interpréter avec justesse. En somme, posséder l'intelligence d'une chose, signifie saisir les liens qui l'unissent à toutes les autres dans un réseau sémantique ou « treillis conceptuel » : une ontologie. Ainsi, par exemple : un chien -(est un)- animal - quadrupède – meilleur ami de l'homme - (n'est-pas)- un chat, etc.

La *signification* d'un concept dans une ontologie correspond à l'*extension* de ce concept, soit l'ensemble des individus ou « instances » appartenant à une certaine classe. L'unité de cette classe sera spécifiée quant à elle par la définition du concept correspondant. On appellera *intension* (à la suite de Rudolf Carnap reprenant la distinction de Frege évoquée plus haut) l'ensemble des propriétés que doivent nécessairement posséder tous les individus entrant sous cette classe (ou stipulant les conditions qu'ils doivent remplir). Notons toutefois qu'au sein d'une ontologie, l'intension d'un concept ne se réduit pas à la définition donnée sous forme de liste de propriétés ou conditions nécessaires et suffisantes, mais dépend également de la position de ce même concept à l'intérieur du réseau sémantique qui relie ce concept aux autres par le biais des relations d'héritage et d'association : dans l'ontologie ordinaire du citoyen d'Athènes, la classe « Athénien » hérite des propriétés de la classe « Homme » (par subsomption), mais s'oppose à la classe « spartiate » et plus encore à la classe « barbare » (association).

Tim Berners-Lee semble avoir été le premier à percevoir l'étendue des perspectives qu'une telle conception du sens ouvrait au traitement intelligent des contenus sur le web. Dans une conférence de 1994, le célèbre co-inventeur du web exposa ainsi l'audacieux projet d'un « *semantic web* », conférant aux machines la capacité d'accéder au sens des données, moyennant leur structuration

---

13 Titre d'un ouvrage de Willard Van Orman Quine (philosophe et logicien américain, 1908-2000) *The Web of beliefs*, McGraw-Hill Education, 1978.

par des formats, schémas et syntaxes adéquats. La conviction qui se fait jour alors introduit une rupture qui inaugurerait le nouvel âge du « web 3.0 » : le web ne saurait se réduire à sa seule dimension documentaire, car ce ne sont pas tant des documents que l'utilisateur recherche, que l'information qu'ils contiennent. Pour lui délivrer cette information, il suffirait à une machine de posséder la « carte mentale » des représentations de l'utilisateur, de la déplier et de la parcourir, afin de déterminer le chemin débouchant sur la réponse adéquate à sa requête. Tout l'objet du développement d'une ontologie informatique et documentaire consiste ainsi à capturer la représentation d'un domaine du réel en tâchant d'en préciser les concepts (unités sémantiques), par-delà la variété de leurs dénominations en langues naturelles (unités lexicales).

Ce schéma d'imbrication logique des classes peut être formalisé à l'aide d'outils informatiques, de telle sorte qu'en les parcourant, une machine peut simuler le raisonnement humain.

Pour emprunter une illustration proposée par Fabien Gandon dans son support de cours en ligne sur le web sémantique de 2016 (plateforme FUN), chacun admettra que pour répondre à la question « qu'ai-je lu cette semaine ? », il lui faut consulter sa mémoire et passer en revue l'ensemble des choses lisibles (revues, articles, magazines, livres, etc) que referme son ontologie personnelle et les associer à son interrogation. Une ontologie informatique et documentaire n'a pas d'autre dessein que d'emprunter le même cheminement conceptuel (voir figure ci-dessous).<sup>14</sup> Historiquement, poursuit Gandon, l'encodage de cette conceptualisation s'est d'abord effectué au moyen du langage des « frames », puis des logiques de description (DL), avant d'aboutir aujourd'hui au modèle des graphes RDF.

---

<sup>14</sup> Les ontologies du web s'apparentent à cet égard aux « réseaux sémantiques » à l'aide desquels certains chercheurs en psychologie cognitive (telle Eleanor Rosch) tentèrent dans les années 1970 de modéliser la manière dont les connaissances d'un individu s'organisent dans sa mémoire, et se retrouvent par le même chemin. Voir François Raster, « Ontologie(s) », *Revue des sciences et technologies de l'information, série Revue d'Intelligence artificielle*, vol. 18 (1), 2004, p. 15-40.

- Manque une connaissance → **identification**
- Types de documents → **acquisition**
- Modéliser et formaliser → **représentation**

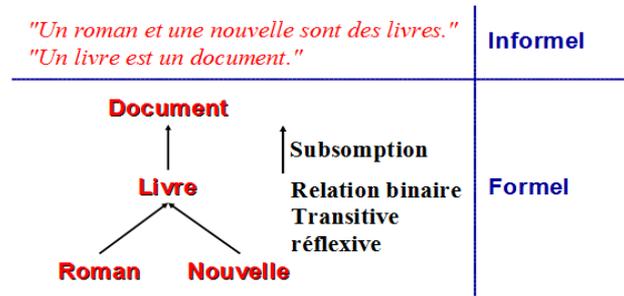


Figure 4 . Des données informelles à leur modélisation formelle : l'exemple du livre. Lorsqu'un individu s'interroge sur ce qu'il a lu dernièrement, il identifie un manque de connaissance qui le pousse à passer en revue les divers types de documents qui représentent pour lui des choses lisibles (un roman, une nouvelle, etc). Ce qu'il comprend par « chose lisible » ou « lue dernièrement » dépend directement des réalités qu'il y associe. Une ontologie de domaine se propose de reproduire cet itinéraire humain de l'interprétation du sens des énoncés en modélisant et formalisant sa représentation du réel.

Précisons toutefois, au passage, qu'une ontologie ne saurait prétendre constituer une représentation, même partielle, du monde lui-même. Il s'agit plus sobrement pour elle de représenter adéquatement une conceptualisation portant sur un domaine circonscrit du savoir. Leur rôle est de décrire les relations *formelles* entre des concepts, aucunement des relations *matérielles* (i-e réelles) entre les objets qu'ils dénotent. De sorte que, comme l'indique Barry Smith<sup>15</sup>, on se gardera d'y importer trop hâtivement les questions traditionnelles du réalisme ontologique issues de la tradition métaphysique : les ontologies de domaine sont des objets *documentaires* qui appellent une approche strictement *pragmatique*. Umberto Eco formule quant à lui la même idée : « une ontologie, aussi grossière et naïve soit-elle, est la représentation locale d'une portion de connaissance encyclopédique pertinente pour un univers de discours déterminé. »<sup>16</sup>. C'est en ce sens uniquement que l'on peut dire qu'une ontologie de domaine

15 Smith Barry, « The Relevance of Philosophical Ontology to Information and Computer Science », *Philosophy, Computing and Information Science*, London: Pickering and Chatto, 2014, 75-83

16 Eco Umberto, *De l'arbre au labyrinthe*, trad. Hélène Sauvage, Le Livre de Poche, 2011, p.97.

*représente* un champ du savoir. Ce caractère modeste et pragmatique dans leur objet et leur usages explique en partie<sup>17</sup> la relative indifférence de l'érudit de Milan aux ontologies informatiques et documentaires, dont l'emprunt terminologique à la philosophie s'avère donc, de ce point de vue, relativement « impropre » (ibid).

### 1.3.2. Définitions

Si l'on peut s'accorder sur une définition large de l'ontologie comme artefact informatique permettant la représentation structurée d'un domaine de connaissance lisible par une machine, une spécification plus précise suppose toutefois l'adoption de certains présupposés théoriques sur lesquels le consensus ne va pas toujours soi. Or, cette variété de points de vue ne relève pas seulement d'un scrupule terminologique, mais engage la façon de concevoir une ontologie (ses finalités, ses principes et sa structure). C'est ce que nous apprend la lecture des premiers théoriciens, Thom Gruber, John Sowa, Nicolas Guarino et, de ceux qui, dans le contexte français, prolongent leurs analyses : Natalie Aussenac-Gilles, Bruno Bachimont, Pierre Charlet, Alexandre Monnin, ou encore Raphaël Troncy.

La définition la plus synthétique nous vient de Tom Gruber<sup>18</sup> (1993), principalement sous la forme légèrement modifiée qu'en a tiré Studer<sup>19</sup> : une ontologie informatique désigne « une spécification formelle et explicite d'une conceptualisation partagée ». Cette définition signifie d'abord que la construction d'une ontologie repose sur l'accord d'une communauté d'individus (« conceptualisation partagée ») sur la définition des concepts et relations d'un domaine de connaissance déterminé. Une fois ces éléments validés, un second temps consiste à les structurer à l'aide d'outils de formalisation logique puis informatique (« spécification formelle ») qui rendent ces concepts et relations lisibles par des machines (« machine readable ») : les programmes ou

---

17 L'autre partie des réticences d'Umberto Eco semble tenir au fait que les ontologies reposent sur une conception fermée, ordonnée et hiérarchique du savoir qu'elles héritent de la structure de l'arbre de Porphyre, image d'un monde clos, caractéristique de la cosmologie médiévale (Eco, op.cit., p.60). A cette figure arborescente du savoir fermé, Eco oppose la figure labyrinthique (rhizomatique) d'un réseau ouvert des connaissances et de l'information (dont l'image sur le web serait alors l'indexation par les tags). Cette remarque met en cause directement la capacité d'une ontologie à épouser les évolution des savoirs du domaine qu'elle entend couvrir. Représentation fixiste impropre à rendre compte d'un savoir vivant, selon la vision de leurs adversaires, les ontologies ouvriraient alors une perspective moins prometteuse qu'ont l'eût pensé.

18 Gruber Tom, « A translation approach to portable ontology specifications », *Knowledge acquisition*. 1993. Vol. 5, n°2, p. 199–220.

19 Studer, Benjamins, Fensel, « Knowledge engineering: Principles and methods », *Data & Knowledge Engineering*. 1998

agent logiciels peuvent répondre par des inférences logiques et sémantiquement pertinentes aux besoins des utilisateurs.

Ces deux dimensions sont inséparables : si le caractère technique (informatique) et logique des ontologies s'avère central, on se gardera toutefois de réduire ces outils à de purs produits indépendants d'un contexte d'utilisation précis et d'une délibération collective concernant la conceptualisation du domaine de connaissance considéré. En d'autres termes, une ontologie n'existe qu'en vertu d'un contexte humain de délibération et d'action qui lui donne son sens et sa structure. Deux dimensions de la signification, formelle (au sens de la logique formelle) d'un côté, socio-sémantique (c'est-à-dire relative à un contexte collectif d'interprétation) de l'autre, se superposent ici, sans qu'il soit assuré qu'elles forment un tout cohérent. De sorte que la structuration logico-formelle des connaissances d'un domaine relevant des sciences interprétatives ou herméneutiques (les sciences humaines), pourrait constituer une difficulté fondamentale, comme nous le verrons à l'issue de ce chapitre. Cette dimension que nous appelons délibérative et herméneutique est d'ailleurs soulignée par Bruno Bachimont, Pierre Charlet et Raphaël Troncy, dans un article publié en 2004<sup>20</sup>, qui rappellent dans leur « Définition 1 » que « Construire une ontologie, c'est *décider* de la manière d'être et d'exister des objets. » (souligné par nous). Et d'ajouter dans leur « Définition 2 » que toute ontologie « implique ou comprend une certaine vue du monde par rapport à un domaine donné ».<sup>21</sup>

En tant que conceptualisation, une ontologie s'appuie sur des définitions univoques tirées du corpus des connaissances d'un domaine. Mais en tant qu'elle est « formelle » (conçue dans l'élément de la logique des inférences valides) l'ontologie se heurte à la variété des interprétations des concepts du domaine, ce qui impose d'en circonscrire les limites en spécifiant chaque ontologie par rapport à ses usages. De ce point de vue, différentes espèces d'ontologies sont envisageables. Tâchons donc à présent d'en dresser une typologie.

### 1.3.3. Typologie

En suivant la suggestion de Van Heijst (cité par Charlet, Bachimont et Troncy, op.cit.), on peut

<sup>20</sup> Charlet Jean, Bachimont Bruno, Troncy Raphaël, « Ontologies pour le Web sémantique ». *Revue Information, Interaction, Intelligence 13* [En ligne]. 2004.

<sup>21</sup> Cette dimension constructiviste (déliurgique?) des ontologies est également soulignée par Tim Berners-Lee :

dégager quatre grandes familles d'ontologies :

- **Les ontologies de domaine**

Ces ontologies (qui font, par ailleurs, l'objet principal de notre propos) consistent en des modélisations des concepts et relations propres à une discipline particulière, un domaine du savoir (les sciences, les techniques). L'objectif de ce type d'ontologie est d'offrir une représentation structurée des connaissances d'un domaine de manière à rendre possible leurs exploitations par diverses applications.

- **Les ontologies d'applications**

Cette famille d'ontologies (dites aussi applicatives), plus spécifique, rassemble les connaissances requises pour une application particulière. Si elles peuvent inclure une ontologie de domaine, elles ne sont en revanche pas pensées pour être réutilisables.

- **Les ontologies génériques ou ontologies de haut niveau**

Ces ontologies articulent des concepts et relations d'un haut degré de généralité, tels que l'espace, le temps, la matière, l'objet, l'événement, l'action. Ce niveau d'abstraction leur confère un caractère transversal et surplombant par rapport aux ontologies circonscrites à des ensembles plus spécifiques de concepts, telles les ontologies de domaines, qu'elles viennent donc « coiffer » au sommet de l'architecture conceptuelle.

- **Les ontologies de représentation.**

Cette classe d'ontologies vise à regrouper les concepts nécessaires à la formalisation des connaissances. Tel est le cas, par exemple de la « Frame-Ontology », dans laquelle se trouvent décrites un ensemble de notions permettant de spécifier les connaissances, telles que la substance, les relations, les concepts, etc. Cette ontologie modélise les concepts utilisés dans les langages à base de « frames » : classes, sous-classes, attributs, valeurs, relations et axiomes. Ce caractère générique des ontologies de représentation les rend indépendantes des ontologies de domaines auxquelles elles fournissent un matériel conceptuel constituant un cadre commun aux diverses représentations des connaissances particulières.

A cette première approche typologique, centrée essentiellement sur le degré de généralisation des concepts et relations des ontologies, Uschold et Gruninger<sup>22</sup> ajoutent une autre distinction, selon le langage utilisé et son degré de formalisation. Quatre ensembles se dégagent alors :

- **Les ontologies informelles**

Cette qualification d' « informelle » signifie que ce type d'ontologies repose uniquement sur une langue naturelle. Leur avantage principal est donc d'être aisément compréhensibles et manipulables par un utilisateur humain. Toutefois, cet atout est contrebalancé par un inconvénient corrélatif, qui tient à la difficulté de contrôler la cohérence d'un ensemble non formalisé (ambiguïtés, redondances, contradictions). Si l'usage de ces ontologies est facilité de prime abord, leur validation pose donc plus de difficultés.

- **Les ontologies semi-informelles.**

Dans le cas de ces ontologies, les ambiguïtés inhérentes à la langue naturelle sont corrigées par l'adoption d'une forme plus structurée qui clarifie l'architecture d'ensemble.

- **Les ontologies formelles**

Le choix est fait, cette fois, de s'affranchir des langues naturelles et de s'appuyer sur un langage artificiel gouverné par une sémantique formelle stricte. D'accès plus incommode en raison de son caractère non intuitif, ces ontologies ont en revanche l'avantage d'offrir une architecture dont la cohérence peut être vérifiée de façon plus rigoureuse, écartant tout risque d'inférences invalides ou non pertinentes.

- **Les ontologies semi-formelles**

Ces ontologies reposent, comme les précédentes, sur un langage artificiel et formel qui au

Le niveau de formalisation d'une ontologie commande par ailleurs sa capacité à effectuer les opérations automatisées qui lui sont confiées. A cet égard, comme l'indiquent Uschold et Gruninger, un choix s'impose : « Si une ontologie est une aide à la communication entre personnes, alors la représentation de l'ontologie peut être informelle du moment qu'elle est précise et qu'elle capture les intuitions de chacun. Cependant, si l'ontologie doit être employée par des outils logiciels ou des agents intelligents, alors la sémantique de l'ontologie doit être rendue beaucoup plus précise ». La décision repose donc sur la finalité et l'usage de l'ontologie, en

---

<sup>22</sup> Uschold Mike et Gruninger Michael, « Ontologies : principes, Methods and Applications », preprint to appear in *Knowledge Engineering Review*, Vol. 11, Num. 2, Juin 1996

prenant en compte l'agent, humain ou logiciel, auquel elle est prioritairement destinée.

Enfin, un dernier critère entre en jeu pour compéter cette typologie : le degré du détail des objets impliqués dans la conceptualisation de l'ontologie, aussi appelé niveau de granularité. C'est ici l'objectif opérationnel de l'ontologie qui permet d'en décider. Suivant les contextes d'utilisation, il pourra s'avérer indispensable de mobiliser une conceptualisation très détaillée, ou au contraire de se concentrer sur des caractères plus généraux en évacuant des détails trop fins. On parlera ainsi de granularité fine et de granularité large :

- **Granularité fine**

Ce niveau de détail dans la conceptualisation d'un domaine est exigé par les ontologies très spécialisées et dont l'usage s'adresse à des utilisateurs initiés au domaine en question.

- **Granularité large**

Ce degré moins élevé du détail conceptuel de l'ontologie s'observe essentiellement dans les ontologies de haut niveau : leur positionnement au sommet de la hiérarchie conceptuelle impose en effet de s'en tenir à des notions suffisamment larges pour maintenir la généralité des notions qu'elles contiennent.

En tout état de cause, cette typologie fait apparaître une tension entre deux grandes approches de l'ontologie qui engendrent une certaine ambiguïté. Si Charlet, Bachimont et Troncy assignent aux ontologies la fonction du raisonnement automatique, Uschold et Gruninger envisagent des modèles moins formels d'ontologie. Cette divergence d'approches concernant l'importance de la formalisation des ontologies est cause d'un certain flou terminologique. Pour notre part, nous estimons que les ontologies info-documentaires ne trouvent leur sens et leur valeur qu'en tant qu'elles présentent une structure sémantique compréhensible par des machines, basées sur les technologies du Web des données. Cette définition est au reste la plus généralement admise par la communauté des chercheurs en ingénierie de la connaissance. Toutefois, ce parti pris ne va pas sans poser des difficultés lorsqu'il est question d'élaborer et de manipuler des ontologies de domaines, en particulier lorsque le secteur de connaissances envisagé peine à s'affranchir des formulations en langue naturelle. De ce point de vue, une forme de contradiction semble s'élever entre le formalisme logico-formel des ontologies et la rationalité propre des sciences humaines et

plus largement des humanités, attachée aux langues naturelles et à des contextes d'énonciation et d'interprétation variables.

On peut toutefois considérer que toute construction d'ontologie consiste à passer d'un contenu originellement informel à une structure formelle et opérationnelle. C'est ainsi que Frédéric Fürst propose de présenter le processus d'élaboration d'une ontologie. Le schéma suivant résume la démarche :

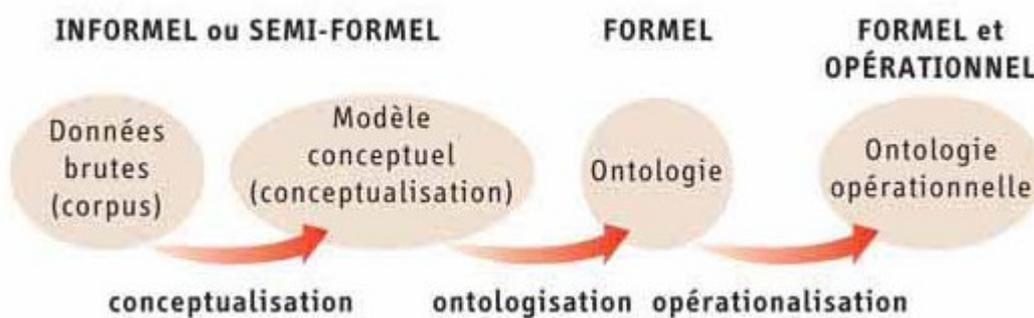


Figure 5. Schéma de construction d'une ontologie opérationnelle, d'après Frédéric Fürst.

Le processus de construction d'une ontologie de domaine consiste en une opération de formalisation plus ou moins stricte suivant les objectifs visés et la nature du domaine considéré. Toutefois, pour entrer davantage dans ce processus de construction, il faut préciser les éléments constitutifs d'une ontologie.

#### 1.3.4. Les constituants des ontologies

Les constituants d'une ontologie peuvent se déduire de la structure en triplet qui sert à représenter la sémantique d'un contenu quelconque : classe – propriété – instance. Une ontologie présente donc un ensemble de concepts et de propriétés (relations) permettant d'articuler des connaissances particulières, qu'on pourra rattacher à l'ontologie par instanciation.

- **Les concepts** représentent les « entités » articulées dans les ontologies, autrement dit les contenus de significations en jeu (objets ou concepts d'un domaine) appelés « instances

dans le langage OWL. Contrairement aux thésaurus qui renferment des unités lexicales (un recueil de termes), les ontologies renferment des unités sémantiques que l'on nomme assez indifféremment classes, catégories ou concepts.

- **Les classes** désignent les concepts primitifs, identifiés par des termes singuliers ou des groupes de mots (labels). Il peut ainsi exister plusieurs labels pour un même concept : une relation de synonymie est alors posée qui élargit la portée du concept en augmentant sa richesse linguistique. Pour compléter cette caractérisation des concepts, des définitions et des indications d'usage peuvent les accompagner. Une classe peut enfin regrouper plusieurs autres classes en les combinant dans une relation d'équivalence. On parle alors de « classe définie ».
  - **Les instances** constituent quant à elles la définition extensionnelle de l'ontologie (c'est-à-dire le domaine de réalité couvert par chaque concept). Elles correspondent à l'ensemble des éléments particuliers, ou « individus », subsumés sous le concept défini en compréhension (ou intension). Ce sont donc elles qui renferment les connaissances qui « peuplent » l'ontologie.
- **Les propriétés** désignent les liens qui articulent les concepts les uns aux autres. C'est ici que la structuration proprement dite des contenus formalisés s'opère et que l'ontologie marque sa différence avec d'autres langages contrôlés tels que les thésaurus ou les taxonomies. La richesse des relations est en effet caractéristique des ontologies et constitue un de leurs apports les plus intéressants.
    - **Les relations hiérarchiques** de subsomption d'une espèce sous un genre (« est\_un », « is\_a ») sont la forme logique des relations linguistiques dites d'hyperonymie et d'hyponymie, communes aux ontologies, thésaurus et taxonomies. Dans la formalisation de l'ontologie en RDF et OWL, on distingue les hiérarchies de classes désignées par la propriété `rdfs:subClassOf` et les hiérarchies de propriétés désignées par la propriété `rdfs:subPropertyOf`. Les hiérarchies peuvent enfin se croiser pour constituer des polyhiérarchies dans lesquelles les instances sont dérivées de plusieurs concepts.
    - **Les relations de partonomie.** Contrairement à la relation de subsomption évoquée

précédemment, la paronymie désigne une relation spécifiquement ontologique qui distingue par conséquent les ontologies des taxinomies et thésaurus. Souvent confondues avec la *subsumption* qui spécifie un rapport d'identité (*is\_a*) cette relation « tout/partie » spécifie un rapport de *composition* entre deux catégories (*is\_composed\_of*). Par exemple, la classe *automobile* peut s'analyser ainsi, en listant l'ensemble de ses composants (*moteur*, etc).

- ➔ Les relations d'associations temporelles. Spécifiques, là encore des ontologies, ce type de relation peut donner lieu à des formalisations particulièrement riches. L'ontologie du CIDOC-CRM propose à cet égard de nombreuses propriétés et entités temporelles.

L'association de ces constituants confèrent aux ontologies leur puissance sémantique : c'est cette architecture formelle, détaillant le sens des concepts et des liens d'une ontologie, qui fournit le matériau sur lequel on pourra effectuer des inférences valides et pertinentes. Un moteur d'inférence pourra aussi bien contrôler la cohérence de la représentation formalisée du domaine et cheminer intelligemment dans les données pour répondre à une requête d'utilisateur.

### 1.3.5. Implémentation informatique

L'enjeu de la structuration ontologique reste l'exploitation des données par des machines. C'est aujourd'hui les technologies développées dans le cadre du Web sémantique qui permettent cette modélisation. En premier lieu, la syntaxe utilisée repose sur les langages à balises, HTML et XML, à partir desquels les données deviennent échangeables. En second lieu, la modélisation s'appuie sur un cadre de description des ressources, le RDF, qui propose une structure en triplets (sujet, prédicat, objet) intégrant des identifiants uniques compréhensibles par les machines (URI).

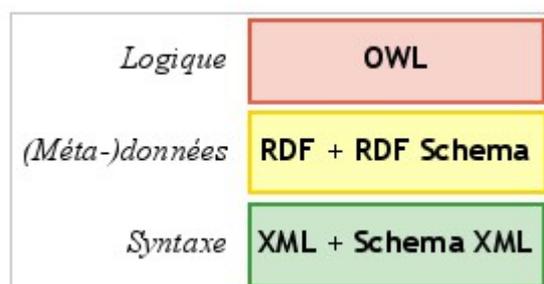


Figure 6. Les trois couches de base du Web sémantique.

En 2001, le W3C s'est engagé dans le développement d'un langage à balise pour les ontologies compatible avec le Web sémantique. C'est ainsi que le langage OWL est apparu et a pu être positionné dans l'architecture générale du Web sémantique.

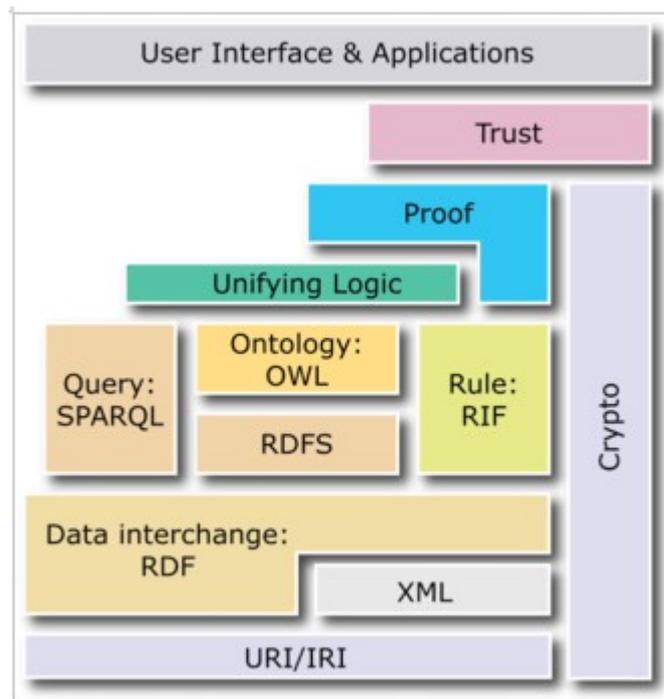


Figure 7. Le « sandwich » du Web sémantique, d'après Tim Berners-Lee (2006)

OWL recouvre aujourd'hui une pluralité de langages : OWL Lite (lightweight ontologies) et OWL Full (heavyweight ontologies) ou encore OWL DL (description logic). Chacun de ces langages permet d'opérationnaliser les ontologies de telle sorte que la machine puisse opérer des raisonnements relativement complexes.

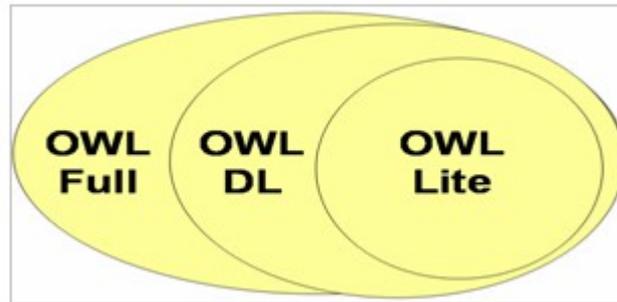


Figure 8. Les trois langages d'OWL

#### 1.4. La construction d'une ontologie : méthodes et étapes

La construction d'ontologie est un processus relativement complexe qui requiert d'adopter un certain nombre de principes méthodologiques, dictant des procédures de travail étape par étape. On distinguera ici les réflexions théoriques, d'ordre linguistiques et épistémologiques, et les approches plus opérationnelles, telles que celles développées notamment par Grüber.

Nous nous appuyerons ici sur les séquences décrites par Charlet, Bachimont et Troncy, dans leur article de 2004, pour la construction d'ontologie de domaine à partir de corpus de connaissances.

La première étape consiste à établir un corpus rassemblant les données du domaine que l'on cherche à modéliser dans l'ontologie. Il s'agit ici essentiellement de corpus textuels en langue naturelle qui explicitent les connaissances du domaine considéré (concepts et relations qui les unissent).

La constitution d'un tel corpus est essentielle, car c'est d'elles qu'émergeront les notions et tout le contenu sémantique à modéliser. On adopte alors une « démarche de corpus » et des outils terminologiques permettant de repérer différentes formes syntaxiques exprimant les concepts du domaine : syntagmes nominaux pour les termes candidats, relations syntaxiques pour identifier des relations sémantiques, etc.

A l'issue de ce travail d'analyse, on dispose alors d'un certain nombre de candidats termes dont il faut vérifier l'univocité. Or, le repérage linguistique n'écarte pas, bien au contraire, la multiplicité

des significations et les ambiguïtés qui en résultent. Exprimées en langue naturelle, les significations dépendent largement d'un contexte particulier d'énonciation, d'usage et d'interprétation, propres aux locuteurs. Or, tout l'enjeu de la formalisation ontologique consiste à dépasser cette polysémie pour aboutir à des primitives univoques, non ambiguës et indépendantes d'un contexte particulier d'interprétation. Cette seconde étape permet d'aborder la dimension proprement sémantique du travail.

Conscient de la difficulté que nous évoquons, Bruno Bachimont propose une « sémantique différentielle » qui permet, à partir du repérage de proximités d'usages, d'aboutir à la construction d'un « arbre de primitives conceptuelles » pour un secteur donné de la réalité. Reprenant une expression de Husserl, Bachimont la nomme une « ontologie régionale ».

## 2. Les humanités numériques : une nouvelle donne pour les ontologies ?

### 2.1. Des ontologies de *quel* domaine ?

#### 2.1.1. Les sciences de la nature à l'honneur. Des outils de modélisations des connaissances essentiellement tournés vers les sciences de la nature et les techniques.

Après avoir retracé la genèse de l'ingénierie ontologique du point de vue de son histoire intellectuelle et technologique d'une part, et de sa conception opérationnelle d'autre part, il importe à présent d'en préciser le champ d'application effectif. Or, à cet égard, il n'est pas indifférent d'observer que les ontologies de domaine, depuis leur origine, se sont prioritairement développées dans le champ des savoirs scientifiques et techniques. On jugera sans doute cette rencontre naturelle. Il existe en effet une conjonction forte entre les sciences expérimentales (et les techniques) et ces outils de modélisation logique des connaissances. Les ontologies de domaines ont en somme investi un champ qui leur était et leur reste favorable à bien des égards :

- Les sciences de la nature (ou sciences expérimentales) sont indéfectiblement liées à l'*environnement technique* qu'elles ont le plus souvent elles-mêmes engendrées et qui, en retour, les rendent possibles. Ce contexte est aujourd'hui dominé par l'informatique, à travers le large éventail de ses applications, dont le traitement automatique des données dans des systèmes à base de connaissances.
- A cette « culture » de l'informatique, correspond une *culture de l'information et de la donnée* qui tend à supplanter le recours traditionnel aux documents et aux œuvres.
- Les sciences et les techniques s'organisent autour de communautés de recherche engagées dans une recherche *collective* de la vérité. Les échanges de savoirs constituent dès lors une nécessité impérieuse qui appelle le développement d'outils interopérables.
- La *profusion* exponentielle des données de la recherche rend indispensable l'adoption de nouveaux outils permettant d'en optimiser l'indexation et l'interrogation *via* des bases de données de plus en plus structurées.

- Le régime de rationalité propre aux sciences de la nature et aux techniques autorise la formalisation logique de leurs contenus : explicatives, et non interprétatives, celles-ci renferment des concepts essentiellement *univoques*, affranchis des langues naturelles et de la variabilité (sociologique ou historique) de leur sens.

### 2.1.2. Les humanités et sciences humaines en reste

Face à cette union naturelle de l'ingénierie ontologique et des disciplines scientifiques et techniques, les humanités et sciences humaines semblent être restées en marge de ces développements. De façon significative, on ne dénombre que très peu d'ontologies dans ce champ du savoir. Cet état de fait peut s'expliquer par des raisons relativement évidentes qui tiennent à l'histoire, à l'épistémologie et aux usages de ces disciplines :

- Les humanités, d'essence et d'origine grecque et romaine, reposent sur le lien insécable de la culture et des *œuvres*. Une œuvre désigne une totalité solidaire de parties indécomposables. De sorte que la notion « atomiste » de donnée ou d'information ne semble avoir aucun sens, ou du moins aucune valeur culturelle.
- Nos humanités modernes (dans le contexte de la fondation des sciences humaines aux XIX<sup>e</sup> et XX<sup>e</sup> siècle) se pensent comme des disciplines ou des sciences de l'*interprétation*. Ce qu'ailleurs on appelle « information » ne désigne jamais pour elles une donnée brute, mais le résultat d'une construction opérée par le travail patient et méthodique de l'interprète : une information particulière n'a de sens qu'à l'intérieur de l'ensemble du document, de l'œuvre et du corpus où elle s'insère.
- Des pratiques de recherche le plus souvent solitaires, qui dérivent de la centralité de la notion d'auteur. Une antinomie subsiste entre l'*autorité* d'un savoir (sa valeur) et son éventuelle construction collective. Le livre prime sur l'article, la production d'un traité ou d'une somme totalisante sur un sujet prime sur la contribution partielle et modeste d'une publication dans une revue, pierre incertaine d'un plus vaste édifice. A la rigueur, il s'agit davantage de faire œuvre, que de faire science.
- Un climat de défiance à l'égard de toute technologie de l'automatisation du raisonnement et du sens, perçue comme une réduction profondément dénaturante de l'homme à la

machine, débouchant, à terme, sur une négation pure et simple de l'objet des sciences humaines.

Ainsi comprises, les humanités semblent donc par nature singulièrement éloignées des enjeux d'une structuration élevée des contenus de savoirs et des outils informatiques d'échanges.

A cet argument de fond, s'ajoute également des contraintes liées à la mise en œuvre relativement coûteuse d'une ontologie en terme de mobilisation de compétences, de ressources humaines et, par conséquent, d'engagement financier.

### **2.1.3. Les humanités numériques : nouvelle donne pour les ontologies ?**

Sur la foi de ce rapide état des lieux, on imaginera difficilement deux champs plus irréconciliables que celui, pluriséculaire, des humanités, fondé sur le paradigme de la culture livresque, étude patiente, érudite et solitaire d'oeuvres-monuments, et l'univers émergent des technologies numériques, marqué par la préséance de l'innovation, la célérité et l'effervescence collective.

Pourtant, à cette nouvelle querelle de l'Ancien et du Moderne, le mouvement des *Digital Humanities* entend aujourd'hui ouvrir la perspective d'une synthèse réconciliatrice. L'essor d'une communauté de recherche et de pratiques en humanités, fécondée par les technologies informatiques et numériques, se veut annonciatrice d'un nouvel âge des sciences de l'homme et des savoirs humanistes, « augmentés » par la puissance computationnelle des ordinateurs et communicationnelle des réseaux.

Pour saisir le sens de ce croisement inattendu des humanités et du numérique et d'en proposer une définition, tâchons d'abord d'en retracer la genèse.

## **2.2. L'émergence des Humanités numériques**

### **2.2.1. Literacy and Linguistic Computing**

On s'accorde généralement (McCarty, Burnard, Dacos, Mounier, Zighed) pour situer l'origine de l'approche des *Digital Humanities* aux années 1940, dans les travaux de Roberto Busa. Cet exégète de l'oeuvre de Thomas d'Aquin fut en effet le premier à s'appuyer sur l'informatique pour proposer

un système de repérage automatique des occurrences du lexique d'une œuvre, et jeter ainsi les bases d'un index intégralement informatisé du corpus thomiste (*Index Thomisticus*). Ce projet inaugural du premier âge des humanités numériques est caractéristique de ce que Lou Burnard nomme « Literacy and Linguistic Computing ». De nombreuses réalisations prolongeront cette perspective, telles que le traitement automatique de la langue (TAL) ou les différents corpus linguistiques de références (le Brown corpus pour la langue anglaise, ou encore la base Frantext dans le contexte francophone).

Cette période voit ainsi s'affirmer une approche quantitative de l'information, qui se traduit par l'essor du traitement calculatoire de sources textuelles (analyses statistiques).

### **2.2.2. Humanities Computing**

Au cours de années 1980, s'amorce une deuxième séquence où la question de la modélisation des connaissances constitue l'enjeu majeur. A l'approche quantitative et calculatoire qui caractérisait essentiellement le premier âge de l'informatique appliqué aux humanités, succède ainsi une perspective plus soucieuse des formes du raisonnement scientifique et de leur modélisation en vue de leur implémentation informatique. L'analyse statistique visant à exploiter de grandes quantités de données cède le pas au travail de mise en forme conceptuelle des différents champs du savoir. L'effort tend alors à affranchir les contenus d'information de leurs contextes linguistiques respectifs tout en levant les ambiguïtés inhérentes aux langues naturelles. Cette approche oriente dès lors les humanités numériques vers l'interdisciplinarité, comme en témoigne l'un des projets les plus significatifs de cette période : la Text Encoding Initiative (1987). En proposant un standard pour l'encodage et la structuration des textes numériques, la TEI constitue en effet un outil adapté à tous types de corpus textuels, quelles que soit leur nature et leur origine disciplinaire. Son importance s'avère donc considérable au regard de l'enjeu que représente la communication des différents savoirs : cette syntaxe universelle constitue un maillon essentiel de l'interopérabilité des données en humanités.

### 2.2.3. *Digital Humanities*

Enfin, avec l'apparition du web, ce qu'on appelle encore « humanities computing » amorcent leur dernière évolution majeure qui débouchera sur l'émergence de ce qu'il est convenu aujourd'hui de désigner sous le nom de « *Digital humanities* »<sup>23</sup>. L'approche computationnelle s'enrichit alors d'une dimension communicationnelle, que met en œuvre le vaste mouvement de diffusion des corpus en ligne, à l'image du William Blake Archive ou du Cranach Digital Archive. De nouvelles questions relatives à la publication des contenus sur le web se posent, engageant de profonds bouleversement sur le terrain de l'édition, du droit, de l'économie, de la conception d'interfaces, ou encore du positionnement des institutions de la bibliothèque ou du musée, confrontées à un environnement qui délocalise les ressources et renouvelle les pratiques de l'utilisateur. De simple *outil* au service des humanités, le numérique devient ainsi progressivement l'*élément* à l'intérieur duquel celles-ci s'exercent, se développent, voire se pensent elle-mêmes.

« quel que soit le sujet d'étude, la recherche s'accomplit désormais systématiquement sur des objets numériques, qu'ils soient le résultat d'opérations de numérisation ou qu'ils soient nativement numériques. Dans tous les cas et quelle que soit leur spécialité, les chercheurs sont aujourd'hui conduits à concevoir, construire, manipuler de grandes bases de données bibliographiques, quantitatives, textuelles, d'images ou de sons ; ils sont conduits à travailler au sein d'équipes multilocalisées et en réseau, souvent à l'échelle internationale ; ils sont conduits à publier les résultats de leurs travaux, que ce soit les données ou les interprétations de ces données, sur le Web ouvert. »<sup>24</sup>

L'essor de cet environnement informatique et numérique est rapidement perçu dans le monde anglophone de la recherche en sciences humaines comme l'occasion de progrès majeurs, tant du point de vue des pratiques des chercheurs que des problématiques de recherche.

Dans ce contexte, les humanités numériques recouvrent essentiellement un mouvement visant à unir dans une vaste communauté de recherche et d'échanges, les différents laboratoires travaillant

<sup>23</sup> La première occurrence de l'expression semble remonter au titre d'un ouvrage de synthèse, *A Companion to Digital Humanities*, paru en 2004 (source : Marin Dacos et Pierre Mounier, *Humanités numériques*, p.12]

<sup>24</sup> Dacos Marin, Mounier Pierre, *Humanités numériques. Etat des lieux et positionnement de la recherche française dans le contexte international*, Institut français, p. 14.

en sciences humaines au moyen d'outils numériques.

Ces nouvelles pratiques apparaissent en France dans la décennie suivante. La TGE Adonis et la TGIR, puis la plateforme Huma-Num (2013), en constituent le principal support institutionnel.

### 2.3. Approche d'une définition

Réalité mouvante et diverse, les Humanités numériques se prêtent difficilement au travail de définition. Toutefois, à défaut d'y saisir de prime abord l'unité d'une signification, on peut y discerner une certaine unité d'orientation. Celle-ci s'articule autour des thèmes suivants :

- conservation, organisation et structuration des données et des corpus, offrant la possibilité d'un traitement intelligent (*i-e* sémantiquement riche) des données (*data mining*)
- édition et diffusion des contenus sous un format numérique sur le web
- création de communautés virtuelles de recherche, d'échange et de débat, s'appuyant sur l'architecture en réseau typique du web

A l'intersection de ces trois axes fondamentaux, on peut tenter de cerner les premiers éléments d'une définition précise des humanités numériques. Djamel Abdelkader Zighed la formule ainsi :

« Les Humanités Numériques sont un cadre méthodologique et technologique qui opère sur des sources de données de SHS et permettant :

- la création, la numérisation et la structuration de toutes les sources de la connaissance
- l'exploration, l'analyse et l'interprétation des informations numériques
- la diffusion, le partage et la capitalisation des connaissances »<sup>25</sup>

En dépit du caractère délibérément ouvert de cette définition, Zighed offre une détermination assez rigoureuse des humanités numériques qui précise :

1. leur nature (une réalité à la fois technologique et méthodologique)

<sup>25</sup> Zighed Djamel Abdelkader, *Les Humanités Numériques en Sciences Humaines et Sociales*, Institut des Sciences de l'Homme, CNRS éditions.

2. leur objet (des sources de données numériques)
3. leur projet (structuration, interprétation, partage des données numériques)

Or, l'élément qui nous paraît ici décisif est contenu dans la notion de « cadre méthodologique ».

Le simple recours, rare ou fréquent, à des outils numériques dans les humanités, ne saurait justifier l'usage de cette appellation qui implique l'idée d'un véritable changement de paradigme en sciences humaines. Forçons le trait : utiliser un service de messagerie électronique sur le web ou un logiciel dans le cadre d'une recherche en humanité, n'autorise en rien à parler d'un quelconque renouveau de leur « cadre méthodologique » (Zighed, ci-dessus). Un nouvel âge ne s'annonce que lorsque le cadre technologique déborde son caractère purement utilitaire et, de simple instrument, devient une véritable matrice de sens, une source génératrice d'hypothèses, de problèmes, de méthodes et de savoirs. Or tel est bien le renversement qu'opère le passage de l'autorité traditionnelle du document au primat info-documentaire de la donnée : les humanités ne deviennent à proprement parler numériques que lorsque le paradigme technologique (info-documentaire) de la donnée féconde (ou corrompt?) la façon dont ce champ du savoir pense son objet et ses méthodes (son épistémologie). Considérer que l'objet des humanités se laisse réduire à des données numériquement structurées, autorisant l'automatisation du traitement et l'échange dans des communautés virtuelles, c'est là s'engager dans le mouvement des humanités numériques.

En somme, les humanités numériques ne consistent pas à faire des humanités *avec* du numérique, mais à faire *numériquement* des humanités.

Ce profond bouleversement des savoirs humanistes correspond bien à ce que Barry Smith propose d'appeler le « tournant computationnel » (« computational turn ») des humanités. Celles-ci se définissent par leur adoption du paradigme numérique et de sa catégorie fondamentale, l'*information*, comprise comme objet de calcul (au sens logique et informatique d'inférence).

S'il en est ainsi, la synthèse du savoir humaniste et de la technologie numérique désigne davantage qu'un point de connexion entre deux champs distincts voire opposés, mais une véritable hybridation dans laquelle les humanités changent pour ainsi dire de nature :

« A tous les niveaux et à toutes les étapes de son déroulement, la recherche en sciences humaines est désormais immergée dans le numérique et ses usages. C'est la différence essentielle qui fait que les humanités numériques

se démarquent des périodes qui ont précédées. L'informatique (computing), outil au service de la recherche est devenu le numérique (digital), véritable environnement global au sein duquel la recherche est effectuée. » (Marin Dacos et Pierre Mounier, op. cit., p.13)

La « digitalisation » des humanités consiste à jeter les savoirs humanistes dans le creuset du numérique afin d'en modifier la forme même, et non d'en élargir simplement le champ d'application.

### 3. Des ontologies pour les humanités : un contresens ?

#### 3.1. L'argument de la *coupure sémantique*.

Le mouvement des humanités numériques se déploie aujourd'hui sur toutes les dimensions du web, conjuguant ainsi l'approche communicationnelle du web social et l'approche sémantique du web des données liées. De ce fait, le devenir numérique des humanités investit un terrain où cohabitent deux dimensions distinctes du sens : une dimension sociale, fortement dépendante des contextes d'énonciation, d'interlocution et des langues naturelles, et une dimension logique et formelle où la signification des concepts demeure indépendante du contexte de leur usage.

Or, les ontologies de domaine relèvent intégralement de cette dernière conception purement référentielle et logique du sens. De sorte que les connaissances qu'elle entreprennent de modéliser doivent pouvoir être ramenées à ce régime de signification. Concevoir des ontologies pour les humanités revient donc à postuler une compatibilité de leurs sémantiques respectives. Or, peut-on représenter à l'aide d'une sémantique logique et formelle les contenus de significations essentiellement mouvants et contextualisés des disciplines interprétatives (humanités) ? N'existe-t-il pas une incompatibilité de principe entre ces deux sémantiques ?

Cette objection qui repose en dernière instance sur un positionnement relatif à l'épistémologie des sciences humaines et des humanités en général, n'a pas manqué d'être soulevée, dans le contexte anglosaxon par Veltman, puis en France par Rastier et plus récemment par Manuel Zacklad.

L'ensemble de ces critiques revient à opposer ce qu'on peut appeler l'argument de la coupure sémantique. Un article de Manuel Zacklad<sup>26</sup> offre à cet égard une présentation claire, synthétique et actualisée de cet argument.

Toute représentation des concepts d'un domaine de connaissance dans une ontologie, repose

---

<sup>26</sup> Zacklad Manuel, « Quelle formalisation pour les contenus culturels ? », *Documentaliste-Sciences de l'Information*. Vol. 48, n°4, p. 40-41

nécessairement, quoique d'une façon souvent implicite, sur une conception référentielle du sens : le sens d'un concept n'est pas à chercher dans le contexte qui en règle l'usage, mais dans sa référence, *i-e* le contenu de réalité qu'il dénote.

On comprend alors pourquoi les ontologies se montrent particulièrement bien adaptées à la représentation du domaine des sciences de la nature et des savoirs techniques, dont les concepts se prêtent aisément à une interprétation référentielle : tous renvoient à un contenu de réalité que l'on peut désigner dans le monde d'une façon univoque et insensible aux variations de contexte. C'est aussi pourquoi les ontologies trouvent un débouché naturel dans les pratiques « logistiques » (Zacklad, *ibid.*).

En levant les ambiguïtés d'un domaine du savoir, les ontologies de domaine permettent un traitement des données à la fois univoque et indépendant de tout contexte d'interprétation, d'énonciation et d'interlocution particulier. Comme le souligne Bruno Bachimont, les énoncés exprimés en langage formel sont « non-contextuels : si, en effet, l'interprétation dépend du contexte, la forme de l'énoncé ne suffira pas à déterminer sa signification. Le fait que la sémantique soit formelle implique que l'interprétation soit la même quel que soit le contexte, et n'est déterminé que par la forme seule. »<sup>27</sup>

La représentation ontologique rabat tout contenu d'information sur l'alternative du vrai et du faux, en excluant *a priori* les modalités du possible ou de l'incertain, n'autorisant dès lors que des « opérations ne nécessitant aucune composante interprétative : des calculs »<sup>28</sup>

En revanche, les concepts mobilisés en humanités offrent rarement une telle univocité et insensibilité aux contextes de signification :

« Quand la relation aux objets et aux personnes ne relève pas de pratiques logistiques liées à leur dimension *matérielle* mais de pratiques liées à la lecture, à l'évaluation qualitative, au raisonnement juridique, à l'histoire, et à la géographie, aux relations sociales, à l'art, aux activités scientifiques dans leurs dimensions inventives, etc., alors

---

27 BACHIMONT Bruno, *Arts et Sciences du numérique : Ingénierie des connaissances et critique de la raison computationnelle*, Mémoire d'habilitation à diriger des recherches, Université de Technologie de Compiègne, 2004.

28 DECLERCK G. et CHARLET J., *Intelligence Artificielle, ontologies et connaissances en médecine. Les limites de la mécanisation de la pensée*, in *Revue des Sciences et Technologies de l'Information - Serie RIA : Revue d'Intelligence Artificielle*, Lavoisier, 2011, 25 (4), pp.445-472. <inserm-00916651>

le niveau de signification des *symboles* utilisés dans le communication n'est plus référentiel mais rhétorico-herméneutique. » (Zacklad, *ibid* ; nous soulignons)

Dire que les humanités ne relèvent pas d'une sémantique référentielle, n'implique nullement qu'elles ne traitent pas du réel et n'en délivrent aucune connaissance. Mais tandis que le scientifique et le technicien envisagent le réel comme un ensemble d'objets *matériellement* localisable et mesurable (« dimension matérielle »), le chercheur en humanité l'envisage comme un ensemble de phénomènes porteur de *sens* (des « symboles »). Or, cet aspect du réel appelle une autre rigueur qui est celle, non plus du mécanicien (au sens scientifique et technique du terme), mais de l'interprète. Le sens qu'il s'agit ici de dégager appelle une approche rhétorique et herméneutique. Son objet est en effet de dégager le sens d'un concept en le resituant d'une part dans son contexte de communication (*rhétorique*) et d'autre part, dans l'ensemble plus vaste du texte, document ou corpus auquel il appartient (*herméneutique*).

Une ontologie d'un domaine des humanités désignerait donc l'hybridation impossible de deux formalités inconciliables : logico-mathématique d'un côté, socio-sémantique de l'autre.

### **3.2. Qu'est-ce que la musique bretonne ? Un exemple de cacophonie des concepts.**

Pour illustrer ce point, Manuel Zacklad nous propose l'exemple particulièrement parlant d'un catalogue musical :

« Pour savoir si « musique bretonne » (terme objet) est bien un « genre musical », « représentatif de » « musique bretonne » (trois termes de la métasémiotique <sup>29</sup>), il faut interpréter les notions de « genre musical » (selon la source sonore, le lieu de destination, la durée du morceau, etc.), de « représentatif » (un exemple, un type majoritaire, un prototype, une tendance actuelle, etc.) et la notion de « musique celtique » (par exemple définition géographique ou ethno-musicologique). L'interprétation peut laisser subsister des ambiguïtés ou des sens multiples qui

---

29 Source du sens des termes utilisés.

pourront être résolus (ou non) en contexte et être considérés, selon les cas, comme une richesse ou un défaut. »

Et de poursuivre :

« « Musique bretonne », le terme objet doit lui-même être interprété selon le contexte ou la situation d'interlocution d'où il émane pour savoir s'il désigne, par exemple, la « musique folklorique bretonne » ou la « musique produite par des artistes bretons ».

Ce cas concret, que l'on peut aisément généraliser, illustre bien la spécificité sémantique des humanités : le sens des concepts qui y ont cours dépend d'interprétations qui véhiculent des intentions diverses possédant chacune sa légitimité. Si cet état de fait conduit à maintenir une certaine ambiguïté des concepts et nourrir des conflits d'interprétations, on ne saurait toutefois les supprimer sans faire preuve d'arbitraire. De ce point de vue, adopter l'exigence d'une sémantique purement référentielle visant à lutter contre toute variété des significations d'un concept (« désambiguïser »), ne serait-ce pas commettre un important contresens quant à la nature des concepts en humanités ?

### 3.3. La *coupure* doit-elle être *tranchée* ? Deux versions de l'argument

Ces courts extraits de l'article de Manuel Zacklad nous permettent ainsi de formuler le problème que renferme le projet d'ontologies pour les humanités (contradiction entre deux registres sémantiques) que l'auteur ramène à la question suivante : la multiplicité des significations d'un même concept en humanités constitue-t-elle « une richesse ou un défaut ? ». Des ontologies en humanités auront-elles pour vertu de corriger une sémantique défectueuse parce que flottante et ambiguë, ou bien au contraire tâcheront-elles de relever le défi de représenter une sémantique plus riche, complexe et variable dans le temps et l'espace qu'à l'accoutumée (registre des sciences de la nature et des techniques ou un terme renvoie à un concept univoque et stable) ?

Les contempteurs des ontologies en humanités, au rang desquels Zacklad semble se compter lui-même, développent une **version forte** de l'argument de la coupure sémantique, qui oppose un refus de principe des ontologies pour les humanités (utopie « techno-centrée » dit Zacklad) et, face au danger d'une dénaturation de la sémantique spécifique des humanités, plaide, selon l'expression de Rastier, « pour une sémiotique sans ontologie ».

Mais une **version faible** du même argument est possible. Car la polysémie inhérentes à la plupart des concepts en humanités ne désignent pas une **imprécision sémantique**, mais plutôt une **variété de significations**. Aucun savoir qui se voudra rigoureux ne peut s'édifier sur une quelconque imprécision sémantique. Si les humanités enveloppaient un tel flou conceptuel, il y aurait lieu de les affliger du peu flatteur (et peu rigoureux, voire contradictoire !) qualificatif de « sciences molles ». Or, tel n'est pas le cas. Ce que l'on observe, c'est une variété de significations : l'hésitation réside dans le choix de l'interprétation, non dans le sens lui-même. L'impression de flou notionnel vient de la variété des concepts entre lesquels on ne sait pas choisir. Il ne faut pas confondre multiplicité de significations et absence de signification.

De ce point de vue, la coupure sémantique existe, mais loin de ruiner l'entreprise des ontologies du domaine des humanités, elles en préciseraient la portée exacte : non bâtir une représentation transparente, définitive et neutre sur le plan des choix sémantiques et logiques, mais une représentation indexée sur des finalités particulières, engageant des partis-pris interprétatifs dûment assumés. Des ontologies pour les humanités deviendraient dès lors à la fois possibles et légitimes dans la mesure où elles ne se fonderaient pas sur l'illusion d'un absolu du sens : les ontologies sont et demeurent des systèmes contextualisés.

### **3.4. A tower of Babel problem (Barry Smith)**

L'argument de la coupure sémantique sonne-t-il le glas du projet d'ontologies pour le humanités ou en précise-t-il le sens afin de l'orienter adéquatement ? Faut-il congédier l'ingénierie ontologique dans le domaine des humanités ou au contraire encourager son développement moyennant une adaptation de ses méthodes au champ qu'elle ambitionne de représenter ?

A l'épreuve des humanités, l'ingénierie ontologique doit se poser la question suivante : comment une représentation logique (informatique et documentaire) des savoirs humanistes est-elle possible, compte-tenu de la sémantique singulière qui est la leur (rhétorico-herméneutique) ?

Ce moment de problématisation appelle à présent une confrontation avec la réalité concrète des ontologies en humanités, afin d'en examiner la structure, les choix de méthodes, les apports et les limites.

## Partie 2. Des ontologies pour les humanités numériques

« La signification d'un mot est son *usage* dans le langage »

Ludwig Wittgenstein

*Investigations philosophiques*, § 43

# 1. Emergence du thème des ontologies en humanités numériques dans la littérature spécialisée en ligne : données statistiques et analyse

## 1.1. Introduction : objet de l'étude et questionnement

Après avoir examiné la *notion* d'ontologie de domaine en la rattachant à son histoire intellectuelle et aux principes de sa construction, il importe à présent d'en saisir la *réalité* concrète.

Les trois études qui suivent réunissent un ensemble de données factuelles, issues de trois sources de données différentes, relatives au nombre et à la nature des contenus spécialisés en ligne traitant du thème des ontologies de domaine pour les humanités numériques. Les questions suivantes nous serviront de fil directeur pour l'analyse :

- 1) Les développements théoriques et pratiques autour des ontologies de domaine pour les humanités numériques constituent-ils un *phénomène d'ampleur* ?
- 2) Quelle *évolution* générale traduit la publication de contenus spécialisés en ligne à leurs sujets ?
- 3) Quelle est la *nature* de ces contenus ? Quelle part les réflexions théoriques (recherche, prospective) occupent-elles par rapport aux développements d'ordre pratique (projets, retour d'expériences, etc.) ?
- 4) Quelle est l'*origine* de ces contenus ? Qui sont les acteurs à l'initiative des réflexions et développements sur les ontologies de domaine pour les humanités numériques ? S'agit-il de chercheurs en sciences ou humanités, de professionnels de la documentation et des bibliothèques, d'informaticiens ?
- 5) Quels *domaines* des humanités apparaissent les plus représentés ?

Une réponse étayée à ces questions et hypothèses implique de recueillir des données à la fois quantitatives (statistiques) et qualitatives (nature et origine des articles).

## **1.2. Etat des lieux de la publication des contenus scientifiques traitant des ontologies info-documentaires dans Google Scholar**

### **1.2.1. Source de données**

Lancé par la société Google en 2004, Google Scholar est un outil de recherche de contenus scientifiques (articles, livres) référencés sur le web, publiés en ligne par des éditeurs scientifiques, des universités ou des organismes de recherches. Cette source de données présente un certain nombre d'avantages :

- le champ documentaire qu'elle recouvre est particulièrement large, de sorte qu'il est possible d'effectuer des recherches sur un très grand nombre de ressources en ligne
- la recherche, multilingue, porte sur l'intégralité du texte des ressources (titre et/ou texte)
- la littérature « grise », qui n'a pas fait l'objet de publication, apparaît dans les résultats des recherches
- le contenu des ressources en accès libre y est directement consultable par téléchargement
- des fonctions de recherche avancées permettent d'appliquer des critères de filtrage pour plus de pertinence des résultats
- des données chiffrées facilitent l'étude statistique des résultats d'une recherche
- son utilisation est gratuite

Toutefois, Google Scholar est un outil limité à plusieurs égards :

- il ne couvre pas la totalité des contenus scientifiques disponibles en ligne
- les critères de référencement utilisés restent opaques
- Google demeure un acteur économique du web qui, à ce titre, doit adosser son projet (réel ou fictif) de promotion des savoirs, à une stratégie financière qui demeure, en tant que telle étrangère, sinon incompatible, à l'approche désintéressé des sciences.

### **1.2.2. Méthode : Choix des mots-clés et critères de recherche**

L'objet de notre recherche pose une difficulté en raison d'une polysémie présente dans toutes les langues : le terme « ontologie » renvoyant tantôt à la philosophie, tantôt à l'informatique et aux sciences de l'information. Il convient donc d'étoffer l'expression afin d'écarter les contenus traitant de l'ontologie au sens philosophique du terme (voir ci-dessous).

Nous avons par ailleurs choisis de formuler nos requêtes en anglais. Ceci pour deux raisons : la recherche internationale publie majoritairement ses articles dans cette langue, et une deuxième étude, rapportée plus loin dans ce chapitre, nous permettra de sonder l'état des publications dans le domaine spécifiquement francophone.

Ceci étant, les expressions que nous avons retenues sont les suivantes :

- « digital humanities » ET « ontology » : le terme « ontology » associé à l'expression « digital humanities » écarte *ipso facto* les contenus d'ordre philosophiques.
- « digital humanities » : afin d'effectuer une comparaison entre l'état des publications sur les Digital Humanities et celle consacré spécifiquement aux ontologies de domaine.
- « domain ontology » ET « humanities » : ce croisement permet de cibler spécifiquement les ontologies de domaine et conjurer en partie les usages impropres ou ambigus du terme ontology dans la littérature informatique ou documentaire.
- « domain ontology » : dresser un état des lieux des publications sur les ontologies de domaine, afin d'identifier, par contraste, la part qu'y occupent les ontologies de domaine en humanités.

### **1.1.3. Résultats : données quantitatives et commentaires**

Le tableau suivant (figure 9) met en évidence l'ampleur et l'évolution du phénomène des publications spécialisées sur les ontologies et les *Digital Humanities*. Si les premiers contenus apparaissent en 1990, il faut attendre le début des années 2000 pour observer une hausse sensible des publications qui rapidement épouse une courbe de croissance proche d'une exponentielle. Ainsi, en l'espace d'une quinzaine d'années (2001-2015), le nombre de publications est passée de moins de 10 à plus de 800.

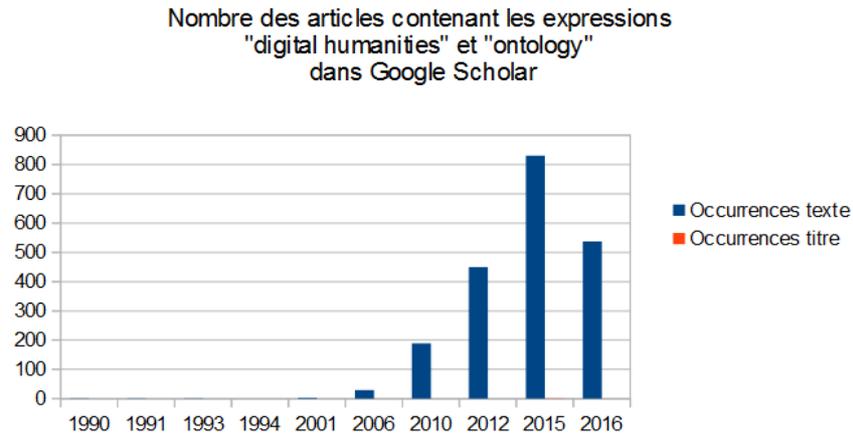


Figure 9. Nombre des articles contenant les expressions « digital humanities » et « ontology » dans Google Scholar, par année, de 1990 à août 2016.

Une évolution de profil analogue se dessine dans le tableau suivant (figure 10) pour les ontologies de domaine (« domain ontologies »). De façon assez nette toutefois, le volume des publications traitant des ontologies de domaine reste inférieur à celui des ontologies en général associé aux humanités numériques. Autrement dit, même sans restriction de contexte, les publications sur les ontologies de domaine restent inférieures à celles portant sur les ontologies dans le contexte des *Digital Humanities*.

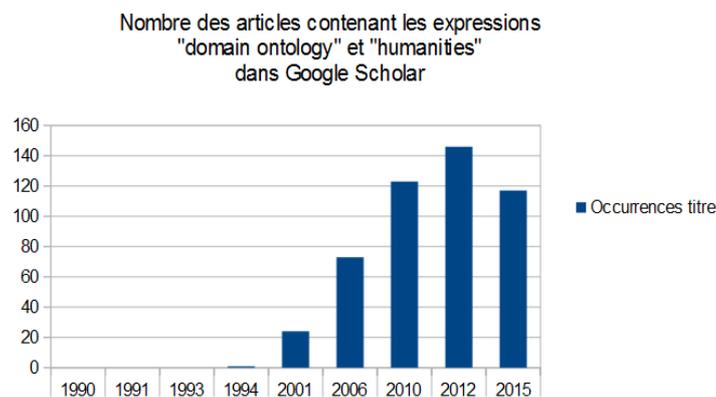


Figure 11. Nombre des articles contenant les expressions « domain ontology » et « humanities » dans Google Scholar, par année, de 1990 à août 2016.

Pour apprécier à sa juste mesure l'ampleur et l'évolution identifiée précédemment, il convient de la superposer avec celle des *Digital Humanities* en général. Toute proportion gardée en terme de volume, on remarque que les deux évolutions suivent une même courbe ascendante : après un moment de relative stagnation au début des années 1990, le volume des publications augmente

subitement à partir de la deuxième moitié de cette décennie passant d'une centaine de contenus en 1994 à près de 4000 dès l'année 2012. Cette évolution spectaculaire est à mettre en perspective avec les événements jalonnant l'histoire des *Digital Humanities* et leur inscription dans le paysage de la recherche et institutionnel. De ce point de vue, la courbe représentée dans le diagramme ci-dessous atteste de la vigueur du mouvement des *Digital Humanities*, bien que l'évolution semble marquer à partir de 2012 un ralentissement voire une stagnation des publications en la matière.

La croissance des articles consacrés aux ontologies dans le contexte des humanités numériques est cohérente avec l'évolution de la publication en humanités numériques, moyennant une différence quantitative d'environ 1 pour 5 et un léger « retard » du thème des ontologies par rapport à l'émergence et l'essor des *Digital Humanities*.

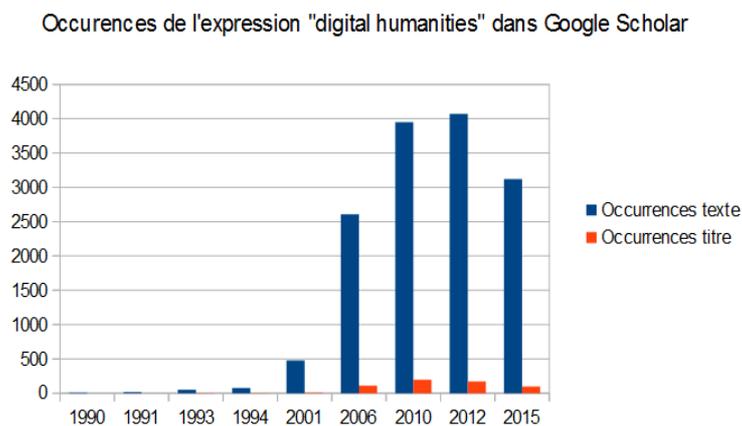


Figure 10. Occurrences de l'expression « digital humanities » dans Google Scholar, par année, de 1990 à août 2016.

Ces éléments superposés les uns aux autres, nous permettent la synthèse suivante, récapitulée dans le tableau ci-dessous (figure 11). L'écart entre le volume des publications consacrées aux ontologies en général et le volume des publications traitant des ontologies de domaine dans le contexte particulier des humanités numériques, se creuse d'abord dans la première décennie des années 2000, pour diminuer par la suite : tandis que la masse des publications sur les ontologies de domaine décline sensiblement entre 2012 et 2015, le nombre des contenus publiés sur les ontologies dans le contexte des humanités numériques atteint son niveau le plus haut. Les publications sur les ontologies en humanités numériques occupent ainsi une part nettement plus importante à partir de 2015 qu'auparavant. Les humanités numériques semblent dès lors avoir

emboîté le pas des autres secteurs de la connaissance en terme de développement sur les ontologies.

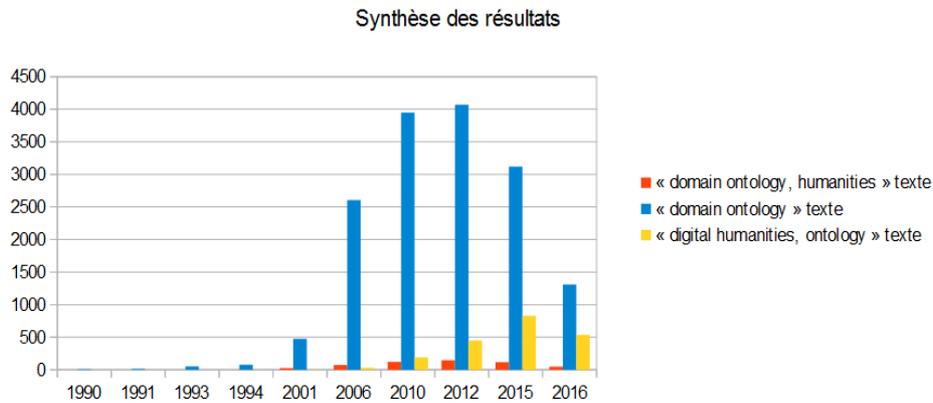


Figure 12. Synthèse des résultats des recherches sur Google Scholar.

A titre de comparaison, on pourra rapprocher les résultats précédents du nombre d'articles référencés sur Google Scholar traitant des ontologies pour le domaine bio-médical, synthétisé dans le graphique ci-dessous. Ce champ du savoir et de la recherche scientifique semble avoir en effet occupé une place centrale dans le développement des ontologies de domaine, à l'image de la GeneOntology, projet pionnier et abondamment commenté en la matière<sup>30</sup>.

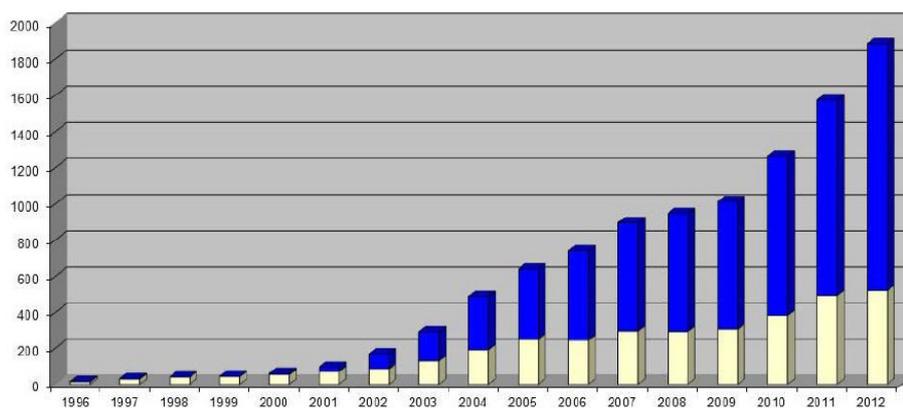


Figure 13. Nombre d'articles sur l'ontologie ou les ontologies dans PubMed/MEDLINE (GO -Gene Ontology- en bleu foncé, les autres ontologies en jaune clair). Source : Barry Smith, The relevance

<sup>30</sup> Voir par exemple Stevens, Noble et Bechhofer, *Ontology-based knowledge representation for bioinformatics*, in *Brief Bioinformatics*, 1: 398–414, PubMed, 2000.

of philosophical ontology to information and computer science, *In Philosophy, Computing and Information Science, Chatto et Pickering, Ruth Hagengruber & Uwe Riss (éditeurs), pp. 75-83, 2014*

### **1.3. Etat des lieux de la publication des contenus scientifiques traitant des ontologies info-documentaires dans les HAL et HAL-SHS.**

#### **1.3.1. Source de données**

Hyper Article en Ligne (HAL) est un service de dépôt et de consultation d'articles de recherche en ligne créé par le CCSD (Centre pour la Communication Scientifique Directe) du CNRS en 2001. Cette plate-forme d'archives ouvertes accueille de très nombreuses disciplines du domaine universitaire et des grandes écoles françaises. Depuis un protocole signé en 2006, les HAL sont adoptées par la plupart des établissements d'enseignement supérieurs et de recherche français comme une référence privilégiée en matière de publication en ligne.

Cette base de données ouvertes constitue dès lors une source fiable du point de vue de la valeur scientifique des contenus référencés (bien qu'ils n'aient pas nécessairement reçu le *satisfecit* d'un comité de lecture), et représentative de l'état de la recherche compte tenu de l'adhésion massive des établissements de recherche dont ce service fait l'objet.

#### **1.3.2. Méthode : critères de recherches et mots-clés retenus**

Pour mener à bien notre étude, nous avons choisi de resserrer notre recherche aux seuls titres des ressources (champ « Titres »). Cette restriction offre plusieurs avantages :

- Pertinence des contenus : on s'assure que les contenus retournés par le moteur de recherche traitent en propre des ontologies, et non de façon secondaire ou allusive.
- Possibilité d'une analyse qualitative plus fine compte-tenu du nombre moins abondant de contenus référencés à examiner.
- Le nombre de contenus retournés nous est apparu suffisant pour procéder à une étude significative.
- L'évolution générale de la publication de ces contenus s'est avérée recouper l'évolution

identifiée avec Google Scholar sur un panel beaucoup plus important et ouvert. Dans cette mesure, nous tenons donc notre échantillon pour représentatif.

Nous avons choisis d'effectuer notre recherche sur le terme « ontologies », en éliminant les mots-clés associés non pertinents parmi la liste indiquée par HAL. En retenant « ontologies », « ontologie » et « ontology », nous sommes ainsi parvenus à affiner nos résultats en fonction de notre objet d'étude. Seules 18 ressources sur 169 traitent de l'ontologie au sens philosophique. Nous avons donc pu aisément les écarter dans le cadre de notre étude.

Nous avons enfin décoché les Notices et Annexes pour ne retenir que les Documents.

### **1.3.3. Résultats : données quantitatives et commentaires**

Le moteur des HAL nous a retourné 169 résultats, dont 99 en français et 70 en anglais. Seuls 18 parmi les résultats en langue française sont d'ordre exclusivement philosophiques ou littéraires.

La répartition de ces contenus par date de publication nous permet d'en percevoir l'évolution générale (voir tableau ci-dessous) :

La tendance d'ensemble fait apparaître une croissance des contenus relatifs aux ontologies entre 2000 et 2015.

- de 2000 à 2003 : de 1 à 2 contenus sont publiés par année, pour un total de 5 en l'espace de 4 ans, soit une moyenne de 1,7 ressource par an.
- de 2004 à 2008 : on repère une première hausse remarquable du nombre des contenus traitant des ontologies, avec une publication de 7 à 14 ressources par an, pour un total de 47, soit une moyenne de 9,4 contenus annuels sur 5 ans.
- De 2009 à 2013 : une période de publication particulièrement abondante, avec un minimum de 12 ressources par an et un maximum de 26 atteint en 2012, soit une moyenne de 19,6 contenus par an sur 5 ans.
- 2014 et 2015 : le nombre des ressources publiées oscille entre 8 et 10 par an, accusant un léger recul qu'on tiendra cependant, faute d'une meilleure visibilité, pour non significatif en

raison de l'absence de congrès nationaux ou internationaux, occasion dans les périodes précédentes des plus abondantes phases de publications.

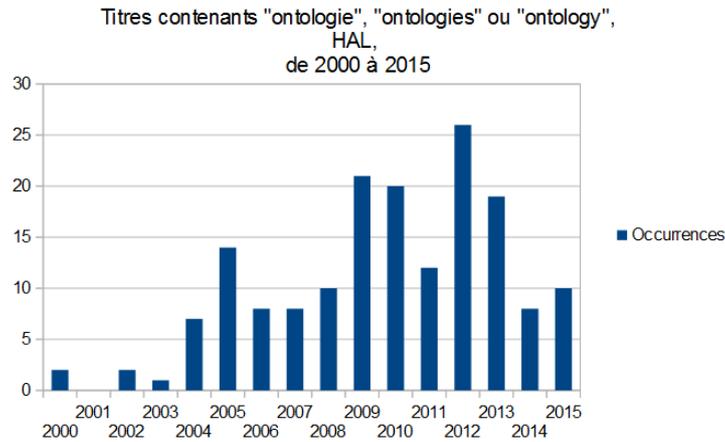


Figure 14. Diagramme représentant le nombre de documents contenant l'un des mots « ontologie », « ontologies » ou « ontology » dans leurs titres, par année depuis 2000. Source : HAL.

Une analyse plus précise nous révèle en effet que les années les plus fastes pour la publication de ressources correspondent essentiellement à la tenue de congrès nationaux ou internationaux initiés par le secteur de la recherche en informatique. Si l'on ne compte qu'un congrès pour l'année 2005 (les 16èmes Journées francophones d'Ingénierie des Connaissances, à Nice en mai 2005), on en dénombre 15 en 2009 se déroulant essentiellement en France, mais également aux Etats-Unis, au Royaume-Uni, en Italie et en Tunisie.

L'allure non linéaire de l'évolution indiquée dans le tableau ci-dessus doit donc être relativisée : l'occasion d'un congrès rend simplement plus visible les résultats de recherches actives en amont, bien que moins visibles. Les hausses soudaines, représentées par des pics dans le diagramme, portent donc témoignage d'une certaine vitalité de la recherche. Avant d'être des causes d'un développement de la recherche, les congrès en sont d'abord la conséquence.

On peut donc conclure à une augmentation régulière de la recherche sur les ontologies dans le contexte français. Toutefois, d'assez fortes disparités existent quant aux secteurs de la recherche à l'origine des publications référencées.

Le diagramme suivant (figure 15) fait apparaître une prépondérance très nette du secteur de l'enseignement et de la recherche en informatique, responsable de près des  $\frac{3}{4}$  des ressources publiées sur HAL.

Un dernier quart se divise en deux secteurs égaux : les Sciences de l'Homme et de la Société, et les sciences en générales, représentées dans le diagramme ci-dessous par les Sciences du Vivant et les Sciences de l'Ingénieur.

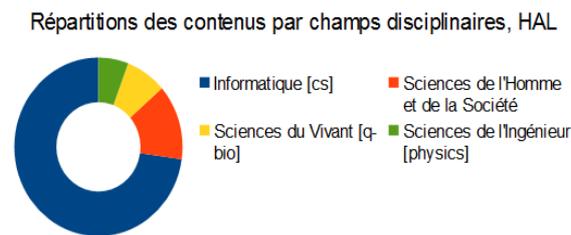


Figure 15. Répartition des contenus comportant le mot « ontologie », « ontologies » ou « ontology » dans leurs titres, par champ disciplinaire. Source : HAL.

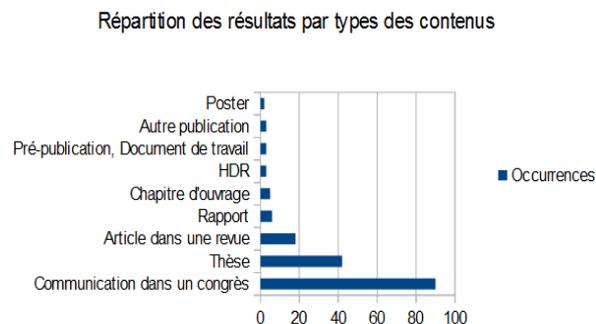


Figure 16. Répartition des résultats obtenus par types de contenus. Source : HAL.

Dans ce diagramme (figure 16), nous représentons les différents types de ressources en fonction de leur nombre dans la base des HAL. L'importance très nette des communications de congrès et des thèses témoigne de la vitalité des recherches relatives aux ontologies, sur un plan à la fois théorique (thèses) que professionnel et interdisciplinaire (congrès).

### 1.3.5. Focus sur les résultats classés SHS dans notre recherche sur les HAL

On remarque d'abord (figure 17) que l'évolution du nombre de ressources traitant des ontologies classés SHS est cohérente avec la tenue des différents congrès. On observe en effet une hausse systématique des articles aux dates des congrès, qui traduit, sinon une participation effective, du moins un écho des congrès dans les sciences humaines.

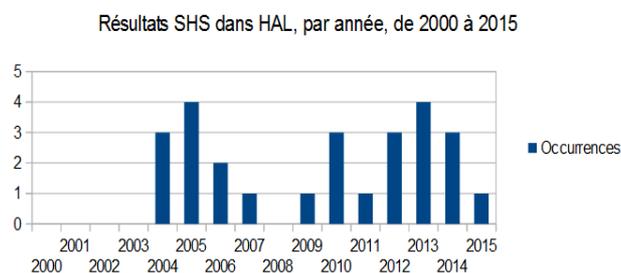


Figure 17. Répartition des résultats obtenus classés SHS dans HAL, par année, de 2000 à 2015.

Source : HAL.

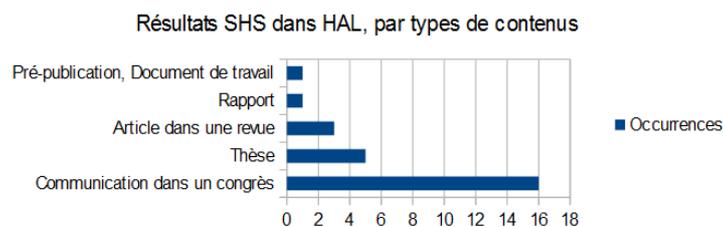


Figure 18. Répartition des résultats obtenus classés SHS dans HAL, par types de contenus. Source :

HAL.

Enfin, le décompte des ressources de ce panel de résultats traitant cette fois explicitement d'ontologie de domaine, nous permet d'identifier les champs de savoir ou de techniques visés par la modélisation ontologique. Les trois domaines les plus représentés sont les suivants :

- Biologie, médecine et santé : 25 occurrences

- Données géographiques : 7 occurrences
- Ingénierie industrielle : 3 occurrences

Ces données factuelles confirment le constat généralement exprimé dans la littérature spécialisée que les domaines prioritairement ou majoritairement modélisés par des ontologies ont en commun :

- d'intégrer une plus ou moins grande part de science expérimentale,
- de renfermer un volume très important de données factuelles,
- d'entretenir un lien essentiel avec un domaine technique.

### 1.3.6. Bilan

L'examen de la publication de contenus sur les ontologies info-documentaires sur les HAL et HAL-SHS permet d'identifier :

- *l'émergence et l'évolution* du phénomène : apparu au début des années 2000, les publications n'ont cessé de croître à raison d'environ 50 % par an entre 2000 et 2015.
- *l'origine* des ressources : l'analyse des principaux secteurs à l'origine des travaux référencés marquent une assez nette avance de l'informatique au travers des centres de recherche tels que les différents sites de l'INRIA.
- *la nature* des contenus publiés s'avère quant à elle plus souvent théorique que pratique et directement opérationnelle : les considérations relatives aux finalités, méthodes et enjeux des ontologies sont plus nombreuses que les analyses de projets concrets par le biais de documents de travail ou de retours d'expériences.
- *les domaines modélisés* : Enfin, les domaines du savoir visés par la structuration ontologique appartiennent majoritairement aux sciences expérimentales appliquées telles que la biologie et la médecine, les savoirs techniques et les secteurs des sciences humaines intégrant des données factuelles importantes (telle que la géographie). Les disciplines

interprétatives et littéraires sont à cet égard moins représentées.

## 1.4. Etat des lieux de la publication de contenus traitant des ontologies info-documentaires dans le secteur des bibliothèques et de la documentation

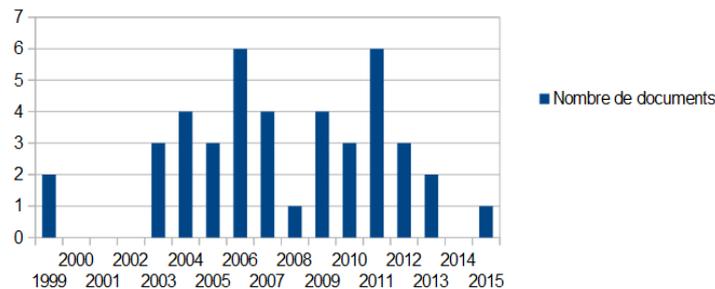
### 1.4.1 Le Bulletin des Bibliothécaires de France (BBF)

Le Bulletin des Bibliothécaires de France

recherche avancée : tous les mots « ontologies »

30 résultats, contre 380 pour les mots « thésaurus » et « taxonomies ».

Nombre de documents contenant le mot "ontologies"  
publiés dans BBF,  
entre 1999 et 2015



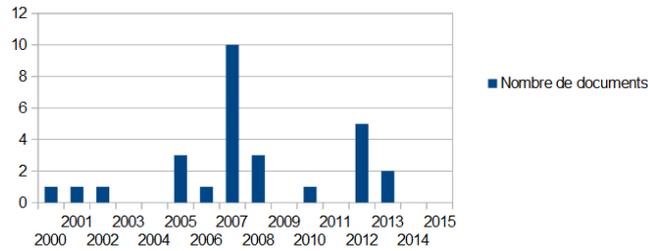
### 1.4.2 Documentaliste - Sciences de l'information

éditée par l'ADBS.

Recherche avancée, Un article de revue, Titre de l'article « ontologies »

22 résultats (36 – 14 traitant de « déontologie »)

Nombre de documents contenant le mot "ontologies"  
publiés sur Documentaliste - Sciences de l'information,  
de 2000 à 2015



### 1.4.3. Archimag

On recense 2 documents en 2015 traitant d'ontologie, contre 14 sur les thésaurus et 4 sur les taxonomies la même année:

- *Etre documentaliste aujourd'hui : évolutions du métier et perspectives d'emploi*, 8 avril 2015. Il s'agit d'un entretien avec Jean-Philippe Accart. Les ontologies y sont évoquées au côté des thésaurus et taxonomies, pour illustrer la nécessaire évolution technologique des métiers de la documentation. Accart y évoque la figure de l'« architecte de l'information », compétent, par exemple, dans les outils de traitement « sémantique » des données.
- *Trouver thésaurus, taxonomie ou ontologies à son pied*, Bruno Texier, 21 septembre 2015 . Prenant acte de l'importance croissante des vocabulaires contrôlés, l'auteur fait le point sur les thésaurus, taxonomies et ontologies, outils devenus « incontournables pour les experts du document » face au défi représenté par un « océan documentaire » grandissant.

## 2. Quelques ontologies du domaine des humanités : examen et appréciation

### 2.1. L'Ontologie Wittgenstein de l'Université de Bergen

Pour aborder la réalité concrète des ontologies de domaine pour les humanités numériques, nous proposons dans un premier temps d'examiner une ontologie modélisant l'œuvre et la pensée du philosophe autrichien Ludwig Wittgenstein (1889 – 1951) élaborée à partir de 2009 par l'équipe des Wittgenstein Archives of Bergen (WAB) de l'Université de Bergen (Norvège)<sup>31</sup>. Nous y serons particulièrement attentif à la façon dont ses concepteurs ont perçu la difficulté de traduire un contenu essentiellement polysémique et ouvert à une grande variété de points de vue et d'interprétations également légitimes, dans la structure logique d'une ontologie, fondée sur une exigence d'univocité et de complétude logique.

Pour rendre compte de la construction de cette ontologie, nous nous appuyerons ici sur quatre articles faisant état des choix opérés par ses deux principaux auteurs, Amélie Zöllner-Weber et Aloïs Pichler (de l'Université de Bergen).<sup>32</sup>

Après une introduction retraçant le contexte du projet de la *Wittgenstein Ontology*, nous tâcherons de restituer les principaux choix ayant présidés à sa conception, pour mesurer, dans un dernier temps, l'intérêt des solutions envisagées au regard du problème qui nous occupe.

---

31 The Wittgenstein Ontology : [http://wab.uib.no/wab\\_philospace.page/wittgenstein.owl](http://wab.uib.no/wab_philospace.page/wittgenstein.owl)

32 PICHLER, A., ZÖLLNER-WEBER, A., *Sharing and debating Wittgenstein by using an ontology*, Literary and Linguistic Computing, Vol. 28, No. 4, 2013. Downloaded from : <http://llc.oxfordjournals.org/>  
 PICHLER, A., ZÖLLNER-WEBER, A. *Towards Wittgenstein on the Semantic Web*. In : *Digital Humanities 2012 Conference Abstract*. Hamburg University Press, 318-321.  
 PICHLER, A. *Wittgenstein Ontology. Technical Report*. University of Bergen, 2013.  
[http://wab.uib.no/wab\\_philospace.page](http://wab.uib.no/wab_philospace.page)  
 MACHA, Jakub, FALCH, Rune J., PICHLER, Aloïs. *Overlapping and Competing Ontologies, DH-CASE'13, proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities*, Article No. 5, *ACM* New York, NY, USA ©2013 ISBN: 978-1-4503-2199-0 doi>[10.1145/2517978.2517984](https://doi.org/10.1145/2517978.2517984). Downloaded from [http://cis.uni-muenchen.de/kurse/max/scholarship/Overlapping and Competing Ontologies.pdf](http://cis.uni-muenchen.de/kurse/max/scholarship/Overlapping%20and%20Competing%20Ontologies.pdf)

### 2.1.1. Contexte du projet : pourquoi une ontologie ?

Le projet d'une ontologie modélisant la philosophie de Wittgenstein s'inscrit d'abord dans la continuité de plusieurs initiatives de numérisation des archives des œuvres du philosophe par l'Université de Bergen. S'emparant des quelques 20 000 pages de manuscrits/tapuscrits du Wittgenstein *Nachlass*, propriété de l'Université de Oxford, l'équipe des Wittgenstein Archives en entreprend en 1990 l'édition numérique *via* son encodage TEI et en publie près de 5000 pages en accès libre dès 2009. Fort du succès de cette base en ligne, abondamment consultée depuis lors par les chercheurs du monde entier, l'Université de Bergen entend alors franchir un pas supplémentaire en proposant d'accéder non plus seulement aux *documents* constituant l'œuvre de Wittgenstein, mais aux *concepts et arguments* produits et discutés par le philosophe autrichien. D'une démarche documentaire, l'équipe norvégienne bascule alors vers une approche proprement sémantique des contenus de la *Wittgenstein Source*. Cette évolution s'explique par la volonté de constituer un outil de référence pour l'étude et l'enseignement de l'œuvre et de la philosophie de Wittgenstein, appelée par les besoins d'une communauté de recherche vivante, mondiale et particulièrement marquée par la pratique du débat argumentatif. Alois Pichler et Amélie Zöllner-Weber, principaux architectes du programme, formulent ainsi un ambitieux projet visant à « partager et débattre sur Wittgenstein ». L'enjeu, de portée proprement scientifique, témoigne d'une adhésion de l'équipe de l'Université de Bergen à une version forte des humanités numériques, comprises comme moteur de la recherche, et non simple courroie de transmission de savoirs disciplinaires au sein d'un espace numérique.

Pour opérer ce changement de paradigme, le développement d'une ontologie apparaît nécessaire dans la mesure d'abord où celle-ci permet une exploration par concepts plutôt que par chaînes de caractères. L'architecture ontologique assure en effet une cohérence d'ensemble qui permet à un utilisateur de naviguer dans la philosophie de Wittgenstein en parcourant les liens unissant ses différents concepts et ainsi être en mesure de les discuter.

Le projet de l'université de Bergen entend donc tirer parti des deux possibilités majeures offertes par une ontologie de domaine : étendre l'accès à un contenu de savoir par un outil interopérable, et approfondir son exploration par inférences.

### 2.1.2. Pourquoi une *nouvelle* ontologie ?

Pour modéliser la pensée de Wittgenstein, il eût certes été possible de s'appuyer sur des ontologies existantes et parfaitement adaptées aux exigences propres des humanités. GOLD a ainsi par exemple été développée pour modéliser des éléments de linguistique, tels que la syntaxe ou la morphologie des textes. WordNet fournit quant à elle une base de données lexicale de l'Anglais qui se présente comme une ontologie. Les chercheurs de l'équipe des Wittgenstein Archives de Bergen ont cependant remarqués, semble-t-il à juste titre, qu'une approche purement linguistique de la pensée de Wittgenstein ne permet pas d'en restituer le caractère conceptuel. De son côté, l'ontologie du CIDOC CRM présente elle aussi une insuffisance majeure. La description des données du patrimoine culturel y fait l'objet d'une approche centrée « événement » : on s'attache à décrire des relations de temps et d'espace. Or, la modélisation d'une œuvre et d'une pensée philosophique appelle un tout autre traitement, centré sur les « objets » tels que des documents, des concepts et des arguments (Pichler et Zöllner-Weber, *ibid.*).

Il importait donc de concevoir une nouvelle ontologie *ad hoc* pour modéliser au plus près la pensée de Wittgenstein en en restituant à la fois l'évolution et les interprétations issues de ses divers commentaires et controverses.

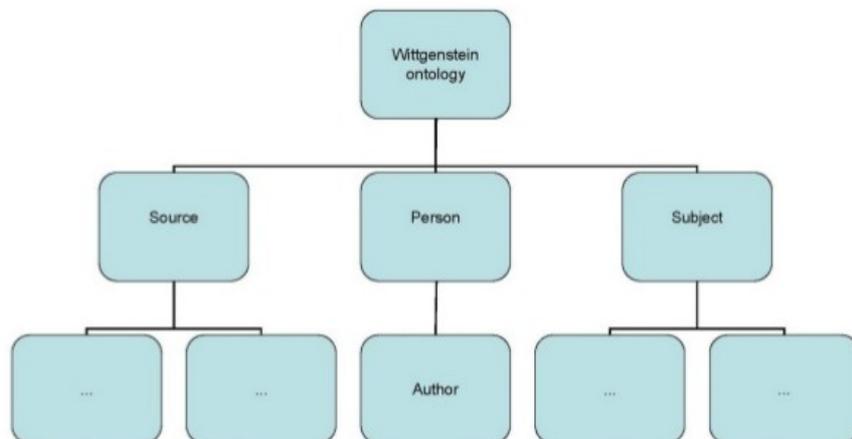
### 2.1.3. Conception de l'ontologie

- **Contraintes et périmètre**

Si la Wittgenstein Ontology a pour but d'élargir l'accès à la philosophie de Wittgenstein et de nourrir la discussion argumentative, Pichler et Zöllner-Weber lui assignent toutefois deux contraintes supplémentaires qui déterminent fortement les choix présidant à sa construction. L'ontologie doit d'abord constituer une « modélisation contrôlée par des experts » (*op. cit.*) afin d'assurer une base commune fiable et solide pour les échanges universitaires. Elle doit ensuite s'inscrire dans le volet édition numérique des Wittgenstein Archives de Bergen. La première exigence implique de constituer le corpus initial à l'intérieur du périmètre des références spécialisées et de l'expertise des chercheurs ; la seconde confirme la nécessité d'une modélisation

orientée objet.

▪ **Structure**



*Figure 18. Structure générale de la Wittgenstein Ontology. Source : Pichler Aloïs, The Wittgenstein Incubator and Swicky Notes, en ligne à l'adresse <http://www.slideshare.net/DM2E/berlin-16161631>*

L'ontologie de Wittgenstein présente à son sommet trois classes principales : SOURCE, PERSON et SUBJECT.

- La classe SOURCE renferme les documents primaires (tels que les ressources disponibles en ligne sur la base Wittgenstein Source) et secondaires relatives aux études wittgensteiniennes (issue par exemple des articles publiés par les Wittgenstein Archives de Bergen (<http://wab.uib.no/agora-alws/>)).
- La classe PERSON contient les instances de personnes (fictives, littéraires ou réelles) repérées dans le corpus des œuvres de Wittgenstein
- La classe SUBJECT rassemble les sujets philosophiques impliqués au fil de l'oeuvre de Wittgenstein. Elle contient entre autre autres les sous-classes ISSUE, POINT, FIELD et

PERSPECTIVE.

- ISSUE renvoie aux thèmes explorés par Wittgenstein (par exemple : « proposition élémentaire », « indépendance logique », « image », « état de choses », « philosophie », « analyse logique », etc.).
- POINT recouvre les points discutés par Wittgenstein à travers ses différentes analyses.
- FIELD correspond à des champs de la philosophie générale.
- PERSPECTIVE donne la possibilité à des instances de ISSUE ou de POINT d'entretenir des relations potentiellement conflictuelles. Cette classe permet ainsi d'accueillir des concepts et interprétations divergentes voire contradictoires sans nuire à la cohérence de l'ontologie. Nous la présentons plus bas.

- **Place**
    - Instances: Skjolden; Cambridge
  - **Date**
    - Instances: 11 May 1936
  - **Issue**
    - Instances: philosophy; logical analysis
  - **Point**
    - Example of instance: Logical analysis is essential to philosophy
  - **Field (a field of philosophical discussion)**
    - Has subclasses:
      - Epistemology
        - Scepticism
          - » Rule-FollowingScepticism
- **Perspective**
  - Has subclasses: APichler\_Course\_TLP; APichler\_Course\_PI
  - Instances: contradiction; state\_of\_affairs ...

Figure 19. Eléments de la branche SUBJECT de la Wittgenstein Ontology.

La classe SUBJECT appelle une attention particulière dans la mesure où sa construction implique des choix de méthode particuliers. Nous en retiendrons deux.

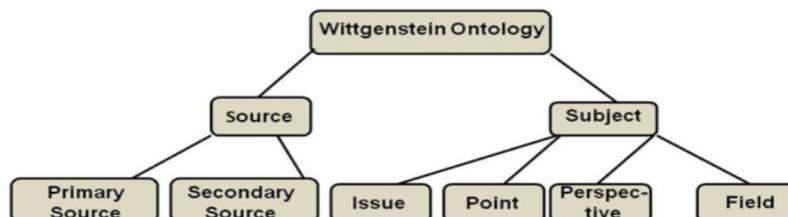


Figure 20. Distribution des sous-classes de SOURCE et SUBJECT de l'Ontologie Wittgenstein.

Source : Zölner-Weber and Pichler, Towards Wittgenstein on the semantic web, *Digital Humanities*, 2012, consultable à l'adresse <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/towards-wittgenstein-on-the-semantic-web.1.html>

Le premier de ces choix de méthode concerne l'établissement de la liste des sujets. Pour opérer une sélection pertinente, l'équipe des Wittgenstein Archives s'est appuyée à la fois sur un corpus de références constitué de dictionnaires et encyclopédies spécialisés<sup>33</sup> (approche « top-down ») et sur une exploration directe des textes du philosophe (approche « bottom-up »). L'équipe a ainsi cherché à refléter aussi fidèlement que possible le paysage de la recherche sur Wittgenstein afin, notamment, de s'assurer l'adhésion de ses futurs utilisateurs.

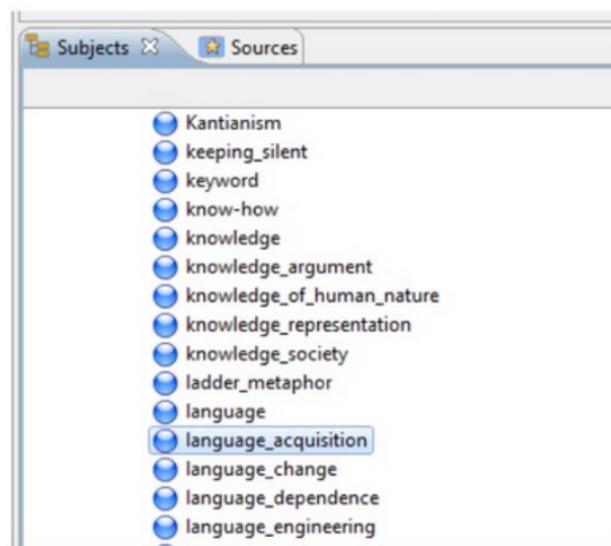


Figure 21. Extrait des sujets de la classe SUBJECT.

Le second choix de méthode effectué pour la classe SUBJECT concerne la difficile mais nécessaire

33 Trois sources : 1) Richter, Duncan. 2014. *Historical Dictionary of Wittgenstein's Philosophy*. Rowman & Littlefield Publishers. 2) Glock, Hans-Johann. 1707. *A Wittgenstein Dictionary 1st edition by Glock, Hans-Johann (1996) Paperback*. Wiley-Blackwell. 3) *Stanford Encyclopedia of Philosophy (SEP)* en ligne.

question de l'ambivalence des concepts dans l'œuvre de Wittgenstein. Celle-ci recèle en effet de nombreux contenus de pensée (concepts, affirmations) dont le sens n'ai pas toujours aisé à fixer, en raison tantôt de l'évolution de la pensée du philosophie, tantôt d'une polysémie intrinsèque de sa terminologie. Cette variété sémantique rend particulièrement délicate la modélisation de la philosophie de Wittgenstein à travers une catégorisation unique et univoque. Cette difficulté majeure, qui constitue le nœud du problème examiné par notre mémoire, a clairement été perçue et assumée par les auteur de la Wittgenstein Ontology : comment concevoir une ontologie accueillant sans ruiner sa propre cohérence, la variété conflictuelle des points de vues qu'elle développe en son sein et suscite chez ses commentateurs ?<sup>34</sup>

Ce point est d'importance : on ne saurait envisager de trancher la polysémie ni même l'équivocité des concepts wittgensteiniens à seule fin de satisfaire l'exigence de la modélisation ontologique. Car cette variété sémantique est précisément cela même qu'il s'agit de modéliser, sans quoi l'ontologie n'offrirait qu'une image, certes univoque, mais tronquée et finalement partielle de la philosophie de Wittgenstein. Il ne peut entrer dans le dessein du concepteur d'ontologie de trancher des controverses disciplinaires, par ailleurs légitimes et nécessaires dans le cas des savoirs humanistes et des sciences humaines. Il lui revient au contraire de trouver une solution permettant de les représenter au sein d'une même ontologie.

Cette position difficile, bien que seule possible, est celle qu'ont tâché de tenir l'équipe des Wittgenstein Archives pour les mêmes raisons épistémologiques de fond relatives au statut des humanités, tout autant que pour des raisons de succès de l'entreprise auprès de chercheurs désireux de nourrir le débat, non de le clore.

La solution avancée par l'équipe de l'Université de Bergen consiste en l'introduction d'une classe supplémentaire baptisée « PERSPECTIVE ». Cette classe, affirme Amélie Zöllner-Weber, permet de tenir ensemble, au sein d'une ontologie qui reste cohérente, des affirmations incompatibles ou contradictoires tirée de la philosophie de Wittgenstein, en les associant à des perspectives d'interprétation différentes.

Comment cette classe fonctionne-t-elle ? Tâchons à présent de le préciser.

---

34 Cette tension de la logique formelle et de la pensée à l'oeuvre dans les sciences humaines, qui constitue pour certains une limite de principe à la construction d'ontologies d'un domaine des humanités (par exemple chez Shirky, Veltman ou encore Zacklad), est soulignée de façon récurrente par Amélie Zöllner-Weber qui s'efforce néanmoins d'envisager une synthèse possible. Voir par exemple son article *Ontologies and Logic Reasoning as Tools in Humanities?*, Digital Humanities Quarterly, 3 (4), 2009.

#### 2.1.4. La classe *PERSPECTIVE*

Tout l'intérêt d'une ontologie consiste à modéliser en logique différents contenus de savoir en vue de les rendre compréhensibles (i-e calculables) par une machine. Celle-ci assurera de la sorte la cohérence du savoir modélisé. En conséquence, deux triplets ne sauraient entrer en contradiction, sans entraîner *ipso facto* la censure du moteur d'inférence, garant de la validité formelle de l'ontologie. Or, s'agissant d'un corpus de données issu des sciences humaines ou des savoirs humanistes (telles, ici, les œuvres de Wittgenstein) on ne peut manquer de rencontrer des contenus impliquant, plus ou moins directement, contradiction. Pour maintenir ces éléments au sein d'une ontologie valide, l'équipe des Wittgenstein Archives ont adopté la position suivante : si l'on ne peut affirmer A et non-A au sein d'une même classe POINT (ou ISSUE), il reste possible de les séparer en rangeant sous deux classes *PERSPECTIVE* distinctes. Le moteur d'inférence, veillant à la cohérence du contenu de chaque classe, ne rencontrera donc jamais, sous le même rapport, les affirmations A et non-A. De cette façon, le triplet « logicalAnalysis – isCentralTo - philosophy » renseigné sous une *PERSPECTIVE*1, n'entrerait pas en conflit avec le triplet « logicalAnalysis – isNotCentralTo - philosophy » renseigné sous la *PERSPECTIVE*2. En ce sens, les des deux thèses se trouvent « confinées » (Zöllner-Weber, *ibid.*), conjurant tout risque de contradiction.

A l'appui de cet élément d'explication, on pourra se référer au schéma que Macha, Falch et Pichler font figurer dans leur article cité en introduction de ce chapitre (en note). Nous le reproduisons ci-dessous.

Dans une explicitation formelle du fonctionnement des classes *PERSPECTIVES* au sein de l'ontologie, les auteurs ont recours à deux définitions. La première définit la classe *PERSPECTIVE* :

« Une *PERSPECTIVE P* désigne une classe contenant plusieurs sous-classes dont chacune représente un groupement spécifique des instances de *ISSUE* (i-e *Problème*) et *POINT* (*Assertion*). Chaque *Perspective* doit être fermée, autrement dit : chaque *POINT* dans *P* (*Perspective*) doit mentionner des *ISSUES* de *P*, et chaque *ISSUE* de *P* doit être mentionné dans quelque *POINT* de *P*. » [nous traduisons].

De façon formelle<sup>35</sup> :

- Pour tout Point  $x$  et pour tout Issue  $i$  : si  $x$  affirme la Perspective  $P$  et  $x$  discute l'Issue  $i$ , alors  $i$  affirme la Perspective  $P$ .
- Pour tout Point  $x$  et toute Issue  $i$  : si  $i$  affirme la Perspective  $P$  et  $x$  discute l'Issue  $i$ , alors  $x$  affirme la Perspective  $P$ .

Quant à la seconde définition, elle précise le type de rapports que peuvent entretenir deux PERSPECTIVES :

« Définition 2. Soit  $P_1$  et  $P_2$  deux Perspectives et  $\alpha$  un mapping de  $P_1$  vers  $P_2$ .  $\alpha$  est un isomorphisme entre elles si et seulement si il mappe tous les Points de  $P_1$  vers les Points de  $P_2$ , toutes les Issues de  $P_1$  vers les Issues de  $P_2$ , tous les Attributs de  $P_1$  vers les Attributs de  $P_2$  et toutes les Relations au sein de  $P_1$  vers les Relations au sein de  $P_2$ , et de même pour les autres composants. Tous les éléments de  $P_2$  ont leurs contreparties dans  $P_1$ , et  $\alpha$  préserve la structure des entités. » [nous traduisons]

Le schéma ci-dessous représente la manière dont la *Wittgenstein Ontology* exprime deux Perspectives conflictuelles, voire contradictoires, sur une même question/problème (*Issue2*), au sein d'une architecture formellement valide et sémantiquement satisfaisante. Pour ce faire, on attache à une classe *Issue2* deux sous-classes, *Issue1* et *Issue3*, renvoyant aux deux affirmations contradictoires ou conflictuelles (*Point1*, *Point2*) par la relation *discussesIssue*.

---

<sup>35</sup> Le passage qui suit est une traduction (libre) des propos de l'article de MACHA, Jakub, FALCH, Rune J., PICHLER, Alois, cité plus haut.

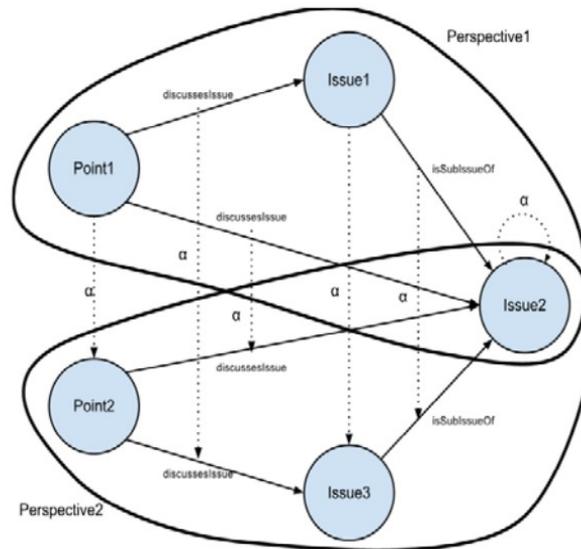


Figure 22 : Exemple de deux Perspectives isomorphiques. Source : MACHA et al.(op.cit)

#### 2.1.4. Bilan

Telle que présentée par ses principaux concepteurs, la Wittgenstein Ontology semble démontrer à la fois la *possibilité* et le *réel intérêt* de la construction d'ontologie pour les humanités numériques. Comme y insistent Pichler et Zöllner-Weber, l'un des principaux défis qu'il importait de relever dans la modélisation du domaine de la philosophie de Wittgenstein, résidait dans les contradictions inévitables qu'elle comporte. La classe PERSPECTIVE y apporte, semble-t-il, une solution opérationnelle, en permettant de représenter des interprétations concurrentes au sein d'une même ontologie. Possible, donc, une ontologie d'un domaine irréductible à première vue à une stricte formalisation logique, présente par ailleurs un réel intérêt. La navigation de l'utilisateur au sein de l'ontologie lui permet d'enrichir sémantiquement sa propre recherche en découvrant des liens qui seraient demeurés inapparents sans cet instrument.

## 2.2. Le choix d'une ontologie de structure : l'enjeu de l'interopérabilité des bases de données en archéologie et patrimoine antique

Un panorama du développement et de l'usage des ontologies pour les humanités numériques ne serait pas complet s'il négligeait de rendre compte d'un certain type d'ontologies sensiblement différent de celui que nous avons jusqu'ici désigné sous le nom d'ontologies de domaines. Tandis que ces dernières ont pour objet de représenter une conceptualisation des connaissances propres d'un domaine du savoir, à l'image de la Wittgenstein Ontology évoquée précédemment, d'autres ontologies s'efforcent, dans une démarche plus générale, de fournir un modèle de données permettant, à qui voudra s'en saisir, de représenter les concepts et informations d'un savoir tombant sous son champ d'application. Parmi ces ontologies de « structure », s'il nous est permis de les nommer ainsi, l'ontologie du CIDOC-CRM s'impose tout particulièrement dans le champ des humanités en raison à la fois de la richesse de sa structure conceptuelle et relationnelle, et de la forte adhésion qu'elle a su recueillir depuis sa création en 1994, auprès d'une communauté active de chercheurs et de professionnels de la documentation. La régularité de ses mises à jour ainsi que les nombreuses extensions dont est fait l'objet depuis lors, témoignent du succès de ce modèle de référence pour la représentation des objets du patrimoine et de musées (« cultural heritage »).

Tirant occasion d'un stage au Laboratoire Archéologie et Territoires de l'Université de Tours/CNRS (UMR 7324 – CITERES, MSH Val-de-Loire) consacré à la mise en correspondance (mapping) des champs de la base de données ArSol (Archives du Sol) vers l'ontologie du CIDOC-CRM et ses extensions CRMarchaeo, CRMsci et CRMba, nous voudrions examiner l'apport et les limites de cette ontologie pour la représentation des concepts d'un domaine des sciences humaines (l'archéologie).

### **2.2.1 Pourquoi une ontologie pour l'archéologie et pour le patrimoine antique ? Interopérabilité des bases de données et granularité de l'information**

Dans le domaine de la recherche archéologique, le recours à une ontologie répond essentiellement à la nécessité de surmonter un obstacle lié d'une part à l'augmentation continue du volume des données documentées, mais encore et surtout à la variété particulièrement disparate des bases de données existantes. Significatif à cet égard est le constat formulé par Xavier Rodier et Olivier Marlet <sup>36</sup> du Laboratoire Archéologie et Territoires :

---

<sup>36</sup> Respectivement : Directeur du Laboratoire Archéologie et Territoires, et Ingénieur d'études dans la même unité de recherche.

« We gave up the fruitless debate about the possibility of adoption by the community of archeologists of a single system that can respond to all situations and all needs a long time ago. But faced with the multiplicity of systems and data structures the question arises, " are most systems interoperable ? " »<sup>37</sup>

« Nous avons abandonnés depuis longtemps le débat infructueux relatif à l'adoption par la communauté des archéologues d'un seul système pouvant répondre à toutes les situations et tous les besoins. Mais face à cette multiplicité de systèmes et de structures de données, la question se pose : " la plupart des systèmes sont-ils interoperables ? " »

Si les négociations entre les différents laboratoires de recherche visant à homogénéiser leurs bases de données se soldèrent par un échec dommageable à la bonne marche des échanges au sein de la communauté des archéologues, on ne saurait se contenter d'accuser la mauvaise volonté de quiconque. Car la particularité de chaque base n'est pas le simple fruit d'une histoire contingente, elle est également le produit d'une nécessité conférant à chacune d'elle une égale légitimité, qui tient à la nature singulière de l'objet et du projet de chaque laboratoire. Une unité de recherche n'a pas affaire ici au même type de vestiges qu'ailleurs, ni ne concevra en conséquence le même programme scientifique<sup>38</sup>. L'hétérogénéité des bases de données ne pouvait dès lors trouver sa solution que dans un système permettant une circulation entre les bases, qui s'affranchisse de leurs structures particulières, sans toutefois les abolir. L'ontologie du CIDOC CRM s'est ainsi imposée comme le pilier principal permettant de construire un tel pont.

Le domaine du patrimoine antique, quant à lui, s'il n'en est pas moins confronté au même enjeu d'interopérabilité, en rencontre un second : la granularité de l'information. Prenant acte des limites de la description des objets du patrimoine antique à l'aide des éléments du schéma de métadonnées *Dublin Core*, Anne-Violaine Szabados, Ingénieure de recherche au CNRS (UMR 7041 – ArScAn, équipe LIMC), démontre la nécessité d'un recours à une ontologie pour affiner la

<sup>37</sup> *Using CIDOC CRM for Dynamically Querying ArSol, a Relational Database, from the Semantic Web*, HAL, 2016

<sup>38</sup> C'est ainsi que le LAT décline un programme scientifique autour de trois axes et trois thématiques transversales intimement liés au type d'objets et de faits archéologiques qu'il lui est donné de recueillir et des connaissances qu'il cherche à établir.

documentation des objets. Dans le texte de son intervention aux *3e Journées d'Informatique et Archéologie de Paris - JIAP 2012*, Jun 2012, Paris, France, p.3, Anne-Violaine Szabados déclare ainsi :

« L'ontologie informatique est une aide précieuse non seulement pour exprimer et analyser des données hétérogènes dans des applications logicielles, mais aussi pour les partager et les documenter, car tout ce qui est « évident » dans la structuration de l'information d'une ressource devient explicite par les définitions mêmes des classes et propriétés utilisées. »

Cette capacité des ontologies à « désambiguïser » l'information permet de la réduire à ses constituants élémentaires, à ses grains les fins, véritables atomes logiques d'un savoir :

« La définition de cette norme ISO [*Definition of the CIDOC Conceptual Reference Model*, v. 5.0.4, 2011, p.i-ii] rappelle [...] notamment que le CRM est destiné "à supporter un niveau de détail et de précision exigés par des professionnels des musées et des chercheurs dans le domaine." »

Dès lors, l'homogénéisation « par le haut » des bases de données disparates et la recherche d'une granularité plus fine de l'information, constituent les deux raisons principales justifiant le recours à l'ontologie du CIDOC CRM pour la documentation des données de l'archéologie et des sciences de l'Antiquité.

### **2.2.2. Comment intégrer une ontologie au domaine de l'archéologie et du patrimoine antique ?**

Comment s'est donc opérée la levée de cet obstacle de la variété que nous évoquions plus haut ? S'agissant du projet porté par le Laboratoire Archéologie et Territoires au sein du Consortium MASA<sup>39</sup>, la construction d'une solution technique au problème de l'hétérogénéité des bases, s'est appuyée sur un certain nombre de positionnements théoriques. Xavier Rodier et Olivier Marlet justifient ainsi l'adoption du modèle de données fourni par l'ontologie du CIDOC-CRM pour le domaine de l'archéologie, en soulignant la grande richesse des entités et propriétés disponibles et

---

39 Mémoire des Archéologues et Sites Archéologiques, fondé en 2012.

en particulier deux d'entre elles, qui permettent d'exprimer une distinction fondamentale de l'archéologie entre ce qu'elle nomme les « faits » et ce qu'elle nomme les « objets ». Pour documenter une sépulture ou une unité stratigraphique, qui ne sont pas à proprement parler des objets, mais des réalités matérialisées « en creux » par des traces, on se servira ainsi de la classe E25\_Man\_Made\_Feature, tandis que pour documenter un élément de mobilier (une parure, etc.), on se servira de la classe E22\_Man\_Made\_Object. Ce processus de mise en correspondance des termes techniques utilisés dans une base de données (« sépulture », « unité stratigraphique », « parure », etc.) dans le langage d'une ontologie (opération nommée « mapping »), débouche sur la construction d'un schéma conceptuel : il ne s'agit pas là d'une correspondance de terme à terme, mais bien de terme à concepts et relations, permettant d'exprimer le sens des termes d'un lexique professionnel. En mappant le terme « sépulture » avec la classe E25\_Man\_Made\_Feature, on lui associe toutes les propriétés sémantiques que l'ontologie prévoit pour cette classe. Ce n'est qu'en passant d'un terme à un concept que l'ontologie peut produire l'interopérabilité escomptée. L'implémentation de ce graphe conceptuel consiste ensuite en un exercice de traduction de chaque entité et relation en un triplet RDF lié à tous ceux qui se trouvent sur la même branche pour constituer un graphe sur lequel on pourra ensuite effectuer des requêtes avec SPARQL (pour le détail de ce processus, nous renvoyons à notre rapport en annexe). En prenant soin d'opérer cette mise en correspondance pour chaque champ de leurs bases de données, les différents laboratoires archéologiques pourront ainsi, sans réformer ni leur base de données, ni leurs pratiques de documentation, offrir un point d'accès à leurs données à tout chercheur étranger à la structure et au vocabulaire de cette base spécifique.

Cette solution, dont nous avons pu tester le caractère opératoire, trace nous semble-t-il une voie intéressante pour une utilisation possible des ontologies pour les humanités : différentes interprétations de la recherche archéologique peuvent se côtoyer et s'interroger mutuellement grâce à une ontologie assurant la correspondance de l'une à l'autre. Si un accord fondamental doit être partagé quant à l'adoption du schéma général de l'ontologie du CIDOC CRM, une certaine variété d'approches scientifiques reste possible.

Or, nous pouvons illustrer de façon tout aussi concrète l'intérêt d'une ontologie pour l'obtention d'une granularité optimale, évoquée par Anne-Violaine Szabados. Dans un exemple particulièrement suggestif, l'auteure de l'article déjà cité, montre combien les ressources de

l'ontologie du CIDOC CRM sont incomparablement plus pertinentes pour représenter le savoir d'un objet du patrimoine antique qu'un schéma de métadonnées tel que le Dublin Core :

« une notice [...] sur une statue antique conservée au Musée du Louvre et représentant à l'origine Minerve puer, au XVIII<sup>e</sup> s. et avant restauration, la personnification de Rome, a été rédigée.

[...]

Comment formuler ceci :

*L'objet a été créé au II<sup>e</sup> s. apr. J.-C.*

*... était en porphyre et représentait Minerve*

*... provient du jardin de San Martino de' Monti à Rome*

*... fut acquise en 1646 par le cardinal Mazarin*

*... transférée à Paris, Collection Mazarin, puis dans la Collection de Louis XIV (en 1665)*

*... placée dans les jardins du Trianon*

*... saisie à la Révolution*

*... complétée par du bronze et rénovée par Francuccio au XVII<sup>e</sup> s.,*

*... fut transformée en Rome au XVIII<sup>e</sup> s., par le changement de sa tête et de son bras gauche*

*... a été étudiée par X qui l'a associée à la gravure réalisée en 1681 par Etienne Baudet,*

*... a été comparée à une autre statue par Y ... »*

La documentation de cette statue au moyen des métadonnées Dublin Core demeure trop grossière et « appauvrissante » (Ibid.) en ne permettant aucun discernement entre « les périodes de fabrication et de modification de la statue », « sa date de création » et « la date de modification de l'enregistrement ». L'élément *dc:date* recouvre en effet une indifférenciation que l'on retrouve aussi bien avec l'élément *dc:creator* qu'avec les éléments *dc:title* et *dc:description* qui ne permettent pas de renseigner les divers états de la statue et ses deux identifications : tantôt Minerve, tantôt Rome.

A l'inverse, le réseau sémantique (et non simplement terminologique) d'une ontologie telle que la CRM rend possible un niveau de granularité de l'information documentée, autrement plus

satisfaisant. En particulier, point central de l'exemple précédent de la statue de Minerve/Rome, il devient possible de restituer l'évolution d'un objet en l'inscrivant dans une chronologie précise. Cette illustration pourrait aisément être étendue à l'évolution d'un site archéologique, ou bien d'un édifice où une description diachronique s'avère nécessaire.

Cet exemple permet d'illustrer l'intérêt bien réel que présente le recours à une ontologie pour représenter des contenus de savoirs dans un but de partage, tout en préservant les diverses facettes du savoir sémantisé.

### **2.3. Bilan**

Il ressort de cette étude de cas que la construction d'ontologies dans le contexte des humanités numériques semble à la hauteur des possibilités offertes par cet outil de représentation sémantique des connaissances : si l'enjeu de l'interopérabilité et du partage des données se retrouve dans tous les exemples que nous avons rencontrés, la dimension d'une représentation cohérente (logique) d'un domaine du savoir n'est pas absente, comme en témoigne l'initiative de la Wittgenstein Ontology, misant sur la capacité d'inférence de l'ontologie pour augmenter les capacités cognitives et interprétatives des chercheurs. L'élément crucial qui semble dès lors justifier la construction d'ontologies pour les humanités numériques réside moins dans le caractère massif des données, prédominant dans les ontologies des sciences appliquées, que dans l'obstacle de la variété des langues naturelles et des points de vue interprétatifs. Pour rendre possible une circulation entre ces éléments disparates, le recours à une ontologie se présente comme une solution avantageuse : loin de normaliser la pensée sous une conceptualisation intangible, une ontologie constitue un socle de convergences sémantiques servant de point de passage d'une interprétation à une autre. Les bases de données archéologiques trouvent en l'ontologie du CIDOC CRM un carrefour commun d'échange, tout en conservant leur intégrité et spécificité propre. De la même manière, l'ontologie Wittgenstein propose une solution technique permettant d'accueillir, à l'intérieur d'un espace logique non contradictoire, des interprétations divergentes réelles et légitimes. Si les ontologies apportent en quelque façon une réponse au problème de la Tour de Babel, pour reprendre une expression de Barry Smith, ce n'est toutefois

pas en substituant aux langues naturelles le langage universelle des relations logiques, mais plutôt en constituant pour un champ disciplinaire donné et en fonction de finalités précises et explicites, une grammaire commune servant d'espace de communication aux langues et discours les plus divers.

Plus qu'une médiation utile aux échanges, la construction d'ontologie, exercice de rigueur analytique, offre ainsi, comme le fait remarquer Anne-Violaine Szabados, l'occasion à une discipline de revenir sur ses propres notions et en réinterroger le sens au contact des vues des autres chercheurs. Si cet effort d' « atomisation » d'un savoir en ses briques élémentaires fait courir le risque d'une perte partielle de la richesse de sens des disciplines humanistes, il offre toutefois un moyen efficace pour construire des systèmes à bases de connaissances sensibles à la sémantique et un puissant moteur pour la réflexion d'une discipline sur elle-même.

## Partie 3. Vers Une interprétation fonctionnaliste des ontologies de domaine ?

### 1. Introduction

Notre investigation sur la possibilité, les apports et les enjeux des ontologies de domaine pour les humanités numériques nous a conduit dans un premier temps à en examiner les aspects théoriques : en nous penchant sur l'origine et la nature des ontologies de domaine (Partie 1), nous avons soulevé la question de leur aptitude à capturer les concepts et raisonnements à l'œuvre dans le domaine des savoirs humanistes. Après avoir fait état d'une objection de principe reposant sur l'idée d'une « coupure sémantique » creusant entre le régime de signification logique et computationnel des ontologies et le régime « rhétorique et herméneutique » du sens dans les humanités un abîme *a priori* infranchissable, nous avons tenté d'analyser (Partie 2) quelques cas concrets d'ontologies de domaine pour les humanités numériques afin de confronter notre questionnement théorique à la réalité des pratiques professionnelles. Nous nous sommes alors aperçus que l'ingénierie ontologique renferme bel et bien des ressources d'usage légitime et profitables pour les humanités numériques. Pourtant, en dépit d'un certain dynamisme des recherches dans ce domaine de l'ingénierie des connaissances, les ontologies restent denrées rares au sein des humanités numériques.

Les compétences informatiques manquent dans le secteur des humanités. Si les humanités numériques ont à cet égard ouvert un point de passage, le web sémantique semble constituer davantage un outil d'interopérabilité permettant de s'affranchir des disparités de l'environnement logiciel, qu'un instrument capable de pénétrer la sémantique des données afin d'en tirer de l'information susceptible, en retour, de féconder la recherche. Nous n'avons rencontré que très peu de projets d'ontologies pour les humanités exploitant spécifiquement les capacités

d'inférences des ontologies, comparativement beaucoup plus sollicitées dans le domaine des sciences et des techniques (en particulier biomédicale où cette dimension joue à plein). L'enjeu explicite des projets d'ontologies dans le champ des lettres et sciences humaines se concentre sur l'interopérabilité et l'échange des données. C'est le cas aussi bien de la Wittgenstein Ontology que de l'ontologie du CIDOC-CRM ou encore du projet porté par le consortium MASA. Les technologies importées du web sémantique dans les humanités numériques sont davantage perçues comme un support d'accès et d'échange pour les savoirs que comme une matrice de nouvelles perspectives de recherches.

A cet état de fait nous avons pu identifier des raisons d'ordre pratique : ainsi que l'a souligné à plusieurs reprises Anne-Violaine Szabados lors de l'entretien qu'elle nous a accordé, les compétences informatiques, documentaires et disciplinaires nécessaires à de tels projets sont rarement réunies au sein d'une même unité de recherche. Si des volontés politiques existent pour favoriser l'essor des humanités numériques en général, le point de blocage essentiel se situe, selon elle, au sein des chercheurs et professionnels qui ne perçoivent pas tous l'intérêt de ce type de technologies, en particulier lorsqu'à l'image du web sémantique ou des ontologies, elles renferment un niveau de technicité qui en rendent l'accès difficile. Notre étude des publications scientifiques sur le sujet nous a permis d'établir au demeurant que la dynamique de recherche et de développement autour des ontologies trouve encore son impulsion principale dans le secteur de l'ingénierie informatique d'où sont issues la plupart des auteurs d'ouvrages spécialisés sur la question.

Il est dès lors légitime de considérer qu'une raison forte du « retard » des disciplines humanistes relève en partie d'un hiatus « culturel » tenant éloignées l'une de l'autre les humanités et les technologies de traitement automatique du sens. Comment les humanités ne pourraient-elles pas percevoir, fut-ce à tort, l'irruption de technologies de représentation logique et computationnelles du sens, sans s'inquiéter de leur propre sort ? Car comment la représentation computationnelle du savoir humaniste (ontologies) n'engendrerait-elle par *ipso facto* celle de son objet, à savoir l'homme lui-même dans sa capacité à produire un sens irréductible au calcul ?

L'ingénierie ontologique étendue au domaine des humanités fait-elle courir un risque de dénaturation de l'objet des humanités, engageant par là même leur raison d'être ?

Notre brève et modeste connaissance de la réalité des ontologies dans le domaine des humanités

numériques, nous fait davantage pencher pour une réponse inverse. Nous voudrions dans une dernière partie montrer que ce type de critiques, pour légitimes qu'elles paraissent, reposent au moins en partie sur un malentendu sur l'objet des ontologies de domaine. C'est à cette clarification que nous consacrons la troisième et dernière partie de ce mémoire.

## 2. Réponse à l'objection de la coupure sémantique

### 2.1. « La signification d'un mot est son *usage* dans le langage »<sup>40</sup>

L'examen des fondements et applications des ontologies de domaine que nous avons tenté de mener dans les précédents chapitres nous a permis d'identifier le postulat élémentaire de ce champ de l'ingénierie des connaissances : la réduction méthodique du sens des énoncés d'un savoir à un ensemble d'atomes logiques (classes et propriétés) au sein d'un réseau ordonné de règles autorisant un traitement computationnel des données (le calcul automatique d'inférences). Si cette réduction du sens paraît d'abord appelée par la nature même de l'ingénierie informatique, d'où sont principalement issues les ontologies, nous avons montré qu'elle s'inscrit plus généralement dans une histoire intellectuelle qui, d'Aristote à Carnap, Quine et le premier Wittgenstein, tendent à réduire la pensée à l'opération logique d'un calcul des énoncés. A bien des égards, l'entreprise des ontologues contemporains réactive l'idéal rationaliste classique d'une langue universelle conjurant la malédiction de Babel par l'édification d'un système transparent et univoque de signes qu'il suffirait de combiner pour produire des significations valides et vraies (le web sémantique, pensé à la façon de Tim Berners-Lee, constituant ainsi une sorte de pendant moderne de l'art combinatoire de Leibniz et de l'atomisme logique du Cercle de Vienne).

Pourtant, à l'hégémonie d'un tel formalisme, nous avons dû opposer une double objection. Factuelle d'abord : comment expliquer la disparité manifeste existant entre le nombre des ontologies de domaine dans les sciences et les techniques, et leur relative rareté dans le domaine des humanités numériques ? Objection de principe ensuite : les savoirs humanistes, littéraires et interprétatifs, sont-ils solubles dans l'élément formel des ontologies, marqué du sceau de l'univocité

---

40 Ludwig Wittgenstein, *Investigations philosophiques*, § 43.

des concepts et du calcul des énoncés ? L'apparente désaffection dans laquelle les humanités numériques semblent tenir les ontologies de domaine n'est-elle pas le signe d'une incompatibilité de fond entre des régimes sémantiques irréductibles l'un à l'autre ? Cette objection majeure, à laquelle Manuel Zacklad ou encore François Rastier s'efforcent de prêter leur voix dans le contexte français, tire argument du caractère essentiellement contextualisé du sens dans le domaine des savoirs interprétatifs. On ne peut espérer dégager du réseau notionnel d'une discipline humaniste un ensemble organisé d'atomes de sens qui, à la manière des perles d'un boulier, autoriserait une manipulation purement calculatoire. Comment définir la notion de « style » ou de « personnage littéraire », sans faire état du fait que la signification de ces termes dépend essentiellement d'un contexte d'énonciation qui varie en fonction des connaissances et des intentions des énonciateurs ? Cet arrière-plan fait peser une indétermination irréductible sur le sens des concepts que dès lors aucune combinatoire ne suffirait à dévoiler. Prenant acte d'une telle coupure sémantique, certains concepteurs d'ontologies proposent alors de modérer l'approche logiciste du sens en tâchant d'en restituer la dimension contextuelle. C'est ainsi qu'en marge d'une ingénierie ontologique inspirée du formalisme de Tom Gruber et Tim Berners-Lee, tend à se développer avec John Sowa aux Etats-Unis ou Manuel Zacklad en France, des méthodes « pragmatistes » de construction d'ontologies qui prennent acte, dans le sillage du « second » Wittgenstein, de l'échec du positivisme logique.

## **2.2. De l'ingénierie *ontologique* à l'ingénierie ontologique**

Il se pourrait pourtant qu'en dépit de sa solidité théorique, à nos yeux impeccable, cette objection manque sa cible. Notre aperçu de la réalité des développements et des pratiques de terrain dans le champ des ontologies de domaine, nous convainc en effet qu'il existe à l'origine de cette objection un malentendu qui en atténue considérablement la portée. Cette méprise nous semble enveloppée dans la notion de « représentation » qui sert, au demeurant à bon droit, à définir les ontologies de domaine et à les rattacher à un cadre disciplinaire déterminé (la *représentation* des connaissances). Qu'attendons-nous exactement d'une ontologie à laquelle est assignée la fonction de « représenter » les connaissances d'un domaine ? Attendons-nous qu'elle nous délivre le *décalque* formel de son objet, sorte de réduplication parfaitement fidèle des concepts et des

raisonnements qu'une communauté d'expert de la discipline modélisée met elle-même en œuvre? Ou bien attendons-nous, plus sobrement, qu'elle nous procure un *modèle* des connaissances qui nous en facilite l'usage? Ces deux attentes reposent sur deux façons fort différentes de comprendre ce que signifie « représenter » des connaissances. La première identifie la représentation à une copie fidèle : la modélisation ontologique doit ressembler à son objet. La seconde, en revanche, conçoit la représentation ontologique comme une image susceptible de rendre service à celui qui s'en sert : une projection des connaissances d'un domaine du savoir sur un plan strictement opératoire. En somme, tandis que l'une ordonne l'ontologie à une exigence de vérité (modélisation fidèle de son référent), l'autre l'ordonne à une exigence d'efficacité.

Nous voudrions ici soutenir que l'interprétation référentielle des ontologies (comme représentations adéquates de leur domaine de référence) est, tout compte fait, largement abusive, de sorte que l'objection consistant à pointer l'inadéquation de leur formalisme logique au domaine informel des humanités, tombe d'elle-même. Une ontologie n'est pas un artefact destiné à capturer dans le filet de la logique les objets et la structure d'un domaine de réalité ou les représentations mentales que s'en fait une communauté d'experts, mais un « simple » instrument pour la recherche de connaissances. Sa fonction n'est pas épistémique (délivrer une connaissance), mais heuristique (guider, accompagner la recherche de connaissances). En d'autres termes, comme l'indique Bruno Bachimont (2004), les ontologies n'ont pas vocation « à répliquer [...] les structures du réel (métaphysique) ou de la pensée (cognitivisme) », mais à fournir « l'accès à un fonds culturel partagé permettant de reprendre à son compte des contenus et de les rendre appropriables par d'autres ». Précision qui suffit selon le même auteur à lever le malentendu : les ontologies « sont donc des instruments intellectuels pour l'organisation de nos connaissances et non des modèles du monde » (p.122). L'ingénierie ontologique est d'abord et en définitive *ingénierie* ontologique : son objet n'est pas de représenter l'être ou les êtres (*ontos*, réalités objectives ou réalités du discours, c'est-à-dire les concepts), mais d'offrir un instrument relevant d'une seule et unique exigence d'efficacité.

L'unique question qu'il s'agirait dès lors de poser porterait sur le point de savoir si les ontologies de domaine peuvent rendre des services à la communauté de recherche en humanités, et non de savoir si ces représentations formelles sont adéquates aux genres de concepts et de raisonnements qu'on rencontre dans ce champ disciplinaire.

Il nous revient toutefois en premier lieu d'établir cette position « instrumentaliste » et d'en préciser la portée exacte, en examinant de plus près l'objection que nous désirons écarter.

### 3. Une interprétation fonctionnaliste des ontologies de domaine

#### 3.1. Une ontologie doit-elle être mimétique ?

Nous tenons pour valide l'argument d'une coupure sémantique interdisant toute réduction d'un savoir humaniste à sa formalisation en logique. Toutefois, nous en contestons la portée en considérant qu'une ontologie n'a pas vocation à se substituer au savoir qu'elle représente et qu'elle ne se propose que de *servir*. Pourtant, n'y aurait-il pas un singulier coup de force à soustraire les ontologies de domaine à leur devoir de fidélité envers le savoir qu'elles modélisent ? Comment se pourrait-il en effet qu'une machine accède à la sémantique des données, si l'ontologie ne reproduisait pas en quelque façon le cheminement de la pensée, n'en répliquait pas fidèlement le schéma d'organisation ? Une ontologie des fruits et légumes ne peut me renvoyer des réponses sémantiquement pertinentes aux requêtes que je lui adresse que parce qu'elle parvient à parcourir les *mêmes* liens notionnels que ceux qui s'établissent en mon esprit. Dans ces conditions, la réduction « instrumentaliste » des ontologies (pures et simples instruments dégagés du devoir de fidélité envers le domaine qu'elles modélisent) ne se rendrait pas seulement coupable d'un contresens sémantique, mais se montreraient incapables d'expliquer le caractère opératoire qu'elles assignent elle-même aux ontologies : une modélisation qui n'est pas un décalque de la pensée dans un domaine, ne peut avoir le moindre intérêt sémantique. Une machine ne donne l'apparence d'un comportement intelligent qu'en tant qu'un utilisateur y reconnaît sa propre pensée.

Dans son article consacré à l'examen des ontologies et du raisonnement logique comme outils pour les humanités numériques (*Ontologies and Logic Reasoning as Tools for Digital Humanities*) Amélie Zöllner-Weber semble elle-même partager ce point de vue. Pour construire une ontologie des personnages littéraires, il convient selon elle de recueillir l'état d'esprit du lecteur, de capturer

ses représentations :

« It is aimed at representing the mental information structure of a reader, which (s)he has in mind when reading a book [...] We regard a character as a complex cognitive entity in the reader's mind. »

« On cherche à représenter la structure d'information mentale d'un lecteur, qu'il a l'esprit lorsqu'il lit un livre [...] On considère un personnage comme une entité cognitive complexe dans l'esprit- du lecteur. » (nous traduisons)

A ses yeux, donc, une ontologie de domaine à pour but de déployer une carte permettant à une machine de parcourir, par le moyen du calcul automatique, le même chemin notionnel que l'individu humain dont l'ontologue s'est inspiré. Il faut donc que l'ontologie ait même forme que la pensée qu'elle modélise : une ontologie du domaine littéraire devra reproduire le raisonnement de l'expert en littérature et donc être fidèle à la façon toute informelle et non calculante propre à ce type de savoir. Dès lors ce n'est pas un hasard si dans le même article, l'auteure s'interroge sur la possibilité d'une modélisation logique des savoirs humanistes.

« More generally, when comparing literature studies and computer systems, one has to take into account that semantics in literature studies often do not match semantics in machine-to-machine learning [...]. Often, technical, e.g. machine-to-machine, semantics are more flat and not too complex, whereas in literature semantics could be highly complex. As an example in the present work, the description of the literary characters is dependent on the reader's interpretation. Therefore, the modelling of the ontology must allow for several interpretations side-by-side in the ontology (via instantiations). This is also covered in [[McCarty 2005](#)], where he points out that proper modelling in humanities is important and that the model has to be kept open or flexible, i.e. to be able to include the thinking itself... »

Plus généralement, lorsque l'on compare les études littéraires et les systèmes informatiques, on doit prendre en compte le fait que la sémantique dans les études littéraires ne correspond souvent pas avec la

sémantique dans [...] Souvent, la sémantique technique, par exemple de machine à machine, est très plate et très peu complexe, tandis que la sémantique littéraire peut être hautement complexe. A titre d'exemple, dans cette étude, la description des personnages littéraires dépend de l'interprétation du lecteur. Ainsi, la modélisation de l'ontologie doit autoriser plusieurs interprétations mises côte-à-côte dans l'ontologie (via les instanciations). C'est un point également soulevé par McCarthy, qui souligne qu'une modélisation appropriée en humanités est importante et que ce modèle doit conserver son ouverture et sa flexibilité, c'est-à-dire être capable d'inclure la pensée elle-même. » (nous traduisons)

Si l'ontologie doit saisir la pensée de l'individu qui lui adresse une requête, la question se pose de savoir si elle peut y parvenir lorsque les requêtes en question engagent des concepts dont le sens dépend d'un contexte d'interprétation précis et éventuellement contradictoire avec d'autres.

A cette objection forte, Gunnar Declerck et Jean Charlet nous permettent cependant d'opposer une réponse qui nous paraît très suggestive pour notre propos<sup>41</sup>.

### 3.2. La carte et le territoire

L'argument proposé par ces deux auteurs s'appuie sur une analogie. Nous reproduisons ici le passage dans son intégralité :

Une bonne illustration est la carte routière : le but premier d'une carte routière n'est pas de représenter de manière fidèle l'espace, mais de permettre à son usager de s'orienter. Bien évidemment, la carte doit être fidèle pour assurer cette fonction. Mais fidèle à quoi ? Quelle réalité la carte représente-t-elle ? Si la carte représente l'espace, c'est en appliquant des principes de sélection extrêmement forts (très peu d'éléments de l'espace géographique sont retenus), et un symbolisme des plus abstraits

---

41 DECLERCK Gunnar, CHARLET, Jean, *Pourquoi notre sémantique naïve n'est pas formalisable et pourquoi c'est (presque) sans conséquence sur l'ingénierie ontologique*. In *Intellectica*, 1014/1, pp. 61 sq.

(un code couleur pour distinguer la taille des routes, des lignes de niveau pour l'altitude, des icônes pour indiquer les lieux remarquables, etc.). Et, point essentielle, les principes d'organisation de l'information utilisés pour la conception de la carte n'ont pas à mimer les principes utilisés par le système d'orientation spatial « naturel » de l'être humain.<sup>42</sup>

La modélisation ontologique n'est pas, ne peut être ni ne doit être le double ou l'image du domaine qu'elle représente : la carte n'est pas le territoire et c'est en cela précisément qu'elle est utile. Pour le constructeur d'ontologie, représenter des connaissances ne signifiera donc pas imiter, mais tenir lieu. Dès lors tout change : une ontologie de l'histoire ou de la philosophie, n'imité pas la démarche de l'historien ou du philosophe, mais tiennent lieu du savoir qui est le leur afin d'en offrir une vision synoptique et aisément manipulable, enjeu crucial à l'heure de la massification des données.

« L'IO [ingénierie ontologique] n'évolue donc pas dans un paradigme de reproduction ou de simulation, mais plutôt un paradigme d'augmentation ou de supplémentation : elle exploite l'intelligence des machines pour augmenter celle de l'homme, pallier ses limites intrinsèques, en premier lieu sa finitude mémorielle et les limites de ses capacité de traitement. » (p.21)

### **3.3. Décrire des connaissances pour prescrire des usages : la portée régulatrice des ontologies**

« les ontologies ne doivent pas être vues comme des représentations/modélisations d'un système conceptuel ou d'un ensemble de catégories déjà en place, mais comme un moyen technologique de promouvoir un système de régulation des pratiques de catégorisation justifié par sa fonction opérationnelle : il s'agit d'aide-mémoire et d'aide-à-catégoriser. » (p.24)

#### **L'ingénierie ontologique doit être comprise dans le cadre d'un projet de maîtrise de la diversité et de la complexité des données**

« leur fonction est de réguler la manière de représenter la sémantique des concepts d'un univers

---

<sup>42</sup> *Ibid.*, pp. 22-23

de discours cible, pour assurer la compatibilité et la consistance de différentes ontologies de domaine. Si elles prétendent également être descriptives (leurs édifices conceptuels modélisent un domaine référent), cette optique est dictée par leur volonté prescriptive. » (28)

L'engagement ontologique doit être mesuré à l'aune de ses effets dans le cadre des pratiques des utilisateurs de l'ontologie. La description d'un champ du savoir n'a de valeur qu'en temps qu'elle permet de réguler des pratiques pour explorer et approfondir ce savoir.

Une ontologie vise à « améliorer les performances d'un système d'information » et augmenter les capacités cognitives des individus.

Reste à déterminer quelle méthode de construction, ou du moins quel modèle général, doit guider l'élaboration d'ontologies de domaine pour les humanités numériques.

## **4. Orientations méthodologiques pour la construction d'ontologies dans le champ des humanités numériques**

### **4.1. Une méthodologie orientée utilisateur**

Si une ontologie doit s'efforcer d'assumer avant tout les besoins sémantiques d'une communauté de chercheurs et non la réalité sémantique d'un domaine de savoir, sa construction appelle une approche orientée utilisateur.

Prenant acte de la vocation essentiellement instrumentale de l'ontologie, on s'efforcera de viser l'adéquation de l'ontologie aux usages de ses utilisateurs, plutôt qu'à la nature du contenu du domaine modélisé : s'il s'agit toujours d'épouser d'aussi près que possible la sémantique du domaine qu'on cherche à représenter, les choix devront être guidés en dernière instance par un principe d'efficacité opérationnelle. Au reste, rien ne garantirait qu'une formalisation parfaitement détaillée d'un domaine se traduise par une ontologie réellement maniable et opérationnelle. Quel pourrait être en effet l'intérêt d'une formalisation sensible aux nuances d'une conceptualisation d'un domaine, si celle-ci ne peut pas être implémentée ? Un schéma conceptuel impeccable peut

constituer un réel obstacle à l'implémentation et rendre l'ontologie pratiquement inutilisable. L'option retenue par le Laboratoire Archéologie et Territoires de l'Université de Tours/CNRS s'inscrit dans cette démarche pragmatique : si l'ontologie du CIDOC-CRM et ses extensions permet en principe une formalisation très détaillée d'un domaine de recherche, il peut sembler toutefois judicieux d'observer un principe de simplicité dans l'écriture des mappings, de façon à faciliter la prise en main de l'outil et encourager d'autres laboratoires à rejoindre le mouvement. Pour le reste, ce n'est pas à la nature du savoir archéologique lui-même d'imposer un approfondissement ultérieur de l'ontologie, mais aux besoins qui naissent des usages de la communauté des chercheurs.

Une position équilibrée nous semble exiger de prendre acte des limites du formalisme ontologique, et de concentrer par conséquent l'essentiel de l'effort sur l'efficacité des ontologies (interopérabilité, partage, sensibilité à la sémantique des données). Or, l'Ingénierie ontologique nous est apparue essentiellement soucieuse d'élaborer des formalisations adéquates aux connaissances modélisées, moyennant une complexité qui en rend l'implémentation et la prise en main souvent problématique. Cette recherche d'exactitude dans la représentation des connaissances n'est sans doute pas sans rapport avec la difficile acceptation des ontologies dans le champ des humanités. Une ontologie épousant chaque nuance sémantique d'un savoir humaniste ne représente pas seulement un idéal démesurément grand, mais à supposé qu'il fût réalisé, c'est alors l'appropriation d'une telle ontologie qui ferait problème. Des ontologies de domaines pour les humanités numériques ne sont envisageables à cet égard qu'à la condition qu'elles fassent la démonstration de leur efficacité en nourrissant les échanges d'une communauté de chercheurs et en offrant à celle-ci un outil d'exploitation sémantiquement pertinente au regard de ses besoins propres. L'évaluation d'une ontologie pourrait ici s'appuyer sur le recueil de données statistiques renseignant sur sa fréquentation et son adoption effective dans des démarches de recherches.

## **4.2. Une approche socio-sémantique pour les ontologies de domaine**

Nos développements précédents nous ont permis de poser un premier principe méthodologique général pour la construction d'ontologies de domaine qui renverse notre position de départ :

l'ontologue ne cherchera pas à épouser la réalité sémantique d'un domaine, mais à satisfaire les besoins sémantiques d'une communauté de chercheurs.

Ce principe posé, il resterait à esquisser quelques éléments d'une bonne pratique pour la construction d'ontologie. Nous nous contenterons d'orienter notre lecteur vers deux propositions au reste parfaitement conciliables formulées, la première par Manuel Zacklad<sup>43</sup>, la seconde par Fabien Gandon<sup>44</sup>. Ces deux auteurs nous paraissent en effet occuper une position équilibrée, alliant au souci d'une formalisation adéquate des domaines de connaissances, une visée essentiellement pragmatique. L'un et l'autre plaident pour une approche synthétique qui emprunte à la fois au web sémantique et au web social.

On doit à Manuel Zacklad de nombreux développements visant à promouvoir une synthèse du web sémantique et du web social, motivée par la nécessité, à ses yeux impérieuse, de ne pas cantonner le sens dont l'intelligence et les interactions humaines sont capables à la seule formalité logique. Son principal objet est alors de concevoir des ontologies « sémiotiques », capables en tant que telle, d'intégrer les propriétés sémantiques que les concepts d'un domaine acquièrent en fonction du contexte de leur utilisation. L'article auquel nous renvoyons développe d'intéressantes suggestions sur la façon de construire des « accords définitionnels » sur lesquels édifier les ontologies de domaine. Une semblable prise en compte des contextes de signification (sociaux, linguistiques, etc.) dans la construction d'ontologie nous paraît particulièrement pertinente dans le cadre des disciplines humanistes, dont le caractère spécifique le plus fondamental réside dans la notion d'interprétation. Ce que l'on nomme un peu étroitement sciences humaines renvoie en réalité aux disciplines interprétatives (par opposition aux savoirs explicatifs ou opérationnels dans sciences et des techniques).

#### **4.3. L'apport du *tagging social* pour affiner la sémantique d'une ontologie de domaine**

Fabien Gandon examine quant à lui la possibilité et la fécondité d'un rapprochement, au

---

43 Zacklad Manuel, *Introduction aux ontologies sémiotiques dans le Web Socio Sémantique*. Jun 2005. <sic 00001479>

44 Limpens Freddy, Gandon Fabien, Buffa Michel, « Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art », *19es Journées Francophones d'Ingénierie des Connaissances (IC 2008)*, 2008, Nancy, France. pp.123-134, 2008. Consultable sur : <hal-00416693>

demeurant inattendu, des ontologies et des « folksonomies », en proposant de s'appuyer sur le « tagging social » pour extraire des liens sémantiques susceptibles de nourrir les schémas conceptuels des ontologies de domaine. Ainsi, le formalisme des ontologies se modère en quelque sorte au contact d'une approche sensible à des liens qui ne relèvent pas d'abord de la logique mais de l'usage effectif qu'une communauté déterminée fait des concepts qui lui sont familiers. Cette approche nous paraît particulièrement recommandable aux constructeurs d'ontologies du domaine des humanités, dans la mesure où ces disciplines sont plus fortement marquées que d'autres à la variété des usages terminologiques et conceptuels.

## Conclusion

Nous étions partis dans ce travail de l'idée selon laquelle l'ingénierie ontologique ouvrait un horizon particulièrement prometteur en matière de traitement de l'information. En capturant le réseau des concepts d'un domaine de savoir ou de pratique dans les mailles d'une syntaxe logique et lisible par des machines (RDF), les ontologies constituent en effet des systèmes de traitement des données sensibles à leur sémantique. Ainsi, à la différence des thésaurus et taxonomies, centrées essentiellement sur la dimension lexicale et non conceptuelle d'un domaine, les ontologies permettent de franchir un pas supplémentaire vers l'optimisation de l'accès aux données et de l'échange d'informations pour diverses communautés d'utilisateurs.

A cette promesse technique et documentaire, nous avons cependant du opposer la réalité d'une situation plus contrastée. Tandis que les ontologies de domaine ont très tôt et assez massivement investi le champ des sciences et des techniques, elles ont manifestement tardé à s'étendre aux savoirs humanistes où l'on ne compte qu'un nombre limité d'initiatives. Ce « retard » assez nettement accusé par les ontologies de domaine en humanités peut tenir en partie à la distance séparant historiquement les savoirs humanistes et les technologies de l'informatique de traitement automatisé de l'information. De ce point de vue, le mouvement récent des humanités numériques pourrait alors représenter une nouvelle donne pour les ontologies. L'entrée des sciences humaines dans l'âge des technologies numériques de traitement, d'accès et de partage de l'information ne s'impose-t-elle pas comme une nécessité à l'heure où les données se massifient et où dès lors la question d'un accès sémantiquement pertinent aux données des savoirs s'avère cruciale ? Pourtant, cet argument d'ordre technique semble se heurter à une réalité d'ordre épistémologique : n'existe-t-il pas une incompatibilité de nature entre l'élément purement formel de la syntaxe logique utilisée par les ontologies et l'élément largement informel de la rationalité déployée dans le champ des lettres et sciences humaines ? A une sémantique formelle, procédant à l'analyse logique des énoncés, jusqu'à atomisation complète d'un savoir en concepts et relations élémentaires, Les disciplines humanistes (interprétatives) opposent une sémantique engageant des concepts dont l'essentiel de la signification tient au contexte souvent polémique de leur usage. Un projet d'ontologie de domaine pour les humanités numériques, justifié par ses avantages

techniques et documentaires, ne commettrait-il pas un contresens sur la nature des connaissances à modéliser ? Si une « coupure sémantique » sépare l'ingénierie ontologique et les humanités numériques, comment des ontologies pour ce domaine seraient-elles possible ?

Pour répondre à cette question, nous avons d'abord cherché à déterminer ce que sont les ontologies en tentant d'en retracer la genèse. Pour savoir ce qu'elles sont il importe de savoir d'où elle viennent. Ce point établi, il convenait d'identifier ensuite où vont exactement les ontologies, en précisant par une étude factuelle quels domaines elles investissent prioritairement. Une analyse de l'état des publications scientifiques sur les ontologies de domaine, doublée d'une étude de cas d'ontologies pour les humanités numériques nous a permis à cet égard de mesurer les disparités existantes et d'identifier des solutions techniques convaincantes au problème d'une formalisation des savoirs humanistes. Nous avons dès lors acquis la conviction que les ontologies de domaine pour les humanités numériques sont à la fois possibles, mais encore pleinement légitimes et souhaitables. C'est là ce que nous nous sommes efforcés de montrer dans un dernier temps en renversant notre opinion de départ sur la nature et les possibilités des ontologies de domaines. Loin de s'ordonner à une stricte exigence d'adéquation aux savoirs qu'elle représentent, les ontologies désignent bien davantage des artefacts dont la valeur est d'abord instrumentale. Le constructeur d'ontologie vise en priorité non pas à produire le doublon formel d'un savoir qui en épouse toutes les finesses sémantiques, mais à satisfaire des besoins sémantiques d'une communautés d'utilisateurs déterminée. Dès lors, et en raison même de la place qu'occupent les controverses interprétatives au sein des disciplines humanistes, il nous semble à présent que les ontologies de domaine trouvent à bon droit dans les humanités numériques un terrain d'élection.

## Bibliographie des ouvrages et documents cités et consultés

### Humanités numériques

DACOS Marin, MOUNIER Pierre, *Humanités numériques. Etat des lieux et positionnement de la recherche française dans le contexte international*. Rapport commandé par l'Institut français, opérateur du ministère des Affaires étrangères pour l'action culturelle extérieure de la France, Institut français, Licence Creative Commons, Isbn : 9782354761080 (epub) / 9782354761097 (pdf).

ZIGHED, Djamel Abdelkader, *Les Humanités Numériques en Sciences Humaines et Sociales*, Institut des Sciences de l'Homme, CNRS éditions.

### Web Sémantique

BERNES-LEE, Tim *et al.* « The Semantic Web », *Scientific American*, 2001. Consultable sur : <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

GANDON Fabien, FARON-ZUCKER Christine, CORBY Olivier, *Le Web sémantique. Comment lier les données et les schémas sur le web ?*, Dunod, Paris, 2012. 1 vol. (XIII-206 p.) p. (InfoPro). ISBN : 978-2-10-057294-6.

Cet ouvrage a été à de très nombreuses reprises consulté, tant pour comprendre l'histoire et les enjeux du Web sémantique et de son corollaire le Web de données, que pour consulter des définitions. Les trois auteurs, tous les trois informaticiens, s'adressent aussi aux développeurs web et illustrent leurs propos avec beaucoup d'exemples commentés de code informatique.

### Ontologies info-documentaires

BACHIMONT Bruno, « Qu'est-ce qu'une ontologie ? » [En ligne]. 2004. Consultable sur : [http://www.technolangue.net/imprimer.php3?id\\_article=280](http://www.technolangue.net/imprimer.php3?id_article=280) >

BACHIMONT Bruno, « Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances ». In : *Ingénierie des connaissances: évolutions récentes et nouveaux défis*. Paris : Eyrolles, 2000. p. 305–323.

BENEL Aurélien, « Ontologies du web : histoire refoulée et perspectives paradoxales ». In : *Intellectica*, 2014/1, 61, 123-141.

CHARLET Jean, BACHIMONT Bachimont, TRONCY Raphaël, « Ontologies pour le Web sémantique ». *Revue Information, Interaction, Intelligence I3* [En ligne]. 2004.  
Consultable sur : < [http://www.eurecom.fr/~troncy/Publications/Troncy-revue\\_i304.pdf](http://www.eurecom.fr/~troncy/Publications/Troncy-revue_i304.pdf) >

DECLERCK Gunnar, CHARLET, Jean, « Pourquoi notre sémantique naïve n'est pas formalisable et pourquoi c'est (presque) sans conséquence sur l'ingénierie ontologique », *Intellectica*, 1014/1, pp. 61 sq.

LIMPENS Freddy, GANDON Fabien, BUFFA Michel, « Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art », *19es Journées Francophones d'Ingénierie des Connaissances (IC 2008)*, 2008, Nancy, France. pp.123-134, 2008. Consultable sur : <hal-00416693>

GRENON, Pierre, SMITH, Barry, « Foundation of an ontology of philosophy », *Synthèse*, 2011, 182 (2), pp. 185- 204.  
Consultable sur : [http://www.ontology.buffalo.edu/smith/articles/Ontology\\_of\\_Philosophy.pdf](http://www.ontology.buffalo.edu/smith/articles/Ontology_of_Philosophy.pdf)

GRUBER Tom, « A translation approach to portable ontology specifications », *Knowledge acquisition*. 1993. Vol. 5, n°2, p. 199–220.

GUARINO Nicolas, « Formal ontology in information systems : proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy ». In : *Proceedings of FOIS'98*, Trento, Italy, 6-8 June 1998. Amsterdam : IOS Press, 1998. p. 3-15

MENON Bruno, « Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan », *Documentaliste-Sciences de l'Information* [En ligne]. 2007. Vol. 44, n°1, p. 18. DOI : 10.3917/docsi.441.0018.  
Consultable sur : < <http://www.cairn.info/revue-documentaliste-sciencesde-l-information-2007-1-page-18.htm> >

PICHLER, Aloïs, ZÖLLNER-WEBER, Amélie, « Towards Wittgenstein on the Semantic Web », *Digital Humanities, Conference Abstracts*, Hamburg University Press, 2012, pp.318-321.

PICHLER, Aloïs, *Ludwig Wittgenstein Source*, Bergen Text and Facsimile Edition. In : Pichler, A., Krüger, H. W., Lindebjerg, A., Smith, D.C.P, Bruvik, T.M. et Olstad, V., *Wittgenstein Archives at the University of Bergen*, Bergen. Consultable sur : < <http://www.wittgensteinsource.org/> >

RASTIER François, « Ontologies. » *Revue des sciences et technologies de l'information*, 2004, vol. 18, n°1, p. 15-40

SMITH Barry, « The Relevance of Philosophical Ontology to Information and Computer Science », Preprint version of a paper to appear in Ruth Hagengruber and Uwe Riss (eds.), *Philosophy, Computing and Information Science*, London: Pickering and Chatto, 2014, 75-83

STUDER R., BENJAMINS V. R., FENSEL D. « Knowledge engineering: Principles and methods ». *Data & Knowledge Engineering*. 1998. Vol. 25, no. 1-2, p. 161-197.

Il s'agit ici de l'article à l'origine de la fusion de la définition donnée par [, GRÜBER].et celle de [, BORST].

USHOLD Mike et GRUNINGER Michael, « Ontologies : principes, Methods and Applications », preprint to appear in *Knowledge Engineering Review*, Vol. 11, Num. 2, Juin 1996

ZACKLAD Manuel, *Introduction aux ontologies semiotiques dans le Web Socio Semantique*. Jun 2005. <sic 00001479>

ZACKLAD Manuel, « Quelle formalisation pour les contenus culturels ? », *Documentaliste-Sciences de l'Information*. Vol. 48, n°4, p. 40-41

ZÖLLNER-WEBER, Amélie, PICHLER, Aloïs, « Utilizing OWL for Wittgenstein's Tractatus », *Papers of the 30<sup>th</sup> International Ludwig Wittgenstein Symposium* (Kirchberg am Wechsel, Autriche, 5 août 2011), ALWS, Kirchberg am Wechsel, pp. 248-250.

ZÖLLNER-WEBER, Amélie, « Ontologies and Logic Reasoning as Tools in Humanities ? », *Digital Humanities Quarterly*, 3 (4), 2009.

## Philosophie

ARISTOTE, *Métaphysique*, trad. Tricot, Vrin, tome 1, Paris, 2000.

ARISTOTE, *Catégories. De L'Interprétation*, trad. Tricot, Vrin, Paris, 2000.

DIOGENE LAËRCE, *Vies et doctrines des philosophes illustres*, trad. Marie-Odile Goulet-Cazé, Le Livre de poche. La Pochothèque, 1999, 1398 p.

MONNIN Alexandre, FELIX Edith, « Essai de comparaison des ontologies informatiques et philosophiques : entre êtres et artefacts ». Denis Phan. *Rochebrune'09 : Ontologie et dynamique des systèmes complexes, perspectives interdisciplinaires*, Jan 2009, Megève, France. Chemin de Traverse, pp.14, 2011. Consultable sur < hal-00636120 >

NEF Frédéric, « L'Ontologie au miroir de la terminologie », *Actes de la quatrième conférence TOTh*, Annecy, 3 & 4 juin 2010, Editeur : Institut Porphyre, Savoir et Connaissance. ISBN 978-2-9536168-1-1.

Consultable sur : < [http://www.porphyre.org/toth/files/actes/TOTh-2010\\_01\\_Nef.pdf](http://www.porphyre.org/toth/files/actes/TOTh-2010_01_Nef.pdf) >

PORPHYRE, *Isagogè*, trad. Tricot, Vrin, 1984

QUINE Willard Van Orman, *Le mot et la chose* (1960), Paris, Flammarion, 1999.

WITTGENSTEIN, Ludwig, *Tractatus logico-philosophicus*, trd. Granger, Gallimard, coll. Tel, Paris, 2001.

WITTGENSTEIN, Ludwig, Wittgenstein's Nachlass : The Bergen Electronic Edition. In : Huitfeldt, C. (ed.), Wittgenstein Archives at the University of Bergen under the direction of C. Huitfeldt. Oxford University Press, 2000.

The Wittgenstein Ontology, fichier OWL téléchargeable sur :

<[http://wab.uib.no/wab\\_philospace.page/wittgenstein.owl](http://wab.uib.no/wab_philospace.page/wittgenstein.owl)>

## Table des matières

Partie 1. Les ontologies info-documentaires, architectures du savoir.....	7
1. Ontologie, ontologies.....	8
1.1. Qu'appelle-t-on ontologie ? un mot de philosophie.....	8
1.1.1. Un sens classificatoire : les êtres.....	8
1.1.2. Un sens métaphysique : l'Être.....	8
1.1.3. L'ontologie informatique et documentaire : un emprunt partiel à la philosophie.....	9
1.2. Panorama historique : de l'ontologie philosophique aux ontologies info-documentaires.....	10
1.2.1. L'ontologie avant l' « ontologie » .....	10
1.2.2. Le moment lexical de l'ontologie.....	13
1.2.3. Le moment logiciste de l'ontologie, « clarification logique de la pensée ».....	14
1.3. Les ontologies de domaine en ingénierie des connaissances.....	16
1.3.1. Emergence.....	16
1.3.2. Définitions.....	20
1.3.3. Typologie.....	21
1.3.4. Les constituants des ontologies.....	25
1.3.5. Implémentation informatique .....	27
1.4. La construction d'une ontologie : méthodes et étapes.....	29
2. Les humanités numériques : une nouvelle donne pour les ontologies ? .....	31
2.1. Des ontologies de quel domaine ? .....	31
2.1.1. Les sciences de la nature à l'honneur. Des outils de modélisations des connaissances essentiellement tournés vers les sciences de la nature et les techniques.....	31
2.1.2. Les humanités et sciences humaines en reste.....	32
2.1.3. Les humanités numériques : nouvelle donne pour les ontologies ?.....	33
2.2. L'émergence des Humanités numériques.....	33
2.2.1. Literacy and Linguistic Computing.....	33
2.2.2. Humanities Computing.....	34
2.2.3. Digital Humanities.....	35
2.3. Approche d'une définition.....	36
3. Des ontologies pour les humanités : un contresens ? .....	39
3.1. L'argument de la coupure sémantique.....	39
3.2. Qu'est-ce que la musique bretonne ? Un exemple de cacophonie des concepts. ....	41
3.3. La coupure doit-elle être tranchée ? Deux versions de l'argument.....	42
3.4. A tower of Babel problem (Barry Smith).....	43
Partie 2. Des ontologies pour les humanités numériques.....	44
1. Emergence du thème des ontologies en humanités numériques dans la littérature spécialisée en ligne : données statistiques et analyse .....	45
1.1. Introduction : objet de l'étude et questionnement.....	45
1.2. Etat des lieux de la publication des contenus scientifiques traitant des ontologies info-documentaires dans Google Scholar.....	46
1.2.1. Source de données.....	46
1.2.2. Méthode : Choix des mots-clés et critères de recherche.....	46
1.2.3. Résultats : données quantitatives et commentaires.....	47
1.3. Etat des lieux de la publication des contenus scientifiques traitant des ontologies info-documentaires dans les HAL et HAL-SHS. ....	51
1.3.1. Source de données.....	51
1.3.2. Méthode : critères de recherches et mots-clés retenus.....	51
1.3.3. Résultats : données quantitatives et commentaires.....	52
1.3.5. Focus sur les résultats classés SHS dans notre recherche sur les HAL.....	55

1.3.6. Bilan.....	56
1.4. Etat des lieux de la publication de contenus traitant des ontologies info-documentaires dans le secteur des bibliothèques et de la documentation.....	57
1.4.1 Le Bulletin des Bibliothécaires de France (BBF).....	57
1.4.2 Documentaliste - Sciences de l'information.....	57
1.4.3. Archimag.....	58
1.4.4. Bilan général des trois études .....	58
2. Quelques ontologies du domaine des humanités : examen et appréciation.....	59
2.1. L'Ontologie Wittgenstein de l'Université de Bergen.....	59
2.1.1. Contexte du projet : pourquoi une ontologie ? .....	60
2.1.2. Pourquoi une nouvelle ontologie ? .....	61
2.1.3. Conception de l'ontologie.....	61
Contraintes et périmètre.....	61
Structure.....	62
2.1.4. La classe PERSPECTIVE .....	66
2.1.4. Bilan.....	68
2.2. Le choix d'une ontologie de structure : l'enjeu de l'interopérabilité des bases de données en archéologie et patrimoine antique.....	68
2.2.1 Pourquoi une ontologie pour l'archéologie et pour le patrimoine antique ? Interopérabilité des bases de données et granularité de l'information.....	69
2.2.2. Comment intégrer une ontologie au domaine de l'archéologie et du patrimoine antique ? .....	71
2.3. Bilan.....	74
Partie 3. Vers Une interprétation fonctionnaliste des ontologies de domaine ?.....	76
1. Introduction .....	76
2. Réponse à l'objection de la coupure sémantique.....	78
2.1. « La signification d'un mot est son usage dans le langage » .....	78
2.2. De l'ingénierie ontologique à l'ingénierie ontologique .....	79
3. Une interprétation fonctionnaliste des ontologies de domaine.....	81
3.1. Une ontologie doit-elle être mimétique ?.....	81
3.2. La carte et le territoire.....	83
3.3. Décrire des connaissances pour prescrire des usages : la portée régulatrice des ontologies. .	84
4. Orientations méthodologiques pour la construction d'ontologies dans le champ des humanités numériques.....	85
4.1. Une méthodologie orientée utilisateur.....	85
4.2. Une approche socio-sémantique pour les ontologies de domaine.....	86
4.3. L'apport du tagging social pour affiner la sémantique d'une ontologie de domaine.....	87
Conclusion.....	88
Bibliographie des ouvrages et documents cités et consultés.....	90