



Conception d'une ontologie à partir d'un thésaurus spécialisé dans le domaine de l'archéologie et des sciences de l'Antiquité

Loraine Marcheix

► To cite this version:

Loraine Marcheix. Conception d'une ontologie à partir d'un thésaurus spécialisé dans le domaine de l'archéologie et des sciences de l'Antiquité. domain_shs.info.docu. 2008. mem_00355920

HAL Id: mem_00355920

https://memsic.ccsd.cnrs.fr/mem_00355920

Submitted on 26 Jan 2009

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

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

Université Vincennes – Saint-Denis, Paris 8
Culture et Communication
Département Documentation
2 rue de la Liberté
93526 SAINT-DENIS Cedex

Maison René-Ginouvès,
archéologie et ethnologie
UMS 844 du CNRS
Documentation et FRANTIQ
21 allée de l'Université
92023 NANTERRE Cedex

CONCEPTION D'UNE ONTOLOGIE A PARTIR D'UN THESAURUS SPECIALISE DANS LE DOMAINE DE L'ARCHEOLOGIE ET DES SCIENCES DE L'ANTIQUITE

Mémoire présenté pour le Master II professionnel de Gestion de l'Information et du Document
Spécialité Gestion des Connaissances

Maître de stage, UMS 844 du CNRS
Brigitte Lequeux, Ingénieur de recherche - CNRS

Directeur de mémoire, Paris 8
Bruno Menon, maître de conférence

Loraine MARCHEIX
N° étudiant : 229308

Année 2007-2008

Je tiens à remercier ici l'équipe au sein de laquelle j'ai effectué mon stage et les enseignants du département Documentation de Paris 8 qui ont permis à ce travail de voir le jour.

RESUME

Conception d'une ontologie à partir d'un thésaurus spécialisé dans le domaine de l'archéologie et des sciences de l'Antiquité

Les ontologies informatiques, phénomène relativement récent, peuvent être considérées comme les successeurs logiques des thésaurus, comme l'étape supérieure tant d'un point de vue linguistique qu'informatique et les possibilités que paraît ouvrir la création d'une ontologie sont nombreuses et tentantes : l'amélioration de la recherche documentaire, l'automatisation de l'indexation ou encore l'accessibilité et la visibilité du produit sur le Web.

D'après la mission qui m'a été confiée durant mon stage de fin de master II professionnel GID, au sein de l'UMS 844 Documentation et FRANTIQ du CNRS, il s'agira dans ce mémoire de suivre pas à pas la conception d'une ontologie à partir d'un thésaurus spécialisé dans les sciences de l'Antiquité et l'archéologie. Pour ce faire, le choix a été fait de procéder en deux temps, le premier théorique où nous découvrirons ce qu'est une ontologie et comment elle se construit ; et le second pratique dans lequel nous mettrons la théorie en application à travers la réalisation de la mission : la transformation du thésaurus en ontologie.

Mots-clés : thésaurus, ontologie, ingénierie ontologique, modélisation des connaissances

SUMMARY

Conception of an ontology from a specialized thesaurus in archaeology and antiquity science

Computer ontologies are quite a recent phenomenon which can be considered as logical successors of thesaurus, as a further step from a linguistic point of view as well as a computer one. Moreover, ontology building seems to offer a lot of possibilities such as the improvement of information retrieval, automatic indexing or accessibility and visibility on the Web.

According to the mission I was given for my higher education training course which took place in UMS 844 Documentation and FRANTIQ of CNRS, in this master's paper, we will deal with the building of an ontology from a specialized thesaurus in archaeology and antiquity science and study it step by step. So as to achieve this aim, we will focus on two different aspects: one the one hand the theoretical one and on the other hand the practical one.

Therefore, we will start with discovering what an ontology is and how it is built, and then, we will put the theoretical knowledge into practice in carrying out the mission, that is to say: turning a thesaurus into an ontology.

Key-words: thesaurus, ontology, ontological engineering, knowledge modeling

SOMMAIRE

RESUME	3
SUMMARY	3
SOMMAIRE	5
INTRODUCTION	7
CHAPITRE I : QU'EST-CE QU'UNE ONTOLOGIE ?	17
I. Un peu d'histoire	17
1) <i>Du principe philosophique : la science de l'être en tant qu'être, à la création du terme d'Ontologie</i>	18
2) <i>Évolutions à travers le temps</i>	19
II. L'ontologie en ingénierie des connaissances	21
1) <i>Définition</i>	21
2) <i>Les ontologies</i>	25
III. Ontologie et langage documentaire	28
1) <i>Les langages documentaires</i>	28
2) <i>Les ontologies et les langages documentaires</i>	32
CHAPITRE II : COMMENT CONSTRUIRE UNE ONTOLOGIE ?	37
I. Conception	38
1) <i>Constituants d'une ontologie</i>	39
2) <i>Engagement sémantique et engagement ontologique pour la conception et réalisation d'ontologies</i>	42
II. Langages de représentation des connaissances	45
1) <i>OWL</i>	47
2) <i>SKOS</i>	51
III. Éditeurs d'ontologies	53
1) <i>Protégé</i>	54
2) <i>OilEd</i>	55
3) <i>OntoEdit</i>	56
4) <i>WebODE</i>	57
5) <i>DOE</i>	57
CHAPITRE III : VERS LA TRANSFORMATION DU MICRO-THESAURUS PACTOLS-SUJETS EN ONTOLOGIE	60
I. Description du micro-thésaurus Sujets des PACTOLS	60

1) <i>Structure et thématique de PACTOLS-Sujets</i>	61
2) <i>Les descripteurs dans PACTOLS-Sujets et la gestion des termes</i>	63
3) <i>Représentation du thésaurus PACTOLS et son utilité</i>	69
II. Choix méthodologiques	72
1) <i>Critères de sélection</i>	73
2) <i>Solutions écartées</i>	75
3) <i>Solution retenue : utilisation du logiciel DOE</i>	80
III. Réalisations et perspectives.....	88
1) <i>Réalisations</i>	89
2) <i>Perspectives</i>	112
CONCLUSION	120
Table des matières	122
Table des illustrations	125
Table des acronymes	127
BIBLIOGRAPHIES	130
<i>Travaux cités</i>	131
<i>Indications bibliographiques</i>	134
ANNEXES	139
Annexe I : <i>Organigramme de l'UMS 844 à la MAE</i>	140
Annexe II : <i>Carte des producteurs participants au réseau FRANTIQ</i>	141
Annexe III : <i>Liste des ouvrages et sites Internet de référence utilisés pour la création des PACTOLS et les définitions des Notes d'Application, élaborée par Madame Lequeux</i>	142
Annexe IV : <i>Liste complémentaire des ouvrages et sites Internet de référence utilisés pour compléter les Notes d'Application et les définitions encyclopédiques de l'ontologie</i>	144
Annexe V : <i>Définitions différentielles relatives à l'ontologie de haut niveau de utilisée pour l'ontologie créée à partir du micro-thésaurus PACTOLS-Sujets</i>	145
Annexe VI : <i>Arbre des concepts de l'ontologie différentielle sur l'archéologie et les sciences de l'Antiquité</i>	149
Annexe VII : <i>Diapositives de la présentation des projets FRANTIQ pour l'Université d'été du TGE ADONIS</i>	154

Conception d'une ontologie à partir d'un thésaurus dans le domaine de l'archéologie et des sciences de l'antiquité

INTRODUCTION

Les volumes d'information disponibles ne cessent de croître, les supports d'inscription ne se résument plus à la simple feuille de papier ; les moyens de communiquer ces informations ont considérablement évolué et se sont diversifiés. Face à cette réalité, il est donc normal que les métiers, et par là-même les pratiques, des professionnels de l'information changent et que ces derniers tentent de nouvelles expériences : la création d'ontologies en est une démonstration. De prime abord, la frontière entre **thésaurus** – *ensemble de termes normalisés fondé sur une structuration hiérarchisée et où les termes sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques* – et **ontologie** – *conceptualisation organisée en vue d'aboutir à un objet formel, où les relations entre les concepts sont sémantiques mais formalisées* – semble très mince. En effet, sans entrer dans les détails nous pouvons dire dans un premier temps que tous les deux sont des ressources terminologiques construites dans différents domaines de la connaissance, basées sur des relations à forte dominance hiérarchique et conçues comme un réseau sémantique. Cependant, il existe des spécificités propres à chacun qui nous amène à penser qu'un thésaurus peut servir de base à une ontologie. Les ontologies informatiques, phénomène relativement récent, sont considérées comme les successeurs logiques des thésaurus, comme l'étape supérieure tant d'un point de vue linguistique qu'informatique et les possibilités que paraît ouvrir la création d'une ontologie sont nombreuses et tentantes : l'amélioration de la recherche documentaire, l'automatisation de l'indexation ou encore l'accessibilité et la visibilité du produit sur le Web.

D'ailleurs, le stage que j'ai effectué à la Maison René-Ginouvès, Archéologie et Ethnologie m'a permis d'approfondir particulièrement les questions qui se posent quant à la transformation d'un thésaurus en ontologie, objet de ce mémoire et objectif de la mission qui m'a été confiée par Madame Brigitte Lequeux. Néanmoins, avant d'aborder ce sujet de manière plus précise, il convient de préciser dans quel cadre et dans quel but ce mémoire a été réalisé.

Il s'agit d'un exercice de fin d'études. En effet, ce n'est en aucun cas un document qui a valeur universelle. C'est le résultat d'une part, de deux ans de formation à l'université de Paris 8 au sein du master Gestion de l'Information et du Document (GID), d'échanges avec les professeurs, de lectures personnelles, d'expériences professionnelles dans le monde de la documentation et de discussions avec des professionnels de la documentation. D'autre part, et plus particulièrement, ce travail est le fruit d'une présence active au sein du service Documentation/FRANTIQ de l'UMS 844 à la Maison René-Ginouvès, Archéologie et Ethnologie de Nanterre de février à juillet 2008 pendant une durée totale de trois mois. Au sein de ce service, j'ai été soutenue, conseillée, suivie par Madame Brigitte Lequeux et j'ai également bénéficié des conseils et du soutien de Madame Nathalie Le Tellier-Becquart, membre également du service Documentation/FRANTIQ. Ce mémoire se trouve donc être la synthèse de deux ans de formation universitaire et d'une mission venant clôturer cette même formation. Il s'agit en quelque sorte de confronter la théorie universitaire accumulée pendant deux ans d'apprentissage à la pratique. C'est un exercice de nature académique mêlant d'une part connaissances théoriques sur la documentation en général, et sur un sujet plus précis qui se situe dans le domaine de l'ingénierie des connaissances, à savoir les ontologies, et d'autre part réalisations pratiques, réelles en fonction d'une tâche précise à accomplir dans un contexte professionnel bien spécifique. Enfin, il est possible de considérer ce document à la fois comme venant clôturer deux ans d'études et un stage de trois mois, et comme annonçant un départ dans la vie active dans l'univers de l'information et du document. Il signe une fin universitaire, et annonce un début dans la vie professionnelle.

Maintenant, il reste à présenter à la fois le contexte dans lequel s'est déroulé le stage, ainsi que l'objet de l'étude, à savoir le thésaurus PACTOLS.

Contexte de la mission

Mon stage qui se trouve être le fondement de ce mémoire professionnel de Gestion de l'Information et du Document a eu lieu du mois de février au mois de juillet 2008 pour une durée totale de trois mois sous la direction de Madame Lequeux, ingénieur de recherche, responsable du service documentation et de FRANTIQ (Fédération de Ressources sur l'Antiquité) au sein de l'Unité Mixte de Service 844 (UMS 844) à la Maison René-Ginouvès, Archéologie et Ethnologie¹.

¹ Annexe I *Organigramme de la MAE* pour plus de précision sur l'organisation de la Maison

(MAE). Cette Maison relève actuellement de trois tutelles : le CNRS et les Universités de Paris X et Paris 1. Le ministère de la Culture est en outre le quatrième partenaire de plusieurs unités. La MAE est membre du Réseau national des Maisons des Sciences de l'Homme (Groupement d'intérêt scientifique) comprenant 22 maisons sous la tutelle du ministère de la Recherche et de l'Enseignement supérieur.

La MAE réunit plusieurs disciplines des Sciences humaines et sociales – Archéologie, Préhistoire et Protohistoire, Histoire ancienne et médiévale, Ethnologie et sociologie comparative – au sein de quatre Unités Mixtes de Recherche et de l'UMS 844 qui gère la Maison et propose des services mutualisés pour les équipes UMR. La Maison compte ainsi plus de 1000 personnes dont 300 chercheurs et enseignants-chercheurs, 80 ingénieurs et techniciens et plus de 500 doctorants. L'UMS 844 se divise en trois pôles réunis sous la direction de Monsieur Pierre Rouillard. Le pôle éditorial assure la publication de revues scientifiques de renom international, propres au CNRS ou hébergées : *Gallia*, *Gallia Préhistoire*, *Paléorient*, *Journal de la Société des Américanistes*, *Archives européennes de sociologie*, *Ethnologie française* et trois autres revues vont rejoindre prochainement ce pôle qui assure aussi la publication du colloque annuel de la Maison René-Ginouvès. Les recherches s'appuient sur des équipements et des outils communs issus notamment du pôle documentaire qui associe deux bibliothèques spécialisées en archéologie et en ethnologie, les archives scientifiques, les services de la cartographie et de photographie, le service Documentation/FRANTIQ regroupant le service des bases de données documentaires qui lui est associé. Enfin, le pôle administratif et technique assure le fonctionnement général de la Maison. Le service documentation/FRANTIQ est celui dans lequel j'ai effectué mon stage. Il est dirigé par Madame Lequeux assistée de Madame Le Tellier-Becquart pour ce qui concerne les bases de données documentaires.

FRANTIQ est un réseau de bibliothèques spécialisées en archéologie et en sciences de l'Antiquité. À l'origine, il traitait uniquement de l'Antiquité classique sous tous ses aspects tels que la société, la linguistique, la philosophie, le droit, l'économie ou encore l'épigraphie. Cette thématique répondait à un besoin des chercheurs car le domaine de l'Antiquité était absent aussi bien des *Bulletins signalétiques* du CDSH que des bases de l'INIST du CNRS. Depuis quelques années, en raison de la politique de regroupement des équipes au CNRS et de la participation des revues *Gallia* au réseau FRANTIQ, ce dernier a étendu son champ disciplinaire à la Préhistoire d'un côté, et à l'archéologie du Moyen-Age et des Temps Modernes et industriels de l'autre.

En ce qui concerne l'étendue spatiale, le réseau n'a pas de limite géographique, mais il couvre en priorité l'Europe et le Bassin méditerranéen. Les limites correspondent aux fonds des bibliothèques et aux centres d'intérêts des chercheurs.

Les objectifs principaux de FRANTIQ lors de sa création en 1984 par quatre bibliothèques qui ont réuni leurs catalogues informatisés en instaurant un minimum de champs communs interrogeables, étaient, et restent encore d'actualité aujourd'hui : faciliter la recherche, répartir les tâches documentaires entre les différents centres participant au réseau afin d'économiser du temps et donc du personnel, et répartir l'achat des documents. Pour réaliser ses objectifs, le réseau a été financé jusqu'en 2005 par le CNRS dans le cadre des réseaux documentaires puis des plates-formes technologiques. Depuis 2006, les réseaux doivent exister sur les fonds propres des laboratoires qui les hébergent. FRANTIQ est une coopérative d'équipes d'horizons divers : CNRS, enseignement supérieur (universités), musées, ministère de la Culture, collectivités territoriales et institutions étrangères. Aujourd'hui, 25 producteurs, c'est-à-dire bibliothèques communiquant leurs données bibliographiques, sont membres du réseau FRANTIQ. Ils sont répartis dans toute la France comme le montre la carte visible en annexe² et deux producteurs se situent à l'étranger, l'un en Italie à Bordighera et l'autre en Espagne à Gérone.

Le personnel participant au fonctionnement du réseau, outre les bibliothécaires et documentalistes membres, est détaché de laboratoire pour des plein-temps ou des temps très partiels et la composition de la cellule de pilotage évolue en fonction des programmes. Pierre Rouillard est l'administrateur général assisté de deux administrateurs (le directeur de la Maison de l'Orient Méditerranéen – MOM – à Lyon (actuellement Rémy Boucharlat) et le directeur du CEPAM (Didier Binder), le comité de pilotage est composé de Madame Lequeux, qui est également la documentaliste responsable du réseau, aidée d'un poste à mi-temps de documentaliste (Madame Le Tellier-Becquart), et d'un informaticien Miled Rousset, dédié au réseau, détaché de la MOM à Lyon, où se trouve notamment le serveur.

L'adhésion à la coopérative se fait par simple lettre d'adhésion. Les producteurs pouvaient à l'origine participer à deux types de structures : le Catalogue Collectif Indexé (CCI) ou les bases thématiques : Ensembles Thématiques Signés (ETS). Ces deux structures ont été réunies en 2006 pour ne faire qu'une base de données, le CCI. En échange de cette adhésion et de la participation au réseau, les producteurs bénéficient des évolutions du produit (passage en texte intégral,

² Voir l'annexe II *Carte des producteurs participant au réseau FRANTIQ*. Carte à laquelle il faut ajouter l'UMR6566 : Centre de Recherche en Archéologie, Archéosciences, Histoire (CREAAH) de Rennes.

participation à de nouveaux projets, à de nouveaux réseaux³...) et ont accès à tous les outils proposés par FRANTIQ tels que la maintenance informatique, le serveur, la liste de diffusion, et les thésaurus PACTOLS qui sont l'objet de ma mission. FRANTIQ est à la fois un réseau, une base de données, et des outils ou services. Concrètement et pour résumer, les activités de FRANTIQ se traduisent par :

- Le développement d'un catalogue collectif indexé qui est un catalogue bibliographique répondant aux normes et standards internationaux en documentation et bibliothéconomie. Il intègre les catalogues des différents producteurs tout en assurant l'autonomie de ceux-ci et la visibilité respective de chaque base
- Le développement d'outils documentaires (application et paramétrage du SIGB Koha, PACTOLS et formations) et informatiques communs
- La diffusion de ressources documentaires (tables et bibliographies spécialisées, Missions Archéologiques Françaises à l'Étranger (MAFE), banques d'images)
- L'informatisation et/ou la migration de catalogues de bibliothèques
- La gestion et la maintenance des serveurs informatiques dédiés à ces fonctions dans le cadre des plateformes technologiques du département des Sciences de l'Homme et de la Société (SHS)
- L'élaboration d'un thésaurus spécialisé (PACTOLS) multilingue

Enfin, les activités, outils et autres informations concernant FRANTIQ sont accessibles via un site web : <http://frantiq.mom.fr> . Ce site fait actuellement l'objet d'une refonte totale visant à le rendre dynamique et collaboratif.

Objet d'étude

Le PACTOLS est un thésaurus, c'est-à-dire un système de gestion de termes contrôlés et organisés méthodiquement, géré par le programme FRANTIQ. Il se compose de six micro-thésaurus et d'une liste. Les termes ont été sélectionnés et organisés selon des relations sémantiques : synonymie, antonymie, adjacence ; et/ou hiérarchiques : termes génériques et

³ Le réseau FRANTIQ participe actuellement au projet DAPHNE : **D**onnées en **A**rchéologie, **P**réhistoire et **H**istoire sur le **N**Et. C'est un guichet unique d'accès gratuit à des bases de données bibliographiques thématiques. Daphné est une mutualisation de données, c'est-à-dire une convergence de trois bases des données BAHF (Bulletin Analytique d'Histoire Romaine), FRANCIS (domaines Pré-et protohistoire, art et archéologie et partiellement histoire et sciences des religions), FRANTIQ-CCI. <http://www.daphne.cnrs.fr>

termes spécifiques. Ils servent à interroger plusieurs bases de données administrées par FRANTIQ, mais aussi des bases extérieures, telles que les archives de la Maison René-Ginouvès, les images et les textes du projet « Archéologie de la France-informations » (AdlFI). Les PACTOLS ont aussi été diffusés au Musée Saint-Raymond à Toulouse et dans plusieurs services régionaux ou départements d'archéologie.

PACTOLS est l'acronyme de « Peuples et cultures, Anthroponymes, Chronologie relative, Toponymes, Œuvres, Lieux et Sujets ».

- Les **Toponymes** constituent une liste ouverte de lieu-dit. Il s'agit d'une liste alphabétique avec des règles d'écriture.
- Les **Peuples** regroupent tous les noms des entités culturelles reconnues en Préhistoire et Protohistoire, les peuples antiques et actuels et aussi les noms d'habitants d'une cité mentionnés dans un document historique (537 descripteurs dont 98 synonymes⁴)
- Les **Anthroponymes** regroupent tous les noms de personnes (dieux, héros, écrivains, gouvernants...) dont on parle dans un document. Les chiffres romains ont été transformés en chiffres arabes pour les dynasties et les gouvernants. Ce micro-thésaurus comprend de nombreux synonymes en raison des transcriptions et des graphies diverses des noms propres et des chiffres (3159 descripteurs dont 1077 synonymes)
- La **chronologie** regroupe tous les termes de chronologie relative et les ères géologiques. La chronologie absolue a été laissée de côté car elle aurait nécessité trop de niveaux hiérarchiques pour englober une date dans une décade, un demi-siècle, un siècle et un millénaire ; le tout avec des dates avant et après J.-C. (267 descripteurs et 204 synonymes)
- Le micro-thésaurus des **Œuvres** regroupe des œuvres artistiques et littéraires qu'elles soient religieuses, juridiques ou poétiques. Les œuvres artistiques sont celles dont les écrits parlent le plus : Ara Pacis, Colonne de Trajan, ...etc. (187 descripteurs et 42 synonymes)
- Les **Lieux** regroupent tous les noms de lieux hiérarchisés à partir du continent jusqu'à la commune, la géographie physique et hydrographique des continents ainsi que des rubriques sur les mondes antique et médiéval. L'option pour indiquer les lieux (du continent à la commune) correspond à la géographie administrative actuelle. Elle oblige à suivre l'actualité géopolitique (Allemagne, URSS, Yougoslavie). Les îles sont regroupées sous la rubrique Océans. Ce micro-thésaurus s'est appuyé sur le thésaurus de la base INTERGO du CNRS, en y ajoutant quelques rubriques : les « Lieux imaginaires » :

⁴ Les chiffres mentionnés sont donnés à titre indicatif, les PACTOLS étant évolutifs, les chiffres ne sont pas fixes.

Enfers, Atlantide, Tartessos, et les grands sites (qui font l'objet d'une littérature abondante) : Lascaux, Pincevent, Éphèse, ...etc. Le micro-thésaurus Lieux est en forte et permanente augmentation : il dépend en effet de l'actualité archéologique (13402 descripteurs dont 1937 synonymes).

- Les **Sujets**, vingt-six chapitres, regroupent les domaines d'études sur l'Antiquité (pour les archéologues les parties droit, paléographie, philosophie et même politique sont d'un intérêt moindre). La rubrique « céramologie » qui devrait être une partie du chapitre « art » a été remontée au premier niveau car trop de références traitent de ce type de vestige archéologique ; un chapitre « mort » est aussi présent au premier niveau, car les fouilles de sépulture sont une grande activité des archéologues. Le vocabulaire des Sujets est relativement stable depuis l'intégration de termes spécifiques pour la Préhistoire. La participation au projet DAPHNE, évoqué précédemment⁵, oblige néanmoins à une harmonisation des termes et à des ajouts. (4456 descripteurs dont 888 synonymes)

On remarque qu'en ce qui concerne les thèmes des thésaurus, ceux-ci reprennent ceux du réseau FRANTIQ. Tout comme pour FRANTIQ, le domaine s'est étendu de l'Antiquité classique à la Préhistoire d'une part, et à l'archéologie du Moyen Age et des Temps Modernes d'autre part. Les PACTOLS sont le fruit du travail de Madame Lequeux. Leur création a débuté en 1987. Ils sont nés d'un besoin des chercheurs. Dès 1987, la base du CCI était interrogeable au minitel (3615). Les progrès de l'informatique, ou plutôt le « tout informatique » ont fait oublier un temps (1995-1999) l'utilité des thésaurus au profit des interrogations en « texte intégral ». Très vite les tests n'ont pas convenu aux chercheurs. Par exemple, il fallait interroger en plusieurs langues car il n'y avait pas de traduction automatique des concepts. En outre, il y avait un besoin d'avoir un thésaurus qui couvre le domaine des sciences de l'Antiquité car aucun outil adapté n'existait. Les historiens de l'Antiquité et les archéologues, comme tous les littéraires, ont appris à écrire en français, avec l'obligation d'éviter les répétitions de mots. La richesse de notre langue leur donne cette possibilité ; ainsi, il n'est pas possible de réaliser une indexation à partir de texte comme cela a déjà été fait. On trouve avec la même signification : céramique, vase et poterie. L'indexation avec un vocabulaire dit libre n'est pas non plus envisageable car, par exemple, les littéraires emploient « théâtre » pour l'ensemble des œuvres dramatiques alors que les archéologues l'utilisent pour le bâtiment. De plus, de nombreuses homonymies procuraient beaucoup de bruit à l'interrogation : Alexandre le Grand a créé quinze Alexandrie sur son

⁵ Voir à la page 11 de l'Introduction.

itinéraire d'expansion ; Thésée est à la fois un héros et une commune de France, de nombreuses cultures préhistoriques ont des sites éponymes. Il fallait donc distinguer tous ces termes. En 1987, la base de données bibliographiques CCI comprenait 30000 références bibliographiques et 46000 mots-clés. Il a donc été nécessaire de rationaliser l'existant. Cette rationalisation s'est traduite par la normalisation du catalogage, et par la création du thésaurus PACTOLS avec comme corollaire la formation des producteurs aux nouveaux outils. Ce thésaurus ne devait pas être dédié à un seul type de données, mais être multi-objets : texte, image, bibliographie, cartographie, et même artefacts.

Ce thésaurus est poly-hiérarchique, c'est-à-dire qu'un terme spécifique peut être relié à plusieurs termes génériques, et multilingue. La version actuelle comporte d'une part la traduction quasi-intégrale des Sujets et de la Chronologie en italien et en anglais. Par ailleurs, des termes isolés sont traduits en allemand et en espagnol. Enfin, les concepts sont reliés entre eux par des relations essentielles, définies dans les normes internationales ISO 2788 pour les thésaurus monolingues, et ISO 5964 pour les thésaurus multilingues.

C'est donc ce thésaurus, important, imposant, inédit et à la grande réputation au sein du réseau FRANTIQ, et ailleurs, qui a été l'objet de ma mission intitulée *étude de la transformation du thésaurus PACTOLS en ontologie*.

Cette mission s'inscrit dans le cadre de l'appel à projets lancé par le TGE ADONIS⁶ en août 2007 auprès des laboratoires et centres de recherche pour soutenir et promouvoir les outils et systèmes innovants de valorisation et de diffusion des données de la recherche et permettre leur mutualisation et leur mise à disposition pour l'ensemble de la communauté des SHS, appel à projets auquel a répondu Madame Lequeux. Le souhait d'ADONIS est que l'accent soit mis sur les approches collaboratives et tout particulièrement celles qui se développent dans les démarches « Web 2.0 » et en général sur les outils dont l'adaptation dans le secteur des SHS permet des usages innovants et un progrès dans la circulation, la collaboration, la visibilité et la réutilisation des données. C'est pourquoi, la transformation du thésaurus PACTOLS en ontologie fait partie de ce programme ainsi que la création par l'informaticien du réseau FRANTIQ, Miled Rousset, d'un logiciel de gestion pour le thésaurus : OpenTheso et OpenThesoWeb pour l'Opac utilisateur.

De prime abord, nous pouvons déjà annoncer que le projet de transformer le thésaurus PACTOLS en ontologie s'inscrit totalement dans cet appel à projet. En effet, sans trop déflorer

⁶ TGE signifie Très Grand Équipement (du Département des Sciences de l'Homme et de la Société du CNRS)
ADONIS signifie Accès unifié aux données et documents numériques des sciences humaines et sociales

le sujet, il est possible d'affirmer que d'une part la création d'une ontologie suppose l'utilisation d'un langage recommandé par le W3C⁷, l'autorité visant à régulariser les pratiques sur le Web, OWL⁸, et cela permet donc des progrès dans la circulation, la collaboration, la visibilité et la réutilisation des données sur le Web. Et, d'autre part, beaucoup d'espoirs sont placés dans l'utilisation des ontologies pour réaliser des indexations automatiques, mais plus probablement et plus raisonnablement semble-t-il, dans un avenir proche, pour améliorer la recherche documentaire. De cette manière, cela répondra à la question des usages innovants souhaités par le TGE ADONIS. Néanmoins, étant donné deux contraintes majeures : le temps qui m'a été imparti pour défricher le sujet, m'approprier le thésaurus, mener mes réflexions et réaliser des tests – trois mois – et l'absence d'appui spécialisé au sein du service pour me guider ou du moins me conseiller dans mes choix en matière d'ingénierie des connaissances, l'accent a été mis prioritairement sur la création de l'ontologie à partir du thésaurus et non sur les possibilités d'utilisation qu'elle pouvait offrir.

Ainsi, le mémoire s'interrogera plus particulièrement sur la première phase de création d'une ontologie et non sur ses applications. Il s'agira de prendre connaissance de la ou des méthodologie(s) possible(s) pour réaliser une ontologie et d'étudier quels moyens informatiques sont nécessaires à cette création une fois la méthodologie définie. En outre, il faut garder à l'esprit que toutes ces réflexions seront menées dans l'idée de transformer un thésaurus en ontologie. En d'autres termes cela signifie que, contrairement aux cas les plus courants, nous possédons déjà le contenu de l'ontologie. Notre objectif principal ne sera donc pas de savoir comment définir un domaine de connaissances, mais de savoir comment faire entrer un domaine de connaissances préexistant dans un modèle. En résumé, le mémoire tâchera de répondre à la question suivante : comment construire une ontologie à partir d'un thésaurus ? Pour ce faire, la démarche consistera dans un premier temps à dire ce qu'est une ontologie, à se demander tout d'abord ce que doit être le résultat avant de se demander comment le faire, et à savoir ce qui le différencie d'un thésaurus puisqu'ils semblent si proches. Ensuite, nous pourrons nous interroger sur les manières de créer les ontologies et sur ce que cela implique, théoriquement d'une part, et matériellement, informatiquement même, d'autre part. Enfin, forts de ce savoir, nous exposerons et analyserons les réalisations effectuées pendant le stage ainsi que ce qu'il reste à faire pour achever le projet. Cela reviendra donc à poser les bases du sujet – bases qui ne sont nullement exhaustives puisqu'elles ne sont que le fruit des recherches que j'ai effectuées dans le temps qui m'était

⁷ *World Wide Web Consortium*

⁸ *Web Ontology Language*

imparti pour réaliser ma mission et qui m'ont été nécessaires pour le réaliser – puis peu à peu à dérouler la problématique pour finalement analyser, au regard de la théorie acquise dans les deux premières parties, la réalisation de la mission.

CHAPITRE I : QU'EST-CE QU'UNE ONTOLOGIE⁹ ?

L'ontologie est une notion difficile, complexe, à définir. Si l'on regarde dans le dictionnaire *Le nouveau Petit Robert de la langue française 2007* [LE ROBERT 2006], on trouve la définition suivante : « PHILOS. Partie de la métaphysique qui s'applique à l'être en tant qu'être, indépendamment de ses déterminations particulières. » Il n'y a rien – sinon le nom – de prime abord, qui ne fasse référence à l'ontologie qui nous concerne et qui trouve sa place dans le domaine de l'ingénierie des connaissances. Comment cette notion a-t-elle glissé du domaine de la philosophie à celui de l'ingénierie des connaissances, pourquoi cette homonymie ? Nous tenterons dans un premier temps, de comprendre les racines de cette désignation avant de s'intéresser, dans un second temps, à ce qu'est l'ontologie en ingénierie des connaissances. Enfin, après ces propos, une question ne manquera pas de se poser : quelle est la place de l'ontologie au sein des langages documentaires ? Nous y apporterons donc, en dernier lieu, des éléments de réponses.

I. Un peu d'histoire

Le mot ontologie est construit à partir des racines grecques : *ontos* qui veut dire « ce qui existe », « l'existant », et *logos* pour « le discours », « l'étude ». En d'autres termes, Ontologie signifie l'étude de ce qui existe, la science de l'être.

Pour tenter de définir cette notion d'ontologie, il n'est pas inutile de revenir brièvement sur les origines purement philosophiques de la notion et sur son histoire afin de comprendre pourquoi l'ingénierie des connaissances a choisi de réutiliser un mot préexistant pour désigner l'un de ses objets.

⁹ Par commodité et ce afin de les distinguer, l'Ontologie en tant que science et partie de la philosophie, prendra une majuscule. L'ontologie en ingénierie des connaissances commencera par une minuscule et pourra être au pluriel.

1) Du principe philosophique : la science de l'être en tant qu'être, à la création du terme d'Ontologie

De manière à se constituer des fondements solides sur la notion d'ontologie, dans cette partie, nous nous intéresserons à son sens philosophique, le plus étroit et le plus théorique, où l'Ontologie est définie comme *la théorie de l'être en tant qu'être*.

Bien que le mot ontologie ne date que du XVIIe siècle, le principe peut en être trouvé chez Aristote dès le IIIe siècle avant-notre-ère. En effet, dans la philosophie classique, l'Ontologie correspond en partie à ce qu'Aristote appelait la Philosophie Première ou encore ce qui a été appelé la métaphysique, c'est-à-dire la science de l'être en tant qu'être, par opposition aux philosophies secondes – la physique notamment – qui, elles, s'intéressaient à l'étude des manifestations de l'être (les étants). La métaphysique peut se définir de la manière suivante : premièrement c'est une science qui étudie les premiers principes et les premières causes ; deuxièmement c'est une science qui étudie « l'être en tant qu'être » et non ce qui fait qu'un être est ceci ou cela mais ce qui fait qu'il est un être.

D'après d'Aristote, l'étant se dit de multiples façons. Cela signifie que l'être peut être divisé en catégories. Elles constituent les différentes descriptions associées aux manifestations de l'être dans le monde traduites par des propositions. Ainsi, on peut qualifier l'être selon dix catégories :

- Substance
- Quantité
- Qualité (quels attributs)
- Relation (plus ...que, *etc.*)
- Lieu (où)
- Temps (quand)
- Posture (positionné comment)
- Possession (avec quoi)
- Action (en faisant quoi)
- Souffrance (subissant/étant affecté par)

Parmi ces dix catégories, la Substance a une importance prépondérante car elle constitue l'essence sans quoi une entité ne peut subsister, et qui par le fait même, individualise et différencie une entité par rapport à toutes les autres et elle assure une structure qui reste stable à travers les

changements continuels du monde. Ainsi, il est possible de reconnaître un être (un certain individu par exemple), comme étant en essence le même, en dépit des changements qu'il subit de toutes les autres propriétés (les autres catégories dites accidentelles ou transitoires) au fil du temps, modifiant par exemple son apparence/qualité (nourrisson, enfant, adulte, vieillard), sa posture, ses actions, ses possessions, etc.

En outre, nous l'avons mentionné, le terme d'Ontologie ne date pas d'Aristote ou d'aucun autre de ses contemporains d'ailleurs, seul le principe est existant, et ce jusqu'au XVII^e siècle. En effet, le terme Ontologie naît en 1613 lorsqu'il apparaît, en grec, chez Goclenius (professeur de logique) et Lohrardus (recteur du Gymnasium de Saint-Gall) avant de se trouver sous sa forme latine dans les écrits de Clauberg (théologien et philosophe) en 1647 [CLAUBERG 1647].

Dans *Lexicon philosophicum* de Goclenius [GOCCLENIUS 1613], le terme désigne la science de l'être en général. Il correspond par conséquent à la recherche en tant qu'être d'Aristote, parmi d'autres objets, propres à la philosophie première ou encore à la métaphysique.

Plus tard, Clauberg attribue la même signification au terme Ontologie dans *Metaphysica* et *Ontosophie sive ontologia*, où il l'emploie pour faire référence à une sorte de métaphysique générale qui aurait pour objet les caractéristiques essentielles communes à tous les êtres, à savoir : substance, existence, essence, etc.

Enfin, la diffusion du terme est due à l'*Ontologia* de Christian Wolff [WOLFF 1730], qui dans le concept scolaire de métaphysique, rangeait l'Ontologie en tant que métaphysique générale, puisqu'elle traitait de l'être en général, et la distinguait des trois sciences métaphysiques spéciales que sont la psychologie rationnelle (l'être de *l'âme intellectuelle*), la cosmologie rationnelle (l'être du *monde*) et la théologie rationnelle (l'être de *Dieu*), chacune traitant d'une région déterminée de l'être.

2) Évolutions à travers le temps ¹⁰ [PSYCHE et al. 2003]

À partir de cette première diffusion du terme, on peut dire que la notion d'Ontologie va faire l'objet de divergences, d'accords, d'enrichissements, en un mot évoluer. On a d'ailleurs pu noter qu'il y avait une différence entre la conception wolffienne de l'être et la conception classique. Cette différence dépendrait de ses prémisses leibniziennes qui veulent que le possible

¹⁰ Cette partie est écrite à partir de l'article de PSYCHE, Valéry et al. Apport de l'ingénierie ontologique aux environnements de formation à distance paru en 2003 dans la revue *sticf.org*.

précède le réel, si bien que l'être est défini comme ce qui veut exister, soit qu'il existe effectivement, soit qu'il n'existe pas, l'existence apparaissant comme le complément de la possibilité.

Les principes suprêmes de l'Ontologie sont le principe de non-contradiction et le principe leibnizien de raison suffisante. Les déterminations internes de l'être sont ses attributs essentiels. Pour le reste, l'Ontologie étudie une série de couples conceptuels, comme quantité et qualité, nécessité et contingence, simplicité et composition, finitude et infinitude, identité et diversité, cause et effet, etc.

Peu à peu, la notion d'Ontologie évolue. Kant conçoit son *analytique transcendantale* – première partie de la logique transcendantale, dans la *Critique de la raison pure* [KANT 2001] – de telle manière qu'elle peut prendre la place de la « vieille » Ontologie. Hegel agit de la même façon avec la logique qu'il identifie à la métaphysique. Il écrit dans l'un des textes introductifs à la *Science de la logique* : « la logique objective prend [...] la place de la métaphysique d'autrefois » [HEGEL 1994].

Le terme d'Ontologie réinvestit le discours philosophique dans le cadre du développement de la phénoménologie. Il s'agit de la science des phénomènes, de la science de ce qui existe, de ce qui se manifeste à la conscience, soit directement, soit par l'intermédiaire des sens. Notamment, le projet de phénoménologie pure d'Husserl le conduit à parler d'ontologie régionale, ou sciences idéales de genres d'être qui empiriquement sont l'objet de plusieurs sciences – l'ontologie régionale de la nature physique par exemple. L'école existentialiste, avec Sartre, développe ensuite sa propre vision de l'Ontologie.

Dans la philosophie analytique, l'Ontologie a été étroitement liée à la logique et à la philosophie du langage. Selon Quine, les engagements ontologiques du discours ne sont pas tant déterminés par ses assertions¹¹ d'existence que par le type de variable sur lesquelles le langage admet la quantification : ainsi, une position nominaliste – pour qui il n'existe que des individus – admettra seulement la quantification sur des variables individuelles.

L'Ontologie est donc déterminée par la sémantique de son langage, et coïncide de fait avec les aspects généraux de cette sémantique. Un courant significatif de la philosophie analytique poursuit la construction d'une ontologie formelle, c'est-à-dire d'une théorie formelle des modes d'être. La construction d'une telle théorie coïncide avec la définition d'une sémantique pour un langage logique, dans laquelle peuvent trouver place les types d'entités que la théorie admet (par

¹¹ Assertion : proposition, de forme affirmative ou négative, qui énonce un jugement et que l'on soutient comme vraie absolue. (TLFi – Trésors de la langue française informatisé)

exemple, des individus, des classes ou bien des propriétés, etc.), et où sont définies les relations entre les différents types d'entités. Une telle ontologie formelle implique de soumettre à une reformulation dans le langage logique toutes les théories traditionnelles de l'être substantiel (idéalités mathématiques, réalités phénoménales des sciences naturelles, etc.).

À travers ce large panorama historique de l'Ontologie, nous avons pu remarquer le glissement d'une part de l'Ontologie de la théorie de l'être en tant qu'être à la méditation de l'être en général et d'autre part vers une formalisation dans un langage logique des théories traditionnelles de l'être substantiel. L'Ontologie s'interroge sur ce qu'est l'être, sur ce qui existe, sur ce qu'est la réalité, mais une question peut aller de pair avec tout cela : comment fait-on pour connaître, comment dire qu'une chose est ce qu'elle est ? On se demande ce que sont les choses, les étants, et comment faire pour les représenter ?

II. L'ontologie en ingénierie des connaissances

Après un détour sur l'origine philosophique des ontologies, nous entrons maintenant au cœur du sujet puisque nous allons nous intéresser à ce qu'est l'ontologie, et nous pouvons même dire : à ce que sont les ontologies, en ingénierie des connaissances.

Dans un premier temps, il sera question, comme dans la partie précédente, de tenter de définir la notion d'ontologie en tant que nouvelle approche de modélisation des connaissances, puis nous verrons pourquoi il est possible d'utiliser le pluriel lorsque l'on parle d'une notion comme celle-ci.

Enfin, par soucis de clarifier les choses avant de passer à la définition d'une notion relativement complexe, très schématiquement, précisons que l'ingénierie des connaissances évoque les techniques utilisées pour manipuler des connaissances, notamment sur ordinateur. C'est donc un domaine très proche de l'intelligence artificielle et de l'informatique.

1) Définition

La réapparition et la diffusion du terme d'ontologie dans un domaine autre que la philosophie est un phénomène assez récent puisque qu'il date du début des années 1990.

En ingénierie des connaissances, les ontologies sont apparues dans le cadre des démarches d'acquisition des connaissances pour les systèmes à base de connaissances, les SBC. Les SBC ont succédé à ce qu'on appelle les systèmes experts, c'est-à-dire des outils capables de reproduire les mécanismes cognitifs d'un expert dans un domaine particulier. Plus précisément, un système expert est un logiciel capable de répondre à des questions en effectuant un raisonnement à partir de faits et de règles connus. Par rapport aux systèmes experts qui se composent d'une base de faits déclarative, d'une base de règles et d'un moteur d'inférence, les SBC proposaient de spécifier, d'un côté, des connaissances du domaine modélisé et, de l'autre, des connaissances de raisonnement décrivant les règles d'utilisation des connaissances du domaine. L'idée de cette séparation modulaire était de construire mieux et plus rapidement des SBC en réutilisant le plus possible des composants génériques, que ce soit au niveau du raisonnement ou des connaissances du domaine – ces dernières précisant tout ce qui a trait au domaine. Dans ce contexte, les chercheurs ont proposé de fonder ces connaissances sur la spécification d'une *ontologie*.

Par ailleurs, les ontologies s'inscrivent dans la continuité de nombreux travaux sur la représentation des connaissances, réseaux sémantiques, cartes et graphes conceptuels et on peut préciser que leur popularité dans le monde de la gestion de l'information a principalement bénéficié du développement du web sémantique [BERNERS-LEE 2001]. Le web sémantique ou web de données désigne un ensemble de technologies visant à rendre le contenu des ressources du Web accessible et utilisable par des programmes et agents logiciels grâce à un système de métadonnées formelles utilisant notamment les langages développés par le W3C. En d'autres termes, il s'agit d'une extension du Web actuel dans laquelle les informations, auxquelles on donne une signification bien définie, sont reliées entre elles de manière à ce qu'elles soient comprises par les ordinateurs dans le but de transformer la masse des pages Web en un index hiérarchisé et de permettre de trouver rapidement les informations recherchées.

Ainsi, dans un premier temps, on peut en déduire que le terme d'ontologie désigne des artefacts informatiques qui permettent la représentation d'un domaine de connaissance.

Pour aller plus loin, lorsque l'on tente de définir ce qu'est une ontologie, certains noms reviennent inmanquablement comme notamment : Gruber, Sowa, Chandrasekaran, Uschold et Gruninger ou encore Guarino, parce qu'ils sont considérés comme des précurseurs et/ou théoriciens de l'ontologie. Même si actuellement de nouveaux noms font leur apparition dans le domaine, tels que Bruno Bachimont, Raphaël Troncy, Nathalie Aussenac-Gilles, pour entamer

une définition des ontologies, on revient généralement toujours à ces débuts et aux premières personnes qui en ont parlé.

En effet, dans leurs écrits on peut constater que chacune d'elles apportent une pierre à la constitution d'une définition :

Pour Gruber, « une ontologie est une spécification partagée d'une conceptualisation ». [GRUBER 1993a]

Pour Sowa [SOWA 1999], une ontologie est un catalogue des types de choses supposées exister dans un domaine, du point de vue d'une personne utilisant un langage pour parler du domaine.

Pour Chandrasekaran [CHANDRASEKARAN *et al.*1999], une ontologie est une théorie du contenu sur les sortes d'objets, les propriétés de ces objets et leurs relations possibles dans un domaine spécifié de connaissances.

Pour Uschold et Gruninger [USCHOLD et GRUNINGER 1996], « il s'agit du terme utilisé se référant à la compréhension partagée d'un domaine d'intérêt qui peut être utilisé comme cadre unificateur pour résoudre les problèmes de communication entre les gens et d'interopérabilité entre les systèmes. »

Pour Guarino [GUARINO et POLI 1995] [GUARINO 1997a] [GUARINO 1997b], qui synthétise beaucoup d'éléments qui ont été dits sur les ontologies par d'autres personnes, il peut exister plusieurs ontologies concurrentes du même domaine, et la conceptualisation évoquée par Gruber, peut n'être que partielle. Dans une ontologie, il y a une distinction *a priori* entre les entités du monde, et entre les catégories utilisées pour modéliser le monde.

Ainsi, ces différents éléments, nous font comprendre ce que peut être une ontologie : une modélisation des connaissances ; un système, une théorie pour représenter un domaine de connaissances partagé par une communauté, etc. mais ne nous en donnent pas une définition complète et surtout ne laissent qu'entrevoir tout ce que la conception d'une ontologie implique, sans l'explicitier. Or, pour continuer ce mémoire et saisir son contenu, il faut se mettre d'accord sur une définition précise et rigoureuse. Pour ce faire, j'utiliserai les deux définitions complémentaires proposées par Jean Charlet, Bruno Bachimont et Raphaël Troncy dans un article datant de 2004 [CHARLET *et al.* 2004] :

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

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

[...]

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

[...]

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

La seconde définition propose un point de vue complémentaire. Il est cohérent avec la première mais plus précis, en termes de spécification et par rapport à une application informatique. Elle permet de préciser notamment les contraintes qui s'imposent successivement au concepteur d'une ontologie :

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

D'ailleurs, ce dernier point nous permet de faire la transition avec le deuxième aspect évoqué dans l'introduction de cette partie, à savoir l'explication de l'utilisation du pluriel pour designer une ontologie en ingénierie des connaissances.

2) Les ontologies

Ainsi, il est possible et même conseillé d'utiliser le pluriel pour parler de la notion d'ontologie afin de refléter les multiples facettes qu'elle recouvre.

Selon Van Heijst [VAN HEIJST *et al.*], on peut distinguer quatre types d'ontologies :

- **Les ontologies du domaine** : elles sont appelées de la sorte parce qu'elles expriment des conceptualisations spécifiques à un domaine. Elles rendent compte du vocabulaire d'un domaine spécifique au travers de concepts et de relations qui modélisent les principales activités, les théories et les principes de base du domaine en question. Elles sont réutilisables pour plusieurs applications concernant le domaine pour lequel elles ont été créées car elles ont été conçues de façon aussi indépendante que possible du type de manipulations qui vont être opérées sur ces connaissances.
- **Les ontologies applicatives ou ontologies d'application** : ce sont les ontologies les plus spécifiques, elles contiennent les connaissances requises pour une application particulière et ne sont pas réutilisables. Elles peuvent en outre inclure une ontologie de domaine.
- **Les ontologies génériques ou ontologies de haut niveau** : elles expriment des conceptualisations valables dans différents domaines de valeur relativement générale comme les notions d'objets, de propriété, de valeur, d'état, ou encore de temporalité. Théoriquement, ces ontologies doivent pouvoir être reliées au sommet des ontologies de domaines.
- **Les ontologies de représentation** : ce type d'ontologies regroupe les concepts utilisés pour formaliser les connaissances. Parmi les ontologies de représentation, on trouve des ontologies qui vont décrire les notions utilisées dans toutes les ontologies pour spécifier les connaissances, telles que les substances, les concepts, les relations etc. Par exemple, la « Frame-Ontology » est une ontologie de représentation. Elle définit de manière formelle les concepts utilisés principalement dans les langages à base de *frames* : classes, sous-classes, attributs, valeurs, relations et axiomes. Selon Guarino, les ontologies de représentation sont en fait indépendantes des différents domaines de connaissances, puisqu'elles décrivent des primitives cognitives communes aux différents domaines.

Par ailleurs, Uschold et Gruninger [USCHOLD et GRUNINGER 1996] précisent que les ontologies peuvent être de natures variables, suivant le type de langage utilisé et donc allant d'un degré de formalisation zéro à une formalisation totale. Quatre distinctions sont mises au jour :

- **Les ontologies informelles** : elles sont exprimées en langue naturelle. Ainsi, cela peut les rendre plus compréhensibles par l'utilisateur, mais cela rend plus difficile à vérifier l'absence de redondance ou de contradiction dans les ontologies. En d'autres termes, elles sont plus difficiles à valider.
- **Les ontologies semi-informelles** : elles sont exprimées dans une forme de langue naturelle structurée et limitée. Cela permet d'augmenter la clarté de l'ontologie tout en réduisant l'ambiguïté.
- **Les ontologies semi-formelles** : elles sont exprimées dans un langage artificiel et défini formellement.
- **Les ontologies formelles** : elles sont exprimées dans un langage artificiel disposant d'une sémantique formelle, permettant de prouver des propriétés de cette ontologie. L'intérêt de ces ontologies est la possibilité d'effectuer des vérifications sur l'ontologie : complétude, non-redondance, cohérence, etc.

Uschold et Gruninger expliquent également que du degré de formalisation de l'ontologie, dépend le degré d'automatisation dans les diverses tâches impliquant l'ontologie. « 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 » [USCHOLD et GRUNINGER 1996].

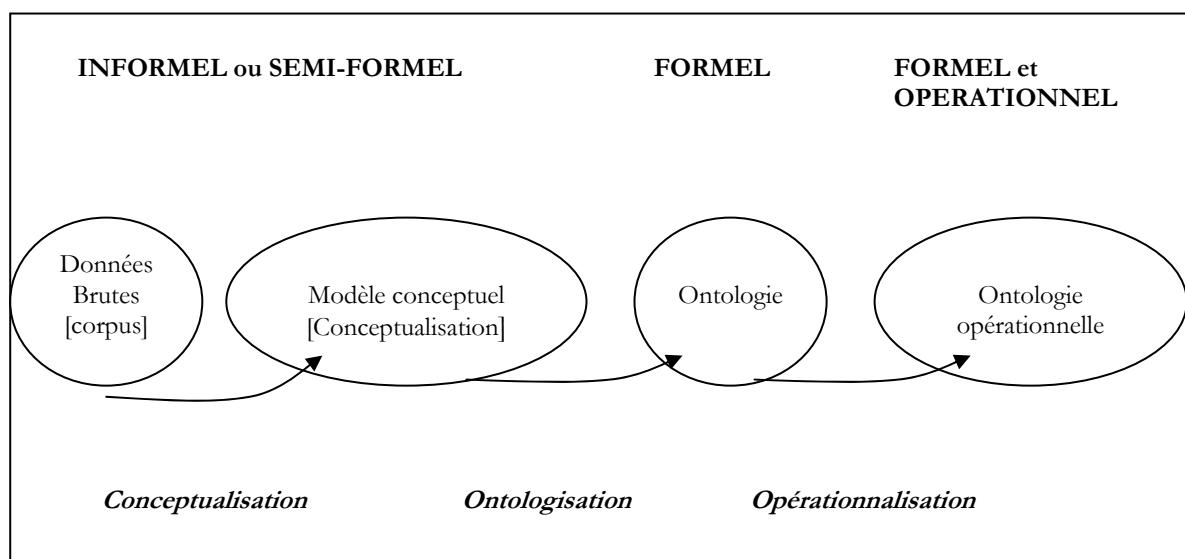
Enfin, une dernière classification peut s'effectuer en fonction du niveau de granularité, c'est-à-dire du niveau de détail des objets de la conceptualisation. Ainsi, selon l'objectif opérationnel de l'ontologie, une connaissance plus ou moins fine du domaine est nécessaire et des propriétés considérées comme accessoires dans certains contextes peuvent se révéler indispensables pour d'autres applications. On peut relever alors deux types de granularités :

- **Granularité fine** : cela correspond à des ontologies très détaillées, possédant un vocabulaire riche capable d'assurer une description détaillée des concepts pertinents d'un domaine
- **Granularité large** : cela correspond à un vocabulaire moins détaillé. Les ontologies de haut niveau ont par exemple une granularité large, car les notions sur lesquelles elles portent peuvent être raffinées par des notions plus spécifiques.

À travers ces différentes typologies qui peuvent chacune selon leur critère de base (composition, nature et granularité) qualifier une ontologie nous voyons qu'il y a parfois contradiction avec la définition que nous avons choisie comme étant celle de référence pour ce mémoire. En effet, pour Jean Charlet, Bruno Bachimont et Raphaël Troncy l'ontologie doit pouvoir se prêter au raisonnement automatique. Il faut donc obligatoirement qu'elle soit conçue dans un langage formel. Or, nous venons de le voir, Uschold et Gruninger conçoivent quant à eux la création d'ontologies non formelles.

Cette divergence de position, parmi d'autres, par rapport à la conception d'une ontologie illustre le flou terminologique qui règne autour des ontologies. On note que la vision correspondant à la définition que nous avons choisie est plus répandue chez les « ingénieurs », que ce soit au sein de la communauté de l'intelligence artificielle ou de celle du web sémantique.

Néanmoins, les étapes de développement d'une ontologie permettent d'avancer une hypothèse pour articuler ces deux acceptations. Sans en dire trop sur le sujet car cela sera développé dans le chapitre suivant, le processus d'élaboration d'une ontologie peut se schématiser selon Frédéric Fürst [FÜRST 2004] de la manière suivante :



1. Construction d'une ontologie opérationnelle

D'après ce schéma, il est donc possible de concilier les deux visions. L'ontologie écrite en langage formelle est le produit fini mais il y a tout un chemin qui mène à cette formalisation.

Ainsi, on peut relever très simplement qu'une ontologie permet de définir, de conceptualiser de manière plus ou moins formelle et plus ou moins précise un domaine de connaissances. Très schématiquement, on pourrait résumer les choses à l'extrême et dire qu'une ontologie est un vocabulaire contrôlé et organisé, pouvant, selon sa formalisation être interprétable par une machine. Mais est-ce vraiment cela une ontologie, n'est-ce que cela ? Pouvons-nous la considérer vraiment comme un nouveau vocabulaire ? Que propose-t-elle de nouveau par rapport aux autres ? Car la définition extrêmement simplifiée, voire simpliste, qui vient d'en être énoncée ne conviendrait-elle pas aussi à une terminologie, à une taxinomie ou encore à un thésaurus ?

III. Ontologie et langage documentaire

Même si pour les besoins du mémoire nous avons décidé d'une définition bien précise pour dire ce qu'est une ontologie, il est indéniable qu'il existe un flou autour de ce terme et de ce qu'il désigne. L'objectif de mon stage, ne le perdons pas de vue, est de transformer le thésaurus PACTOLS en ontologie. Cela sous-entend qu'il y a un lien entre ces deux notions. Le but de cette partie ne sera pas de trancher sur la question : l'ontologie est-elle un nouveau langage documentaire, mais plutôt d'exposer les différentes positions qui peuvent être prises face à cette question puis de situer les ontologies par rapport aux langages documentaires existants et notamment par rapport aux thésaurus. Pour se faire, nous nous intéresserons dans un premier temps à l'évolution des langages documentaires, pour étudier ensuite les rapports qui peuvent exister entre une ontologie et ces derniers.

1) Les langages documentaires

D'après la définition qu'en donne l'AFNOR, *un langage documentaire* est un langage artificiel, constitué de représentations de notions et de relations entre ces notions et destiné, dans

un système documentaire, à formaliser les données contenues dans les documents et dans les demandes des utilisateurs.

Le langage a donc un rôle de médiateur, il doit faire coïncider le langage naturel de l'utilisateur avec le langage d'indexation du document. C'est un outil qui sert de pont entre les utilisateurs ou les indexeurs, et les informations contenues dans les textes qu'ils recherchent ou indexent.

La fonction générale des langages documentaires est de faciliter le repérage des documents par sujets et d'aider à la recherche d'information. Ils ont pour but de répondre aux ambiguïtés du langage naturel et d'éviter ces pièges comme la synonymie ou la polysémie. Ils permettent donc trois choses :

- normaliser l'indexation des documents
- naviguer entre les sujets voisins
- présenter le contenu des documents pour faciliter une sélection

Par ailleurs, il est convenu dans le domaine de la documentation, au sens large du terme, de classer les langages documentaires selon deux catégories : les langages à structure hiérarchique et les langages à structure combinatoire. Les langages à structure hiérarchique sont basés sur la règle logique qui va des concepts génériques aux concepts spécifiques. Il y a donc un rapport de hiérarchies entre les termes, d'où leur nom. En outre, ils n'utilisent pas le langage naturel, mais un code basé sur les chiffres, les lettres et autres symboles, comme c'est le cas pour la classification Dewey ou encore pour la CDU, la Classification Décimale Universelle. Ce premier type de langage se subdivise en plusieurs catégories : les plans de classement, les classifications universelles et les classifications spécialisées, relatives à des domaines particuliers comme les sciences naturelles, les normes, etc. En ce qui concerne les langages à structure combinatoire, ils sont basés sur l'utilisation du langage naturel. Parmi ceux-ci on retrouve les lexiques (ou listes de mots-clés, listes de mots vedettes) et les thésaurus.

Langages documentaires	
Langages à structure hiérarchique	Langages à structure combinatoire
<ul style="list-style-type: none">- Plans de classements- Classifications universelles- Classifications spécialisées	<ul style="list-style-type: none">- Lexiques (listes de mots-clés, listes de mots vedettes)- Thésaurus

2. Tableau récapitulatif du classement des langages documentaires

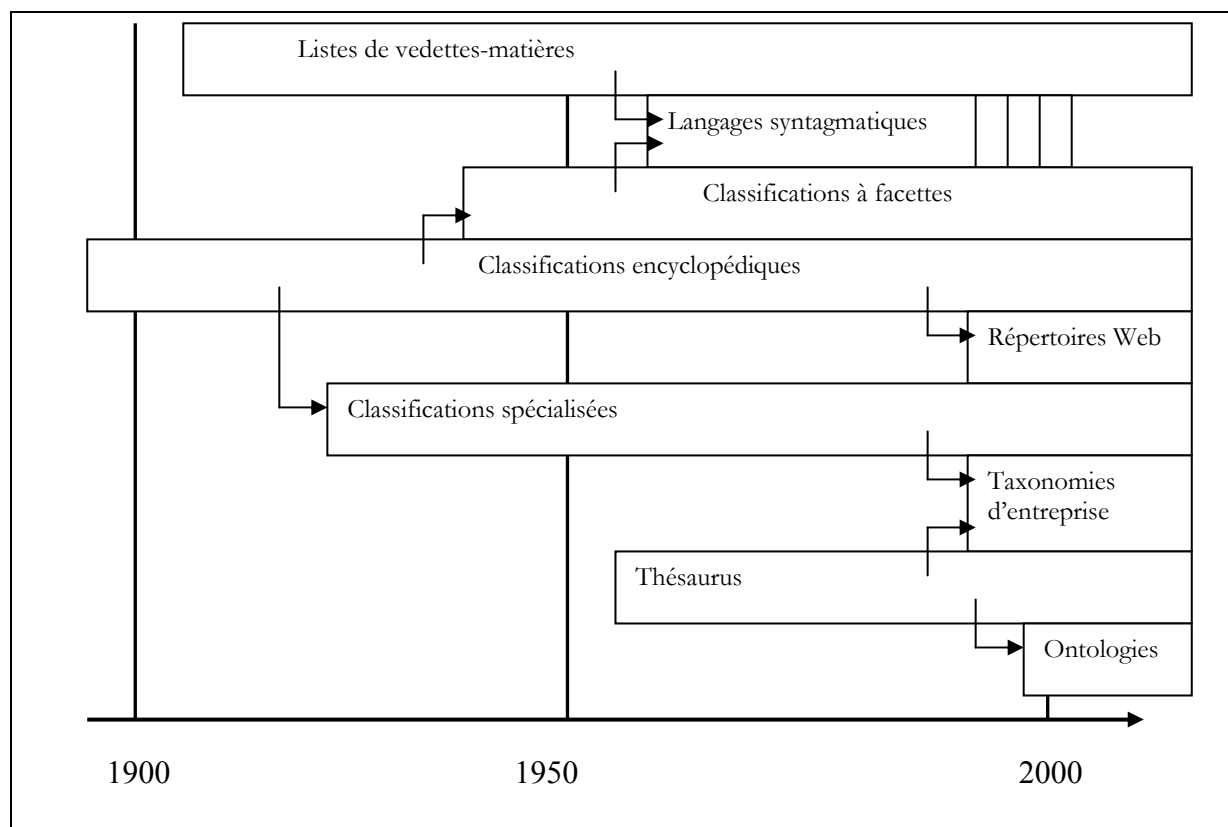
Enfin, pour concevoir ces langages documentaires sous un angle historique, Bruno Menon, dans son article « Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan » paru en 2007 dans la revue *Documentaliste – Sciences de l'information* [MENON 2007] distingue trois types de langages documentaires traditionnels :

- **Les classifications** : elles remontent au début du XIXe siècle. Elles ont été conçues, comme la classification de Dewey, pour permettre un arrangement rationnel des ouvrages sur les rayons des bibliothèques, elles facilitent l'accès thématique aux ouvrages et simplifient également la gestion de ces accroissements.
- **Les listes de vedettes matières** : elles font leur apparition à l'aube du XXe siècle et prospèrent à la faveur de l'accélération de la publication et de la circulation des savoirs sous forme imprimée. L'établissement de catalogues imprimés, avec leur indexation matière, permet la diffusion des informations bibliographiques et favorise l'émergence des premiers « réseaux » de bibliothèques. La publication des travaux scientifiques se fait de plus en plus sous forme d'articles de périodiques, d'où une distinction entre unité documentaire et unité physique qui accentue la nécessité des documents secondaires. Les registres et catalogues imprimés sont de plus en plus délaissés au profit des catalogues sur fiches, pour lesquels un classement alphabétique par sujets devient possible grâce à cette forme d'indexation.
- **Les thésaurus** : leur développement a suivi la diffusion des techniques informatiques, vers la fin des années cinquante. La stabilisation de leur modèle sous la forme qu'on leur connaît aujourd'hui, est l'aboutissement de réflexions autour de trois préoccupations sur l'organisation d'une documentation qui prolifère, qui n'est plus strictement scientifique mais de plus en plus technique¹², et pour laquelle les langages documentaires existants sont inadéquats. En premier lieu, il s'agissait de concevoir un système compact, qui ne demandait pas de répertoire *a priori* de nombreux concepts, mais qui autorisait l'expression d'un très grand nombre de sujets. En deuxième lieu, il importait, en prenant en compte la synonymie, d'harmoniser le vocabulaire des auteurs, celui des indexeurs et celui des utilisateurs. Enfin, il fallait imaginer des moyens de guider indexeurs et utilisateurs dans le choix des termes appropriés, ce que vise la structuration sémantique des termes par les liens hiérarchiques et associatifs. Ainsi, on arrive aux trois caractéristiques définitives du thésaurus : un langage structuré, contrôlé et combinatoire.

¹² Les grandes entreprises deviennent alors des protagonistes du champ documentaire, et c'est en leur sein que les thésaurus et leurs précurseurs sont élaborés.

Ces trois types de langage sont aujourd'hui, encore d'actualité et largement utilisés, même si comme le relève Bruno Menon [MENON 2007], les deux premiers sont davantage maîtrisés par les bibliothécaires, si le troisième demeure une sorte de « domaine réservé » des documentalistes, au moins dans le modèle français de cloisonnement professionnel qui est encore en vigueur. Sur ce point, le cas de FRANTIQ et de son thésaurus PACTOLS semble être l'exception qui confirme la règle, car tous les bibliothécaires du réseau utilisent le thésaurus pour indexer leurs documents.

On peut remarquer et comprendre à la vue de ce bref panorama historique que les langages documentaires ont évolué en fonction d'un besoin. Il y a eu renouvellement, évolution des langages parce que le contexte documentaire et informationnel changeait. En outre, B. Menon voit également un héritage entre les différents langages même s'il n'est pas toujours revendiqué. L'examen des parentés structurelles lui a permis de proposer cette généalogie :



3. Généalogie « des langages documentaires »

Ainsi, même si l'on ne décide pas de tous les nommer « langages documentaires » – comme nous allons le voir dans quelques lignes – ces systèmes témoignent de la renaissance des pratiques de description des contenus et d'organisation des connaissances. On remarque que les uns proviennent des autres, qu'il y a un dynamisme, que l'un se construit par rapport à l'autre. Il y

a une véritable progression mais sans destruction des anciens langages. Ainsi, le schéma nous permet de voir que les ontologies sont les dernières arrivées dans le domaine.

Nous avons relevé, il y a peu, que le renouvellement s'effectuait lorsque l'environnement informationnel évoluait : est-ce bien le cas pour les ontologies ? Y a-t-il un nouveau besoin ? Qu'apportent les ontologies que le thésaurus n'apporte pas ? Autant de questions auxquelles nous allons tenter de répondre dans la partie suivante.

2) Les ontologies et les langages documentaires

Même si nous avons fait notre choix, nous avons vu que la définition des ontologies faisait encore débat aujourd'hui, et qu'il régnait bien un flou terminologique autour de cette notion. C'est de ce flou que peut naître la confusion entre les ontologies et les autres langages documentaires, mais nous pouvons également nous en servir pour « aménager » une définition à notre convenance.

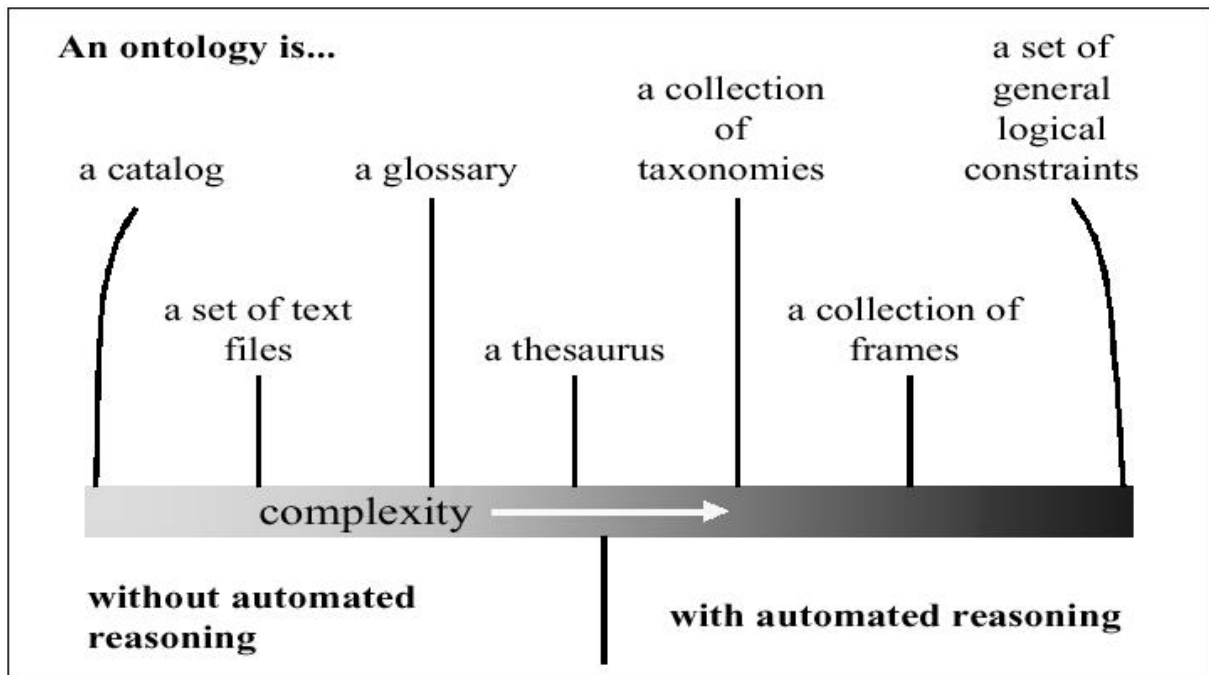
Les définitions les plus souvent citées sont les suivantes :

- « *An ontology is an explicit specification of a conceptualization.* [GRUBER 1993b] » (Une ontologie est une spécification explicite d'une conceptualisation)
- « *An ontology is a formal, explicit specification of a shared conceptualisation.* [STUDER et al. 1998] » (Une ontologie est une spécification formelle explicite d'une conceptualisation partagée.)

Elles se ressemblent beaucoup et se complètent. La seconde introduit les mots : *formal* (formelle) et *shared* (partagée) et en cela elle est plus spécifique et plus contraignante que la première – en outre, la seconde peut contenir la première. Cependant, c'est surtout l'ajout de *formal* qui les différencie mais qui complique également les choses car cette notion étant elle-même ambiguë comme le fait remarquer Bruno Menon [MENON 2007] en citant Guarino : « *An ontology is sometimes called a formal ontology, although we shall use the expression « formal ontology » only to refer to philosophical research field.* » (Une ontologie est parfois appelée ontologie formelle, mais nous utiliserons l'expression « ontologie formelle » uniquement pour nous référer à un champ philosophique de recherches.) En accord avec la première définition proposée par Thomas R. Gruber, nombreux sont les auteurs qui parlent d'ontologies informelles ou d'ontologies semi-

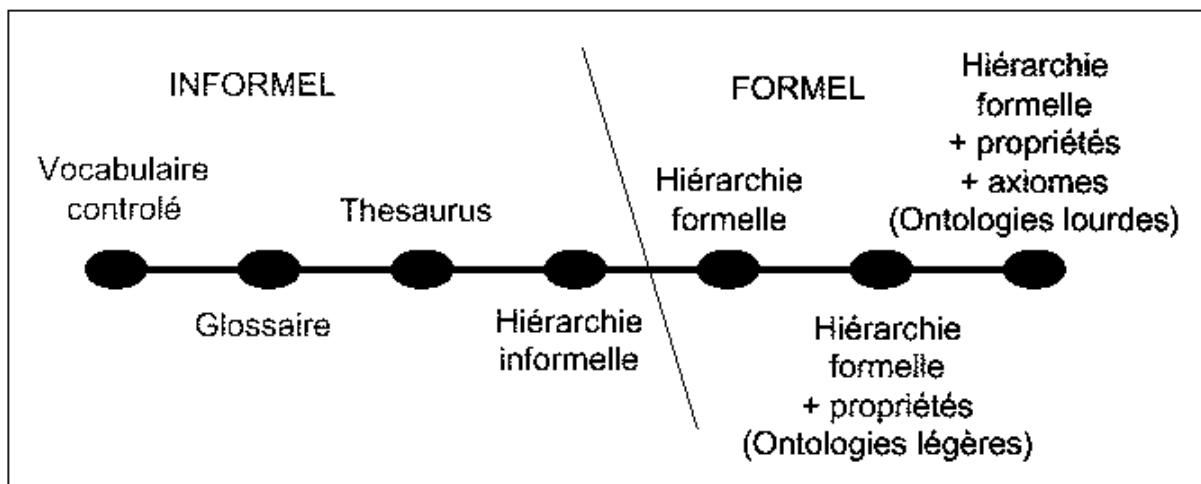
formelles. Nous l'avons d'ailleurs fait remarquer dans la partie précédente, avec Uschold et Gruninger. Ces auteurs se trouvent en général parmi ceux qui revendiquent la continuité entre l'ontologie philosophique et les ontologies en ingénierie des connaissances.

En considérant les choses de cette façon, certains langages documentaires peuvent ainsi être rangés dans la classe des ontologies comme le montre le schéma suivant proposé en particulier par Barry Smith et Christopher Welty [SMITH et WELTY 2001] :



4. Objets pouvant être qualifiés d'ontologies

Ce schéma a ensuite été repris et modifié dans un premier temps par Deborah McGuinness qui y ajoute une délimitation tranchée entre les ressources que l'on peut dénommer *ontologies*, obligatoirement formelles et d'autres types de ressources, puis dans un second temps par Frédéric Fürst [FÜRST 2004] qui stipule de manière plus explicite le caractère obligatoirement formel de l'ontologie.



5. Les différents niveaux de formalisme et d'engagement sémantique en représentation des connaissances

Nous l'avons dit précédemment, cette vision de l'ontologie est davantage répandue parmi les ingénieurs. Ce schéma nous permet notamment d'amorcer une réponse à l'une des questions posée en début de partie : l'ontologie est différente des thésaurus – et fort heureusement, sinon ma mission de stage n'aurait eu aucun sens. Mais avant d'aller plus loin dans cette réponse et de dire ce qui les différencie, souvenons-nous que nous avons également évoqué la possibilité, pour une ontologie, d'être assimilée à un langage documentaire : que pouvons nous en dire ?

L'assimilation des ontologies aux langages documentaires peut s'expliquer par au moins deux raisons. D'une part, il existe une certaine homologie fonctionnelle entre eux. Dans la vision du web sémantique par exemple, les ontologies doivent servir de cadre et de contrôle de référence pour l'expression des métadonnées et leur interprétation par des applications dites « intelligentes ». Elles jouent donc dans ce contexte un rôle similaire à celui des langages documentaires dans les systèmes d'information bibliographiques. D'autre part, il existe une homologie structurelle apparente avec les langages classificatoires. En effet, la hiérarchie de classes de concepts est l'élément central d'une ontologie, en particulier, nous le verrons, dans une ontologie exprimée avec le langage OWL. Mais malgré ces deux raisons, certaines personnes comme Bruno Menon ou encore Roberto Poli pensent qu'il est inexact d'assimiler de quelque manière que ce soit ontologie et langages documentaires : « *An ontology is not a catalogue of the world, a taxonomy, a terminology or a list of objects, things or whatever else. If anything, an ontology is the general framework (= structure) within which catalogues, taxonomies, terminologies may be given suitable organization* [POLI 1996]. » (Une ontologie n'est pas un catalogue du monde, une taxinomie, une terminologie ou une liste d'objets, de choses ou de quoi que ce soit d'autre. Une ontologie serait plutôt le cadre

général – la structure – au sein duquel catalogues, taxinomies, terminologies peuvent recevoir une organisation appropriée.)

Dans cette conception, que je partage également, une ontologie est avant tout un modèle pour la description ou la définition formelle des concepts de diverses natures, de manière à en permettre la manipulation par des automates.

Ainsi, un thésaurus peut être transformé en ontologie. Elle en sera alors le réceptacle selon ses propres conditions, ses propres règles comme nous allons maintenant le voir en nous intéressant à ce qui différencie un thésaurus d'une ontologie, ce qui nous permettra par ailleurs d'achever – si cela est possible – notre définition des ontologies.

Il est indéniable que le thésaurus, dernier produit terminologique en date, entretient des relations étroites avec les ontologies, mais leurs points communs et leur complémentarité ne doivent pas masquer les spécificités de chacun. Pour marquer leur différence, nous nous intéresserons uniquement à leur mode de conception et à leur contenu puisque c'est le sujet sur lequel porte essentiellement le mémoire.

D'après la définition qu'en donne Jean Charlet dans son mémoire d'habilitation à diriger des recherches [CHARLET 2002], *un thésaurus est un ensemble de termes normalisés fondé sur une structuration hiérarchisée. Les termes y sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques. Organisé alphabétiquement [et thématiquement], il forme un répertoire alphabétique de termes normalisés pour l'analyse de contenu, le classement et donc l'indexation de documents d'information (dans de nombreux cas, les thésaurus proposent aussi une définition des termes utilisés).*

Comme nous l'avons dit en introduction, sa construction et son développement sont régis par les normes ISO 2788 [ISO 2788:1986] et ISO 5964 [ISO 5964: 1985]. Celles-ci expriment son rôle comme outil d'indexation, le processus de construction et définissent notamment les types de relations sémantiques pour structurer le champ de connaissances. Ces relations sont de trois types :

- Les relations d'appartenance et d'équivalence qui gèrent les relations de synonymie et d'antonymie
- Les relations hiérarchiques qui se subdivisent en relations génériques et relations partitives
- Les relations associatives qui réunissent des termes qui sont mentalement associés.

Les relations associatives suscitent des extensions très diverses et permettent d'associer deux ou plusieurs entités sous plusieurs angles mais ces extensions ne peuvent jamais être explicitées. On ne peut pas décliner les relations associatives, les spécifier clairement. On ne peut pas dire pour quelle raison la pluie et l'érosion peuvent être liées dans un thésaurus, c'est à l'utilisateur d'interpréter. Par ailleurs, le contenu sémantique du thésaurus se résume à des concepts et classes, et aux relations prédéfinies mentionnées à l'instant.

C'est donc à toutes ces « limites » que l'ontologie peut apporter des réponses, des enrichissements, des améliorations. Très brièvement car nous en reparlerons dans le chapitre suivant, nous pouvons dire d'une part, qu'à la différence des thésaurus, les relations sémantiques dans une ontologie ne sont pas standards ni normalisées. Elles sont conçues en fonction de l'objectif à atteindre et du type d'application à laquelle l'ontologie est destinée. D'autre part, en ce qui concerne le contenu sémantique de l'ontologie, ce dernier est plus riche que celui du thésaurus. On y distingue certes des concepts ou objets, des classes de concepts, des types de relations (plus variées et explicitées) mais aussi des propriétés pouvant être attachées aux objets comme par exemple leur matière, couleur, taille, situation géographique, etc., des instances de concepts ou encore des règles d'inférences.

Ainsi, les ontologies offrent plus de richesse conceptuelle que les thésaurus en structurant de manière plus formelle et en établissant des liens plus profonds entre les concepts.

Peu à peu, au cours de ce premier chapitre nous avons pu définir ce qu'était une ontologie en partant des origines du principe : la représentation de l'être puis des étants, du terme en philosophie pour aller jusqu'au début des années 1990 et découvrir ce que cela signifiait en Ingénierie des connaissances – en considérant cette communauté dans toute son étendue : c'est-à-dire comme domaine touchant aussi bien à l'informatique, à l'intelligence artificielle, ou encore à l'univers de la documentation à travers les langages documentaires. D'ailleurs, au travers de ce dernier point, en établissant les différences entre un thésaurus et une ontologie, nous avons pu entrevoir rapidement de quoi était composé une ontologie. Il va s'agir maintenant, en abordant la construction des ontologies d'aller dans un premier temps plus en avant sur ce point.

CHAPITRE II : COMMENT CONSTRUIRE UNE ONTOLOGIE ?

Ce qui est paradoxal avec les ontologies, c'est que plus on essaye de savoir ce que c'est, plus on se perd. En effet, ce qui ressort incontestablement du premier chapitre c'est la polysémie du mot, et le flou terminologique qui en découle. Par souci de cohérence, nous avons opté pour une définition précise en deux parties proposée par Jean Charlet, Bruno Bachimont et Raphaël Troncy [CHARLET *et al.* 2004], que je rappelle ici :

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

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

[...]

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

[...]

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

C'est donc en nous appuyant sur cette définition que nous allons maintenant aborder le deuxième volet de ce mémoire : Comment construire une ontologie, qui va ensuite nous permettre dans un troisième et dernier chapitre de nous intéresser de manière plus aguerrie et impliquée à la réalisation de ma mission de stage : la transformation du thésaurus PACTOLS en ontologie.

Aux détours de l'introduction et du premier chapitre nous avons eu l'occasion de comprendre que des normes régissaient la construction des thésaurus. Ces normes dont sont d'ailleurs issus les PACTOLS sont les normes ISO 2788 [ISO 2788:1986 et ISO 5964 [ISO

5964: 1985]. Elles s'engagent à exprimer le rôle du thésaurus comme outils d'indexation et garantissent ainsi l'uniformité de la description des fonds documentaires. Elles développent tout le processus et les méthodologies de construction des thésaurus monolingues et multilingues. Ces normes apportent des précisions quant à la nature des entités (descripteur masculin, singulier, noms composés ou unitermes, catégories grammaticales...), et définissent les types de relations sémantiques pour structurer un champ de connaissance.

En revanche, il n'existe pas de normes propres aux ontologies définissant tout le processus et les méthodologies de construction. Néanmoins, nous allons, dans le présent chapitre consacré à ce que l'on peut appeler l'ingénierie ontologique, tenter de le faire en donnant tout d'abord les grandes lignes directrices de la conception d'une ontologie, ce qu'elles doivent contenir, ainsi que la méthode que j'ai utilisée pour réaliser ma mission. Ensuite, nous verrons à travers l'étude des différents langages utilisés autour des ontologies qu'il y a tout de même un souci de normalisation quant à leur création ; et enfin, nous étudierons quels sont les moyens logiciels qui nous permettent de les créer informatiquement.

I. Conception

À l'aide des éléments de définition exposés dans le chapitre précédent, nous avons peu à peu découvert, ou supposé parfois, ce qui composait une ontologie. Néanmoins, à ce stade du mémoire et de nos connaissances, il est temps de se demander comment concevoir une ontologie. Je préfère utiliser le terme *concevoir* plutôt que *construire* pour parler de cette première partie parce que le second sous-entend une action, l'acte de créer, d'élaborer. En revanche, *concevoir* est plus théorique, plus abstrait si l'on peut dire. Dans le dictionnaire le *Petit Robert*, on trouve d'ailleurs cette définition : « Concevoir : avoir une idée de, une représentation de [...] »

C'est exactement l'objectif de cette partie : avoir une idée de comment construire une ontologie ; avoir toutes les notions, avoir « les clés en main » pour ensuite aborder véritablement la construction d'une ontologie. Si nous regardons la figure 1, *Construction d'une ontologie opérationnelle*, présente dans le premier chapitre¹³, nous nous situons dans les deux premières phases, la conceptualisation et l'ontologisation. Ainsi, dans un premier temps, nous détaillerons la constitution d'une ontologie, quels sont les éléments qui la composent, puis, dans un second

¹³ Figure 1, *Construction d'une ontologie opérationnelle*, chapitre I, 2) b, page 27

temps nous verrons comment faire pour concevoir puis construire une ontologie à travers la méthode de Bruno Bachimont.

1) Constituants d'une ontologie

À partir des définitions proposées jusque-là, et notamment à partir de celle de J. Charlet, B. Bachimont et R. Troncy, trois grands types de caractéristiques nous permettent de préciser ce qui peut être représenté dans une ontologie considérée en tant qu'objet informatique :

- Les concepts
- Les propriétés
- Les relations

L'hypothèse de départ, issue de la théorie des modèles¹⁴ est qu'il existe des objets individuels qui peuvent être énumérés. Les concepts et les relations de l'ontologie sont organisés sous une forme hiérarchique qui admet une relation de subsomption¹⁵.

Les concepts

Selon François Rastier, chercheur au CNRS, « un concept est le signifié d'un mot dont on décide de négliger la dimension linguistique. Si bien qu'un concept n'est pas la source du terme, mais le produit de son instauration. C'est le travail terminologique qui transforme la notion en concept. Après avoir relié à une même notion plusieurs expressions jugées synonymes, on institue un concept en choisissant l'une d'entre elles comme terme, et en reléguant les autres au rang de pseudo-termes [RASTIER 1995]. ».

Les connaissances portent sur des objets auxquels on fait référence à travers des concepts – dans certains travaux, on trouvera à la place de Concept, Classe. Un concept, selon Uschold et King [USCHOLD et KING 1995], peut représenter un objet matériel (par exemple une épée, un poignard, etc.), une notion (par exemple, la quantité) ou bien une idée.

¹⁴ C'est une théorie de la vérité mathématique. Elle consiste essentiellement à dire qu'une théorie est mathématiquement valide si on peut définir un univers dans lequel elle est vraie.

¹⁵ C'est une relation « est-un » ou encore « is-a »

Un concept peut se diviser en trois parties : un terme (qui se désigne sous le nom de Label), une notion et un ensemble d'objets. Le label d'un concept est l'expression linguistique utilisée couramment pour y faire référence. La notion désigne ce qui est appelé, au sens de la représentation des connaissances, l'intension d'un concept. Elle contient sa sémantique qui est définie à l'aide de propriétés, d'attributs, de règles et de contraintes. L'ensemble d'objets forme l'extension du concept. Il s'agit des objets auxquels le concept fait référence, autrement dit, ses instances¹⁶. Par exemple, le label du concept « pomme » considérée comme un fruit, renvoie aussi bien à la notion de pomme en tant qu'aliment que l'on mange, qu'à l'ensemble des objets de ce type : Golden, Cybel, Granny Smith, Elstar, etc.

Les propriétés

Les propriétés – des attributs, dans le contexte des langages à objets – sont des caractéristiques liées aux concepts. Ainsi, une ontologie est non seulement le repérage et la classification des concepts mais c'est aussi des caractéristiques qui leur sont attachées. Ces caractéristiques peuvent être évaluées. Pour illustrer ce que sont les propriétés, J. Charlet, B. Bachimont et R. Troncy [CHARLET *et al.* 2004] prennent en exemple les sciences naturelles et leurs taxinomies¹⁷ : les vertébrés ont un tégument (la peau) comportant des poils – pour les mammifères – ou des plumes – pour les oiseaux. Dans une ontologie sur le monde animal, on pourra avoir les concepts de « mammifère » ou d'« oiseau » pour lesquels est précisé le type de tégument, respectivement à poil et à plume. En pratique, un attribut « tégument » pourra être attaché aux concepts et sa valeur variera suivant le concept auquel on fait référence.

Les relations

Les relations dans une ontologie représentent un type d'interaction entre les concepts, on peut en distinguer deux sortes :

- **la relation « est-un »** : cette relation de subsomption qui définit un lien de généralisation est utilisée pour structurer les ontologies. Elle permet fortement l'héritage de propriétés et c'est un choix qui s'impose depuis Aristote. Cette relation doit ensuite être complétée par d'autres relations pour exprimer la sémantique du domaine.

¹⁶ Les instances d'un concept sont des éléments singuliers.

¹⁷ Classifications dont la relation de base est une relation de subsomption.

- **les autres relations** : les relations unissent les concepts ensemble pour construire des représentations conceptuelles complexes. Si la connaissance ainsi construite correspond à un concept dans le monde modélisé, celui-ci est dit *défini*, à l'opposé des concepts insérés dans l'arborescence de l'ontologie qui sont dits *primitifs*. Par exemple, si l'on définit les lunettes comme étant des accessoires localisés sur le nez, c'est un concept défini. La relation « Localisés sur » est une relation binaire qui se définit par les concepts qu'elle relie et par le fait qu'elle est, comme les concepts, insérée dans une hiérarchie, une hiérarchie de relations cette fois-ci.

La relation « est-un » est une relation de même type que les autres, mais elle a de spécifique que c'est elle qui a été choisie comme relation de structuration de l'arborescence ontologique. Elle est implicite dans une ontologie.

En ce qui concerne les autres relations et les concepts, J. Charlet, B. Bachimont et R. Troncy [CHARLET' *et al.* 2004] insistent sur le choix du concepteur de l'ontologie. Ils tiennent à remarquer que les concepts et les relations de l'ontologie sont duals l'un par rapport à l'autre. Un concept primitif pourrait être un concept défini, une relation pourrait se retrouver implicitement définie au sein d'un concept primitif. Ce sont les choix assumés du concepteur de l'ontologie qui auront permis de décider de ce qui est essentiel – et donc primitif – ou non. Pour illustrer leurs propos, les trois auteurs donnent cet exemple : « on peut décider que le fait, pour un être humain, d'être un étudiant est temporaire donc non définitoire. On caractérise alors les êtres humains avec une relation de rôle social qui permettra de préciser une fonction d'étudiant ou de professeur. » Enfin, un autre choix de conception doit être fait durant l'élaboration d'une ontologie : décider si une connaissance doit être modélisée dans une propriété ou à l'aide d'une relation pointant sur un autre concept. Une propriété peut être déclarée dès lors que les valeurs possibles sont d'un type dit primitif (entier, chaîne de caractères), et c'est une relation dès lors que les valeurs possibles sont d'un type dit complexe, c'est-à-dire un autre concept de l'ontologie. Néanmoins, cette frontière peut elle aussi être remise en question.

Ainsi, nous savons exactement de quoi doit se composer une ontologie : de concepts, de propriétés, et de relations. Mais d'autres questions se posent alors : comment créer l'ontologie, comment passer de la théorie à la pratique, y a-t-il des techniques particulières pour collecter les concepts si nous ne les avons pas, comment les classer, comment éviter au maximum les erreurs de modélisation, de structuration, etc. ? Autrement dit, existe-il une ou des méthodes pour créer une ontologie ? La réponse à cette question est Oui. Il existe en effet quelques méthodes ou

éléments de méthode dans le domaine. Cependant, (par manque de temps certes), mais aussi parce que le but de l'exercice n'est pas de faire un relevé de tout ce qui existe dans le domaine de l'ingénierie ontologique, mais de trouver le juste milieu entre théorie et pratique, j'ai fait le choix de ne présenter qu'une méthodologie de construction d'ontologies, celle que j'ai utilisé pour réaliser ma mission.

2) Engagement sémantique et engagement ontologique pour la conception et réalisation d'ontologies [BACHIMONT 2000]

L'ingénierie des connaissances participe à rendre les ontologies interprétables par une machine. Elle est chargée de la « modélisation ontologique » – le fait de définir les primitives de représentation et leur signification qui seront utilisées pour la modélisation formelle des connaissances –, étape préalable à la modélisation formelle, c'est-à-dire, à la représentation dans un langage formel des connaissances du domaine.

Cette méthode que nous allons évoquer à présent, et qui sera celle utilisée pour la conception de l'ontologie à partir du thésaurus PACTOLS, est proposée par Bruno Bachimont¹⁸ qui s'appuie lui-même sur les travaux de François Rastier [RASTIER 1987] [RASTIER *et al.* 1994].

Elle comprend plusieurs étapes permettant de passer de l'expression linguistique des connaissances telles que nous autres, humains, pouvons la considérer à une représentation formelle et calculable des connaissances propres à une exploitation informatique. Cette transformation se fait par l'intermédiaire de deux engagements, l'un sémantique, et l'autre ontologique. En outre, elle permet notamment de résoudre, ou du moins propose des pistes pour résoudre, les problèmes soulevés précédemment au sujet du rôle du concepteur de l'ontologie par rapport au choix des primitives, ce qui relève de l'essentiel ou non.

¹⁸ Bruno Bachimont est Ingénieur Civil des Mines, Docteur en informatique de l'université de Paris 6, Docteur en épistémologie de l'École Polytechnique et habilité à diriger des recherches. Il est notamment enseignant-chercheur à l'Université de Technologie de Compiègne et directeur scientifique de l'Institut National de l'Audiovisuel (INA).

1^{re} étape : extraction de termes et analyse

L'un des premiers objectifs à atteindre lorsque l'on construit une ontologie est de définir les primitives du domaine, tout en sachant qu'il n'existe pas de primitives dans un domaine. Les primitives, comme les termes de tête d'un thésaurus, sont arbitraires. Il faut par conséquent, modéliser les primitives nécessaires à la formalisation et à la représentation du problème à résoudre et des connaissances s'y rapportant. La construction des primitives relève du choix du concepteur, mais comment procéder ? Pour y parvenir, Bruno Bachimont propose de repartir de l'expression linguistique des connaissances du domaine en utilisant l'extraction de termes à partir d'un corpus spécialisé. En effet, un corpus (constitué de documents propres au domaine concerné) comporte l'expression des notions qu'il faut modéliser. On peut ainsi construire un corpus textuel qui sera la source privilégiée permettant de caractériser les notions utiles à la modélisation d'une ontologie et le contenu sémantique qui lui correspond. Pour ce faire, B. Bachimont utilise ce qu'il appelle « une démarche corpus » et des outils terminologiques pour commencer la modélisation du domaine. Ces outils, pour la plupart, reposent sur la recherche de formes syntaxiques particulières manifestant les notions recherchées comme des syntagmes nominaux pour des candidats termes, des relations syntaxiques marqueurs de relations sémantiques, ou des proximités d'usage – comme les contextes partagés – pour des regroupements de notions.

En ce qui nous concerne, cette première étape n'est pas obligatoire puisque nous disposons déjà, à travers le thésaurus PACTOLS des candidats termes¹⁹. En revanche, cette étape peut être un avantage pour un pré-classement des termes et relations.

2^e étape : la normalisation sémantique

À la fin de la première étape, nous avons donc une liste de candidats termes dont les libellés ont un sens pour le spécialiste du domaine. Mais rien n'assure que ce sens soit unique : au contraire argumentent Jean Charlet, B. Bachimont, et Raphaël Troncy [CHARLET *et al.* 2004] car nous sommes dans un fonctionnement linguistique où les significations sont ambiguës, les définitions circulaires dépendent en particulier du contexte interprétatif des locuteurs. Or, dans la

¹⁹ Le travail de regroupement de termes et élaboration des concepts avait déjà été réalisé en 1987 pour la création des PACTOLS. 46 000 mots-clés ont été organisés, regroupés et hiérarchisés pour aboutir à 2500 descripteurs thématiques.

modélisation ontologique, on cherche à construire des primitives dont le sens ne dépend pas des autres primitives et est surtout non contextuel. Il faut dès maintenant prendre le chemin du formel en normalisant les significations des termes pour ne retenir, pour chacun d'eux, qu'une seule signification, qu'une seule interprétation par un être humain. C'est ce que permet l'utilisation de la *sémantique différentielle*, proposée par Bruno Bachimont. Cette sémantique, issue notamment des travaux de François Rastier [RASTIER 1987] [RASTIER *et al.* 1994], permet de décrire les unités entre elles par les identités qui les unissent et les différences qui les distinguent. Cela donne lieu à une définition de l'unité selon quatre principes différentiels :

- **le principe de communauté avec le père** : toute unité se détermine par l'identité qu'elle possède avec l'unité parente. Il faut expliciter en quoi l'unité fille est identique à l'unité parente. Il s'agit du principe aristotélicien de définition par le genre proche.
- **Le principe de différence avec le père** : toute unité se distingue de l'unité parente, sinon il n'y aurait pas lieu de la définir. Il faut donc expliciter la différence qui la distingue de l'unité parente. Il s'agit du principe aristotélicien de définition par la différence spécifique.
- **Le principe de différence avec les frères** : toute unité se distingue de ses frères, sinon il n'y aurait pas lieu de la définir. Il faut donc expliciter la différence de l'unité avec chacune des unités sœurs.
- **Le principe de communauté avec les frères** : toutes les unités filles d'une unité parente possèdent, par définition, un même trait générique, celle qu'elle partage avec l'unité parente. Mais il faut établir une autre communauté entre unités filles : celle qui permet de définir des différences mutuellement exclusives entre les unités filles. Bruno Bachimont donne cet exemple : l'unité parente est *être humain*, et les unités filles sont *homme* et *femme*. Ces unités partagent le fait d'être des humains. Mais cette propriété ne permet pas de définir en quoi les hommes et les femmes sont différents. On choisit alors comme principe de communauté la sexualité, on peut attribuer à *homme* le trait masculin, et à *femme* le trait féminin. Ces deux traits sont mutuellement exclusifs, car ce sont deux valeurs possibles d'une même propriété.

À la fin de cette étape²⁰, nous avons un arbre de primitives conceptuelles valable dans la seule région du monde modélisée où les concepts retenus correspondent bien à ceux de

²⁰ Pour avoir une représentation graphique ainsi que des exemples d'application de cette étape, voir dans les parties 2 et 3 du chapitre III de ce mémoire les propos et figures se rapportant à l'ontologie différentielle.

l'ontologie, par définition décontextualisée. Nous avons ce que B. Bachimont appelle une ontologie régionale.

3^e étape : l'engagement ontologique

L'engagement ontologique correspond à l'évolution de l'ontologie régionale vers une ontologie formelle. La sémantique formelle ne considère plus des notions sémantiques mais des extensions, c'est-à-dire l'ensemble des objets qui vérifient les propriétés définies en intension dans l'étape précédente, propriétés ayant une définition formelle à ce niveau. La structure est alors un treillis²¹.

Ce treillis de concepts doit être vu comme la possibilité de créer des concepts dits définis en combinant les concepts primitifs.

4^e étape : l'opérationnalisation

Il s'agit de la dernière étape de la création d'une ontologie, elle est généralement le fruit du travail d'un logiciel. Elle débouche sur la manipulation de l'ontologie par une machine. Elle devient un objet informatique. Enfin, cette étape consiste en la représentation de l'ontologie dans un langage de représentation de connaissances, ce qui nous amène directement à nous intéresser aux principaux langages de connaissances utilisés pour la création des ontologies formelles.

II. Langages de représentation des connaissances

Le chemin se poursuit vers la création d'une ontologie à partir du thésaurus PACTOLS. Nous savons ce qu'est une ontologie, ce qu'il faut mettre dedans et comment il est possible de l'organiser, c'est-à-dire sur quelle méthode s'appuyer. Nous avons d'ailleurs achevé la partie

²¹ Pour avoir une représentation graphique ainsi que des exemples d'application, voir dans les parties 2 et 3 du chapitre III de ce mémoire les propos et figures se rapportant à l'ontologie référentielle.

précédente sur l'ultime étape de cette méthode qui consistait en l'opérationnalisation de l'ontologie : représenter l'ontologie dans un langage de représentation des connaissances.

Audrey Baneyx, dans sa thèse *Construire une ontologie de la pneumologie : aspects théoriques, modèles et expérimentation* [BANEYX 2007], donne trois exigences auxquelles le langage de représentation des ontologies doit se soumettre :

- **la lisibilité** : le langage doit être compréhensible pour un utilisateur humain et doit donc avoir une certaine continuité avec le langage naturel
- **la portabilité** : le langage choisi doit être le plus standard possible afin de pouvoir être réutilisé dans d'autres systèmes
- **la possibilité de pouvoir faire des inférences** : le langage doit permettre le traitement informatique des données en vue de calculer les déductions logiques possibles

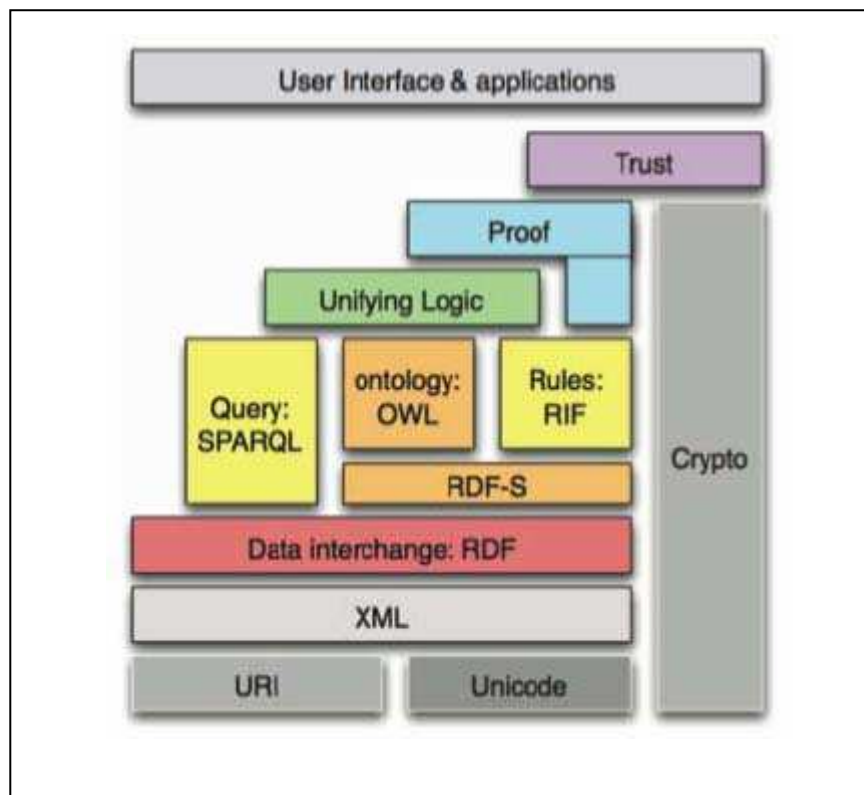
Par ailleurs, un langage d'ontologie doit permettre de signifier l'appartenance d'un objet à une catégorie, de déclarer la relation de généralisation entre catégories et de typer les objets que lie une relation.

Autrement dit, un langage de représentation pour les ontologies doit être capable de représenter toutes les subtilités de l'ontologie (concepts, propriétés et relations) et à l'heure d'Internet où les échanges sont démultipliés, mais aussi simplement par soucis d'interopérabilité et de normalisation, il se doit d'être le plus standard possible, et aussi le plus utilisé. En ce qui concerne la première exigence – la lisibilité – tout est relatif : si l'on n'est pas coutumier des langages de représentation en général, un langage de représentation pour les ontologies ne sera pas plus lisible que les autres.

Un langage répond actuellement à ces critères, il s'agit de OWL (*Ontology Web Language*) issu des travaux sur le web sémantique dirigés par le W3C. Ce langage occupe une place prépondérante dans le paysage des ontologies et est le standard le plus utilisé, c'est pourquoi nous en ferons la description dans un premier temps, avant de nous intéresser à un autre type de langage de représentation des connaissances : SKOS (*Simple Knowledge Organisation System*), qui, comme son nom l'indique est utilisé pour représenter de manière moins lourde les schémas de concepts.

1) OWL

OWL²², est issu d'un groupe de travail dédié au développement de langages standards pour modéliser des ontologies utilisables et interchangeables. Il a acquis le statut de recommandation du W3C le 10 février 2004. C'est un dialecte XML, fondé sur le standard RDF comme on peut le voir sur la figure suivante. RDF et OWL, standards du web sémantique, fournissent un cadre de travail pour la gestion des ressources, l'intégration, le partage et la réutilisation des données sur le Web.



6. Le « gâteau » de Tim Berners-Lee extrait du support de l'exposé de Raphaël Troncy donné à l'occasion d'une journée d'étude sur le Web Sémantique en juillet 2008

La vision du Web sémantique dans lequel l'information serait accessible et manipulable automatiquement par la machine, s'appuie sur une pile de langages jouant chacun un rôle particulier :

- **XML** fournit une manière de représenter des documents structurés, mais il n'impose aucune contrainte sémantique sur les documents produits ;

²² Owl signifie chouette, symbole de Minerve, déesse de la sagesse et de la guerre.

- **RDF** est un modèle de données simple, fondé sur des ressources et des relations entre ces ressources, équipé d'une sémantique et qui peut se représenter en XML ;
- **RDF Schéma** permet de définir le vocabulaire pour décrire des classes et des propriétés hiérarchisées en taxinomies ;
- **OWL** fournit davantage de primitives de modélisation pour décrire des ontologies plus riches sur le web.

Voici une présentation rapide de la structure, intéressons nous maintenant plus en détail à ces différents langages, en partant de XML jusqu'au principal concerné : OWL

XML²³

XML (eXtensible Markup Language) est un langage informatique de balisage *générique*. Il est recommandé par le W3C pour exprimer des langages de balisage *spécifique*, tels que RDF ou OWL. Par rapport au langage HTML (HyperText Markup Language), XML laisse la possibilité à son utilisateur de distinguer les données selon leur sens et leur contenu. Un document XML se présente sous la forme de données « taggées » par un ensemble de balises, chacune pouvant comporter des attributs et des valeurs. Il n'y a pas de définition figée des balises.

Son objectif initial est de faciliter l'échange automatisé de contenus entre systèmes d'information hétérogènes – on parle alors d'interopérabilité.

Il s'agit de ce qu'on peut appeler la « première couche » d'OWL.

RDF²⁴

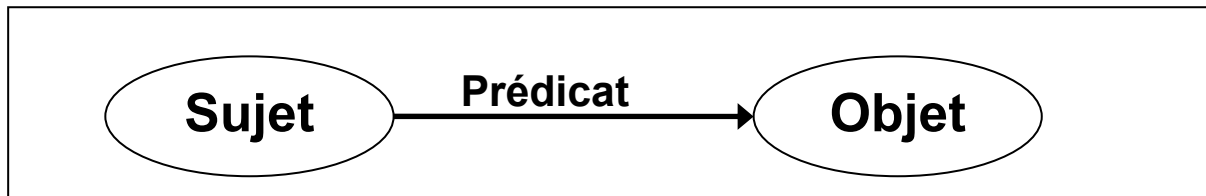
RDF (*Resource Description Framework*) est un modèle de représentation sémantique des informations du Web qui utilise la syntaxe XML. Il permet la mise en place de descriptions simples sur les ressources du Web comme les auteurs de pages Web, leur date de création, etc. Les ressources du Web sont l'élément de base de RDF. Chaque ressource est pourvue d'un identifiant uniforme de ressource (URI, *Uniform Resource Identifier*). Initialement recommandé par

²³ <http://www.w3.org/XML/>

²⁴ <http://www.w3.org/RDF/>

le W3C dans le but de standardiser les définitions et les usages des méta-données, RDF est également utile à la représentation de données elles-mêmes.

La structure fondamentale de toute expression en RDF est une collection de *triplets*, chacun composé d'un sujet, d'un prédicat et d'un objet. Un ensemble de tels triplets est appelé un graphe RDF.



7. Représentation d'un triplet RDF

Dans un graphe, chaque triplet représente l'existence d'une relation entre les choses symbolisées par les nœuds qui sont joints.

Les schémas RDF (RDFS²⁵) permettent de définir le vocabulaire utilisé dans les descriptions RDF. Il confère un formalisme de représentation riche incluant des classes, sous-classes, propriétés, sous-propriétés, des règles d'héritages de propriétés, etc., mais ne normalise pas les inférences que l'on pourrait faire avec. La structure objet-classe des RDFS permet de représenter un modèle du domaine en définissant des objets du domaine et leurs relations pour rendre compte d'une ontologie.

OWL

OWL²⁶ (*Web Ontology Language*) doit permettre de représenter des ontologies, en particulier sur le Web. Il est fondé sur la syntaxe RDF/XML et est dédié totalement à la représentation des ontologies.

OWL est destiné à être utilisé lorsque les informations contenues dans les documents doivent être traitées par des applications logicielles, c'est-à-dire lorsqu'elles ne sont pas simplement montrées à l'utilisateur. Une ontologie OWL est composée d'un en-tête (métadonnées), d'axiomes et de faits. Les axiomes concernent la définition complète ou partielle de concepts et

²⁵ <http://www.w3.org/TR/rdf-schema/>

²⁶ <http://www.w3.org/2004/OWL/>

de relations, la spécification de propriétés sur les relations (propriétés algébriques) et la définition d'axiomes sur les classes et les relations (équivalences, expression booléenne). Parmi les relations, on distingue celles dont le domaine de valeur sera de type primitif (attribut) de celles dont le domaine de valeur sera un autre concept (relation). Les faits concernent des individus pour lesquels on donne des valeurs aux propriétés des classes dont ils sont les instances.

Il existe trois sous-langages de OWL offrant des capacités d'expression croissantes : OWL Lite, OWL DL et OWL Full.

- **OWL Lite** est le sous langage de OWL le plus simple. Son principal intérêt est de permettre la modélisation d'ontologies simples, d'une complexité formelle peu élevée, de sorte qu'il soit facile d'implémenter des raisonneurs corrects et complets.
- **OWL DL** est plus complexe que OWL Lite. Il permet une expressivité plus importante. Il est fondé sur la logique descriptive (d'où son nom, *OWL Description Logics*), un domaine de recherche étudiant la logique, lui conférant son adaptation au raisonnement automatisé. Il garantit la complétude des raisonnements (toutes les inférences sont calculables) et leur décidabilité (leur calcul se fait en une durée finie).
- **OWL Full** est la version la plus complexe d'OWL, mais également celle qui permet le plus haut niveau d'expressivité. Il est destiné aux situations où il est plus important d'avoir un haut niveau de capacité de description, quitte à ne pas pouvoir garantir la complétude et la décidabilité des calculs liés à l'ontologie. OWL Full propose néanmoins la possibilité d'étendre le vocabulaire par défaut de OWL.

Enfin, il existe entre ces trois langages une dépendance de nature hiérarchique : toute ontologie OWL Lite valide est également une ontologie OWL DL valide, et toute ontologie OWL DL valide est également une ontologie OWL Full valide.

Il existe certes d'autres langages de représentations d'ontologies, mais OWL est celui qui tend à s'imposer aujourd'hui. En outre, c'est un *standard* du W3C. C'est-à-dire qu'à défaut d'être une norme ISO, il en est l'équivalent dans le monde des industriels.

Par ailleurs, le choix du langage est fondé principalement sur des critères techniques, notamment en fonction des degrés d'expressivité et de formalisme voulus. OWL répond parfaitement à ces critères et c'est ce que nous recherchons pour la réalisation de la mission, néanmoins il serait

intéressant d'étudier brièvement un autre langage qui constitue une alternative à OWL dans le cas d'un formalisme moins fort des connaissances : SKOS.

2) SKOS

Le projet de SKOS²⁷ (*Simple Knowledge Organisation System*) a été initié par l'Union européenne dans le cadre du projet SWAD-Europe²⁸. Ces travaux ont abouti en 2003 aux premières publications de *SKOS Core Guide* et *SKOS Mapping Guide*, ainsi qu'à une mise en application destinée à valider les technologies retenues en situation réelle. La réflexion a ensuite été reprise par le W3C dans le cadre du groupe de travail sur les bonnes pratiques et le déploiement des standards RDF.

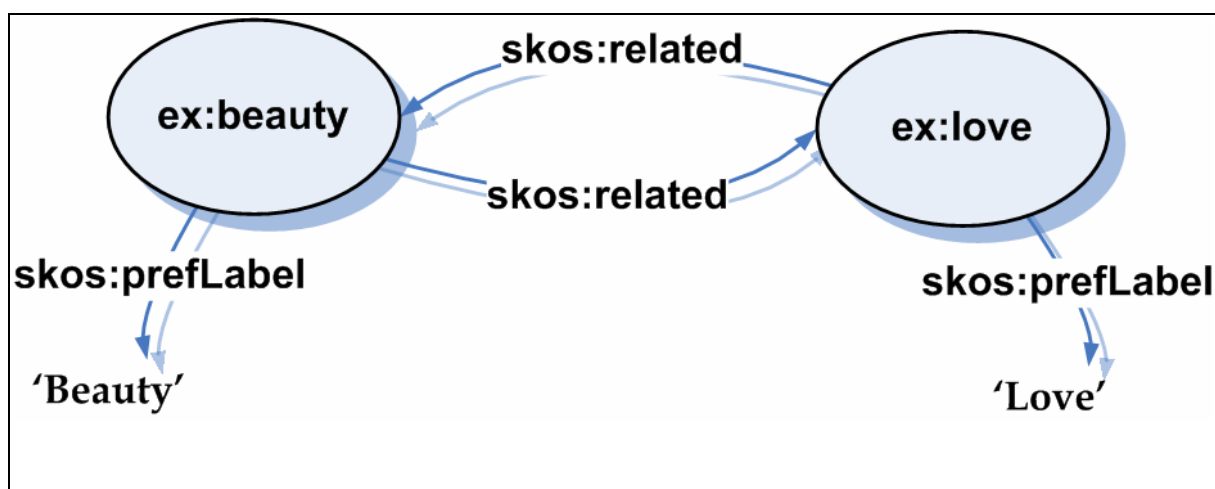
Il s'agit d'un langage de représentation de schémas de concepts. Comme son nom l'indique, il est destiné à proposer un système permettant d'exprimer et de gérer des modèles interprétables par des machines dans la perspective du web sémantique. On le dit « simple » par opposition notamment au langage OWL, qui comme nous l'avons vu est plus à même de représenter des structures plus riches : les ontologies. En effet, SKOS est avant tout destiné à la représentation de thésaurus, classifications, listes de vedettes matières, taxinomies ou autres folksonomies.

Son formalisme de représentation repose sur les graphes RDF. Le concept constitue le centre du graphe auquel peuvent notamment être attachés en tant que propriétés RDF :

- **les indications portant sur le concept lui-même** : des termes préférentiels ou alternatifs, les équivalents dans d'autres langues ; des termes cachés ; et la représentation par une image
- **les différents types de notes** : notes de définition et d'application, exemples, notes historiques, etc.
- **les relations sémantiques** : hiérarchie et association

²⁷ <http://www.w3.org/2004/02/skos/>

²⁸ *Semantic Web Advance Development for Europe*



8. Représentation de la relation associative dans SKOS

Présenté de cette manière, SKOS pourrait convenir pour la transformation du thésaurus PACTOLS en ontologie. Cependant, la figure ci-dessus en nous donnant un exemple de modèle de représentation d'une relation d'association entre deux concepts, nous montre également son principal défaut pour la réalisation de la mission. En effet, on peut voir que la relation associative n'est pas spécifiée, on ne sait pas pourquoi *beauty* et *love* sont associés, et ce même en complétant ce *SKOS core* (le noyau) par les *SKOS Extensions* qui ne permettent pour le moment « que » de préciser si la nature d'une relation de hiérarchie est de type tout/partie ou classe/instance ou encore de préciser certains attributs d'un concept : une note historique par exemple. Or, l'un des critères d'évolution possible vers une ontologie est de pouvoir expliciter les relations entre les concepts.

Cela dit, la répartition entre ce qui figure le noyau et les extensions fait encore l'objet de discussions.

Ainsi, d'après cette brève présentation, nous comprenons bien que SKOS n'est pas destiné à représenter des ontologies qui ont une formalisation trop complexe. Néanmoins, il est intéressant de savoir qu'il existe un langage destiné à la représentation des thésaurus, notamment pour l'avenir du thésaurus PACTOLS qui ne se limite pas à la réalisation de la mission qui m'a été confiée durant trois mois, de le transformer en ontologie.

III. Éditeurs d'ontologies

```
<owl:Class>
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing rdf:about="#Eurasia"/>
    <owl:Thing rdf:about="#Africa"/>
    <owl:Thing rdf:about="#NorthAmerica"/>
    <owl:Thing rdf:about="#SouthAmerica"/>
    <owl:Thing rdf:about="#Australia"/>
    <owl:Thing rdf:about="#Antarctica"/>
  </owl:oneOf>
</owl:Class>
```

```
<owl:Class>
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class>
      <owl:oneOf rdf:parseType="Collection">
        <owl:Thing rdf:about="#Tosca" />
        <owl:Thing rdf:about="#Salome" />
      </owl:oneOf>
    </owl:Class>
    <owl:Class>
      <owl:oneOf rdf:parseType="Collection">
        <owl:Thing rdf:about="#Turandot" />
        <owl:Thing rdf:about="#Tosca" />
      </owl:oneOf>
    </owl:Class>
  </owl:intersectionOf>
</owl:Class>
```

9. Syntaxe OWL

Voici deux exemples de syntaxe en OWL, qui comme nous l'avons vu dans la partie précédente, est le langage standard pour représenter une ontologie. C'est donc sous cette forme que doit être transformée l'ontologie pour devenir interprétable par une machine, ultime étape.

Mais pour ceux qui n'ont pas le temps d'écrire toutes ces lignes ... ou pour ceux, et j'en fais partie, qui n'éprouvent aucun besoin particulier de s'essayer à l'apprentissage de l'écriture de tels langages, il existe fort heureusement des logiciels qui prennent en charge l'écriture des ontologies dans le langage formel de notre choix, ce sont des éditeurs d'ontologies.

Les éditeurs d'ontologies jouent donc un rôle d'intermédiaire entre le concepteur de l'ontologie et le langage informatique, c'est-à-dire qu'ils se chargent de la phase d'opérationnalisation de l'ontologie. Il en existe un certain nombre comme va le montrer l'exposé suivant qui consiste à faire une sorte d'état des lieux des éditeurs d'ontologies. Mais avant de commencer, il convient de préciser que cet exposé n'a pas la prétention d'être exhaustif. Il a pour objectif de nous permettre de voir ce qui existe, les différents modes de saisies proposés, de représentation ainsi que les différents langages de représentation utilisés et supportés.

Par ailleurs, faute de temps, les différents éditeurs présentés dans cette partie n'ont pas fait l'objet de tests de ma part, sauf pour Protégé et DOE qui ont été mes outils de travail lors de la réalisation de ma mission. Pour réaliser cet exposé, j'ai retranscrit la récente étude et synthèse effectuée par Audrey BANEYX dans sa thèse *Construire une ontologie de la pneumologie : Aspects théoriques, modèles et expérimentations*, présentée et soutenue en 2007 [BANEYX 2007].

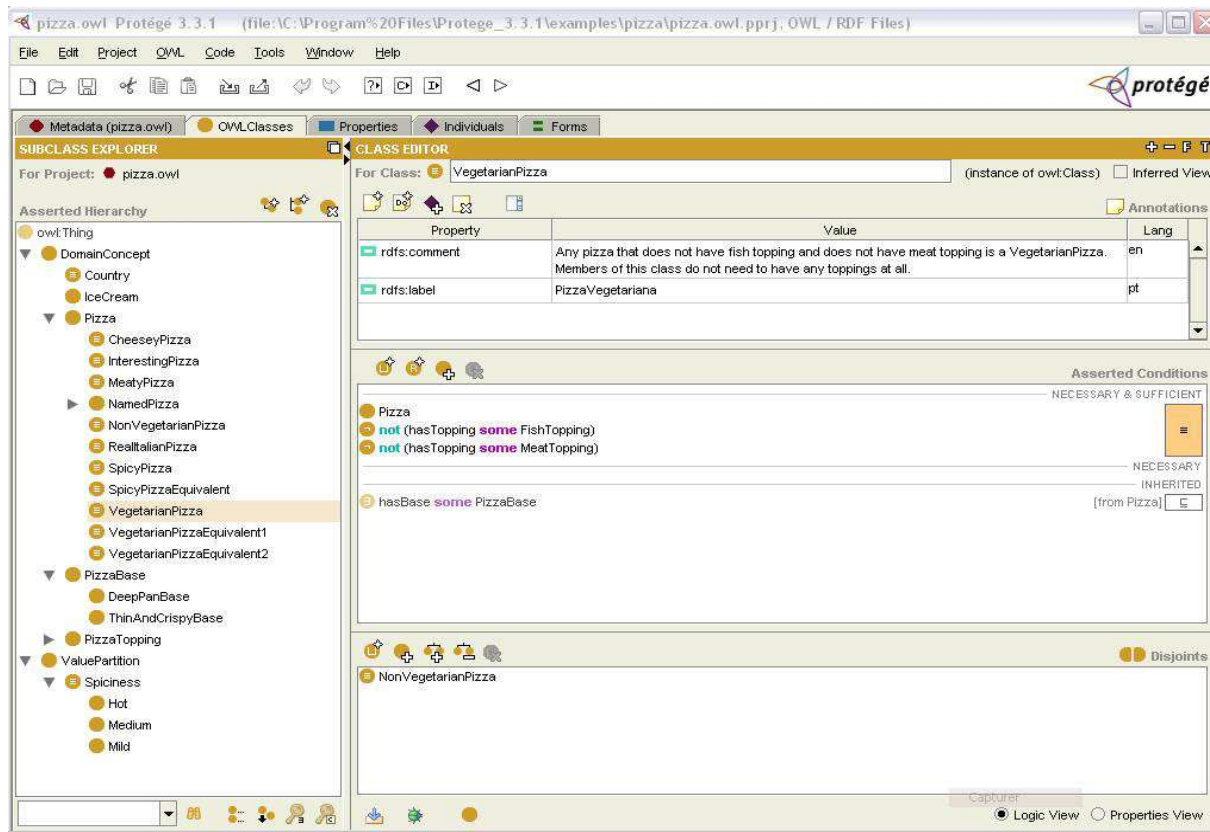
1) Protégé²⁹

Protégé a été développé par le *Stanford Medical Informatics* de l'université de médecine de Stanford depuis 1995. Il est construit autour d'un modèle de connaissances inspiré par le paradigme des frames : classes, slots (attributs) et facets (contraintes sur les attributs) qui sont les primitives de modélisation proposées. Ce modèle autorise une liberté de conception importante, puisque le contenu des formulaires de spécification des classes peut être modifié suivant les besoins, *via* un système de méta-classes, qui constituent des sortes de « patrons » pour les classes du modèle du domaine. Il est adapté à la construction d'ontologies depuis la version Protégé 2000. L'interface très complète ainsi que l'architecture logicielle bien pensée permettant l'insertion de plugins, notamment des plugins pour gérer les représentations sous forme graphique, par exemple OWLViz, ont grandement contribué au succès de Protégé. En quelques années, cet éditeur s'est imposé comme la référence, avec une communauté d'utilisateurs extrêmement importante et active. Ses nombreuses extensions lui permettent en particulier de gérer des langages standards comme RDF et surtout OWL, de créer des axiomes formels de manière intuitive, d'accéder aux ontologies par des interfaces graphiques évoluées, de comparer et fusionner des ontologies avec la suite PROMPT³⁰. Il est également possible de faire fonctionner des raisonneurs, comme RACER³¹ (*Renamed ABox and Concept Expression Reasonner*) pour le langage OWL par exemple, pour vérifier la cohérence et la consistance de la structure ontologique.

²⁹Appelé auparavant Protégé 2000, cet éditeur a repris le nom de l'outil d'acquisition des connaissances qui l'a précédé. <http://protege.stanford.edu/>

³⁰ <http://protege.cim3.net/cgi-bin/wiki.pl?Prompt>

³¹ <http://www.racer-systems.com/products/racerpro/index.phtml>



10. Copie d'écran de l'interface principale de l'éditeur d'ontologies Protégé

2) *Oiled*

L'éditeur OILED³² a été développé par l'université de Manchester pour éditer des ontologies dans les langages de représentations OIL, puis DAML+OIL, les précurseurs d'OWL. Il est donc explicitement orienté vers la représentation en logique de description expressive et, à ce titre, fournit tous les éléments d'interface permettant de spécifier des hiérarchies de concepts et de rôles, ainsi que la construction des expressions complexes définissant ces entités.

À l'origine il n'a pas d'autre ambition que d'illustrer les vertus du langage pour lequel il a été créé. Les versions disponibles de OILED ne constituent pas un environnement complet pour le développement d'ontologies d'envergure. En effet, cet outil n'implémente pas la migration et l'intégration d'ontologies, ne gère pas les différentes versions et autres activités impliquées dans la construction d'ontologies. Néanmoins, la simplicité, la robustesse de cet outil et la présence d'un raisonneur logique de description FaCT³³, capable de tester la faisabilité des ontologies construites ou d'expliquer de nouvelles relations de subsomption entre concepts complexes, en

³² <http://oiled.man.ac.uk/>

³³ FaCT Classification of Terminologies, <http://www.cs.man.ac.uk/horrocks>

font un outil de référence relativement populaire avec plus de 2000 téléchargements. Comme le soulignent les concepteurs, il s'agit plutôt d'un « bloc-notes » offrant assez de fonctionnalités pour permettre à des utilisateurs de construire les ontologies et en assurer l'uniformité. [...]

OILED permet d'exporter des ontologies construites dans des langages standards tels que DAML+OIL, RDFS ou OWL.

3) *OntoEdit*

ONTOEDIT est un outil mis au point par l'institut AIFB de l'université de Karlsruhe et qui est maintenant commercialisé par la société Ontoprise GmbH. Il s'inspire de l'approche par *frames* mais gère de nombreux formats libres de la communauté Web sémantique (FLogic, DAML+OIL, RDFS) et à le mérite de s'appuyer sur une réflexion méthodologique significative, celle d'On-To-Knowledge. Il s'est en effet le premier intéressé à la modélisation « intuitive » des axiomes, indépendamment d'un formalisme ou d'un autre, pour faciliter la traduction d'un langage de représentation à un autre. Cet outil met à la disposition de l'utilisateur plusieurs vues graphiques correspondant aux différentes phases de conception de l'ontologie. Il permet d'éditer une hiérarchie de concepts ou de classes. Ces concepts peuvent être abstraits ou concrets, ce qui indique s'il est permis d'instancier le concept en question.

S'efforçant de mettre en œuvre les propositions du projet On-To-Knowledge, il propose également une gestion originale des « questions de compétences ». Il s'agit de questions auxquelles les connaissances ontologiques doivent apporter une réponse. Un petit outil compare au niveau lexical les termes extraits des différentes questions pour en déduire automatiquement d'éventuelles subsomptions. Comme le fait remarquer Raphaël Troncy, ce procédé semble très peu fiable car il repose sur l'identification du nom du concept dans ses spécialisations. De plus, ONTOEDIT gère la synonymie en admettant plusieurs noms pour un même concept. Visiblement aucune distinction n'est faite entre le terme désignant le concept et ceux désignant les connaissances qu'il recouvre. ONTOEDIT est un des seuls éditeurs que nous connaissons, avec DOE, à s'attaquer au problème de la synonymie.

ONTOEDIT permet d'exporter les ontologies construites dans différents langages : RDF(S), OXML³⁴, DAML+IOL et FLogic.

³⁴ ONTOEDIT's XML-based Ontology representation Language

4) *WebODE*

WebODE³⁵ est une plateforme en ligne développée par le groupe *Ontological Engineering* du département d'Intelligence artificielle de la faculté d'Informatique de l'université polytechnique de Madrid. Elle se place au niveau méthodologique dans la lignée d'ODE, un éditeur qui assurait le support de Methontology, la méthodologie proposée par ce laboratoire. L'ambition nouvelle de WebODE par rapport à ODE est de considérer que les ontologies doivent être construites et mises à disposition *via* le web pour faciliter le développement d'application du web sémantique. WebODE est composé de plusieurs modules : un éditeur d'ontologies qui intègre la plupart des services nécessaires à la construction d'ontologies (édition, navigation, comparaison, fusion, raisonnement...), un système de gestion des connaissances à base ontologique, un générateur automatique de portail du web sémantique, un outil pour annoter les ressources du web et un éditeur de services pour le web sémantique. La plateforme WebODE met l'accent sur la possibilité d'un travail collaboratif et sur la possibilité, comme dans Protégé, d'étendre la plateforme à l'aide de modules complémentaires, comme un moteur d'inférences ou bien l'outil ODEClean, [...], accepte l'export et l'import d'ontologies en RDFS, DAML+OIL et OWL.

5) *DOE*

DOE³⁶ (*Differential Ontologies Editor*) a été développé à l'Institut National de l'Audiovisuel par Raphaël Troncy et Antoine Issac. Si tous les outils précédemment présentés peuvent être considérés, en tout cas dans une première approche, comme satisfaisants en matière d'expressivité ou d'interface, il existe toujours un certain vide méthodologique. La structuration des taxinomies produites est, en particulier, très peu prise en charge : on ne peut pas dire que les outils existants guident réellement l'utilisateur lors de cette étape primordiale. La piste d'un traitement plus complet des informations véhiculées par le langage ne semble pas non plus pouvoir être suivie grâce à ces environnements, puisque les commentaires demeurent toujours accessoires.

L'éditeur DOE offre des interfaces de création, modification et suppression de concepts et de relations, une représentation graphique de l'arbre ontologique, et des fonctionnalités de recherche et de navigation dans la structure créée. Il propose une interface ergonomique pour associer à

³⁵ <http://webode.dia.fi.ump.es/WebODEWeb/index.html>

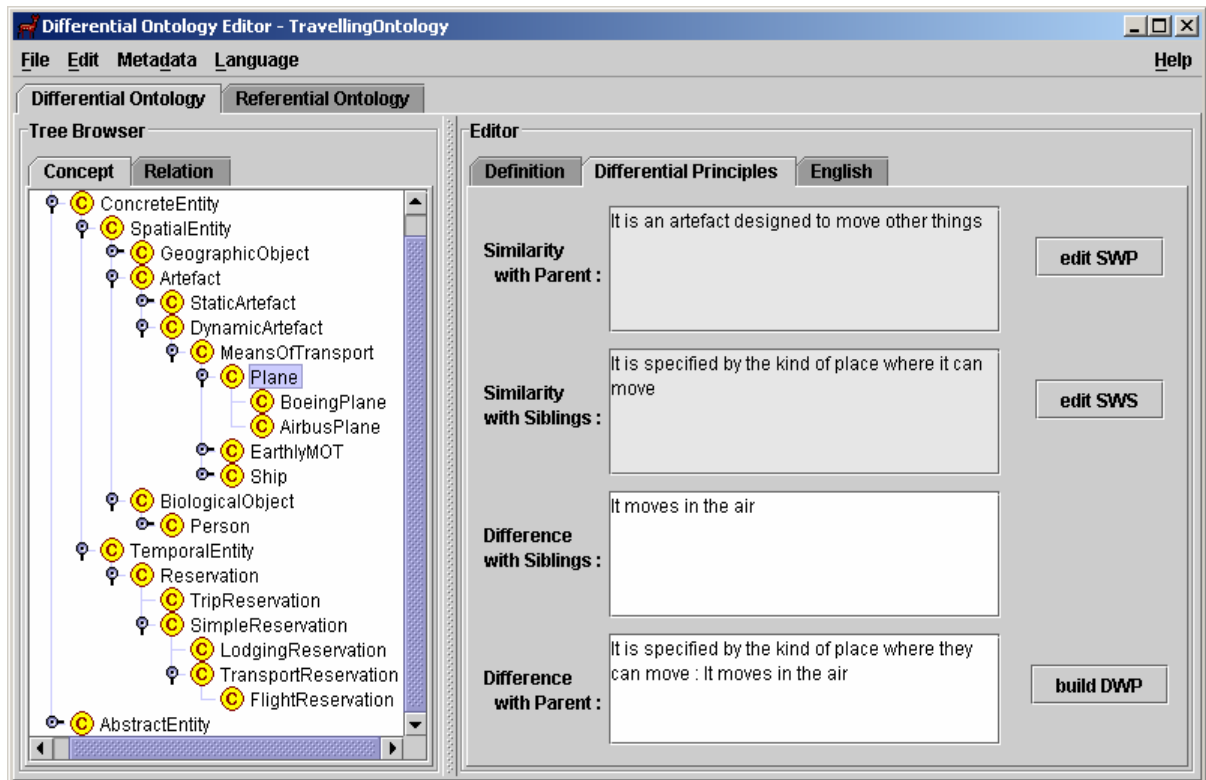
³⁶ <http://homepages.cwi.nl/~troncy/DOE/>

chaque terme sa définition encyclopédique, ses éventuels synonymes et ses principes différentiels³⁷ et cela en plusieurs langues. Cet outil n'a pas pour ambition de concurrencer les grands environnements existants, mais plutôt d'implémenter la méthodologie de structuration différentielle [de Bruno Bachimont] en permettant à l'ontologue de bien faire la distinction entre ontologie différentielle et référentielle. L'outil assiste également la saisie des principes différentiels issus de la méthodologie en automatisant partiellement cette tâche. Le modèle de représentation de l'ontologie est finalement proche de celui du langage RDFS, à ceci près qu'il autorise la modélisation de relations binaires. Dans le domaine formel, l'éditeur est capable de faire quelques inférences en vérifiant la consistance de l'ontologie (propagation de l'arité le long de la hiérarchie des relations et héritages des domaines par exemple).

Cet éditeur permet également d'ajouter des individus à l'ontologie. Finalement, le passage à une ontologie computationnelle s'effectue par un export de l'ontologie formelle dans un certain nombre de langages opérationnels tels que RDFSn DAML+OIL, OWL et CGXML. Cette traduction d'effectue grâce à des feuilles de style XSLT appliquées au format de sauvegarde XML de l'éditeur. De la même façon, DOE peut importer des ontologies modélisées dans d'autres outils grâce à des feuilles XSLT dédiées. Pour être tout à fait complet, il faut noter que l'éditeur permet aussi de sauvegarder un certain nombre de métadonnées concernant l'ontologie elle-même. Il suit ici les propositions du Dublin Core.

L'interopérabilité de DOE avec d'autres outils, tel que Protégé est nécessaire dans la mesure où il ne met pas en œuvre les fonctionnalités d'expressivité formelle prises en charge par les autres environnements. Cet outil ne constitue pas, pour l'instant, un environnement de développement d'ontologies complet mais se situe en amont d'autres éditeurs implémentant, par exemple, les logiques de description. Il met l'accent sur la structuration de taxinomies.

³⁷ L'utilisation des principes différentiels est extraite de la méthodologie de Bruno Bachimont exposée dans la deuxième partie de ce chapitre.



11. Copie d'écran de l'interface principale de l'éditeur d'ontologies DOE

Ainsi, à travers ce rapide tour d'horizon des principaux éditeurs d'ontologies, qui montre bien d'ailleurs la prédominance de OWL comme langage de représentation des connaissances, nous venons de traiter le dernier point indispensable à la compréhension du chapitre III consacré à la réalisation de la mission.

CHAPITRE III : VERS LA TRANSFORMATION DU MICRO-THESAURUS PACTOLS-SUJETS EN ONTOLOGIE

Nous savons maintenant ce qu'est une ontologie en philosophie et en ingénierie des connaissances, comment ce système de représentation des connaissances se situe par rapport aux langages documentaires et notamment par rapport à celui qui lui est le plus proche : le thésaurus. Enfin, nous venons de voir comment construire une ontologie grâce à quelques repères théoriques (composants, exemple de méthode de conception et réalisation d'ontologies, langage de représentation) et pratiques (exposé non exhaustif des éditeurs d'ontologies). C'est donc forts de ce savoir orienté vers la réalisation de la mission, que nous allons, dans le chapitre qui suit, d'une part prendre connaissance en détail de ce qui a été modifié, transformé, à savoir la partie Sujets du thésaurus PACTOLS et de quelle manière cela a été fait, selon quelle méthodologie, selon quels moyens informatiques ; et d'autre part, nous allons pouvoir analyser ce résultat et envisager la suite du projet en fonction de l'évolution des données contextuelles.

I. Description du micro-thésaurus Sujets des PACTOLS

Le micro-thésaurus Sujets, qui correspond au S des PACTOLS, a été choisi par Madame Lequeux pour servir de base à la construction d'une ontologie dans le cadre de l'appel à projets innovants lancé par le TGE ADONIS³⁸, visant à soutenir et promouvoir les outils et systèmes innovants de valorisation et de diffusion des données de la recherche et permettre leur mutualisation et leur mise à disposition de l'ensemble de la communauté des SHS.

Avant de voir comment j'ai procédé pour répondre concrètement à cet appel à projets, ce qui était le but de ma mission de stage, intéressons-nous d'abord en détail à la partie actuelle du thésaurus PACTOLS-Sujets ; car comprendre le fonctionnement, connaître sa composition, et

³⁸ Pour plus de détail sur cet appel à projets innovants lancé par le TGE Adonis, voir en introduction page 14

son utilité ont été les premières démarches qu'il m'a fallu accomplir tout en effectuant des lectures sur l'ingénierie ontologique.

1) Structure et thématique de PACTOLS-Sujets

Le PACTOLS-Sujets, comme les autres micro-thésaurus des PACTOLS, est construit selon les normes internationales de l'Organisation internationale de normalisation (ISO) 2788 [ISO 2788:1986] pour les thésaurus monolingues et 5964 [ISO 5964: 1985] pour les multilingues. Le choix de faire appel aux normes internationales plutôt qu'à celles éditées par l'Association française de normalisation (AFNOR) a été fait par le réseau FRANTIQ car les normes nationales ont été jugées trop limitées par rapport aux normes internationales à l'époque de sa création, vers 1987, notamment en ce qui concerne les normes concernant les thésaurus multilingues. Ces normes sont anciennes, elles datent de 1986 et 1985. , mais sont toujours en vigueur de nos jours.

Du point de vue de sa structure, le thésaurus est un vocabulaire contrôlé, dynamique et évolutif de termes ayant entre eux des relations sémantiques – générique/spécifique, équivalence, associatives – et qui s'applique à un domaine particulier de la connaissance : l'archéologie depuis la Préhistoire jusqu'à l'Ère Industrielle et les sciences de l'Antiquité jusqu'à l'An Mil. Plus concrètement, ce micro-thésaurus se divise en 26 hyperonymes – terme dont le sens inclut le sens d'autres termes qui sont ses hyponymes – regroupant les domaines d'étude sur l'Antiquité. Ces hyperonymes sont les suivants, classés par ordre alphabétique : architecture, art, céramologie, défense, droit, faune, flore, géographie, histoire-civilisation, iconographie, loisirs, matériaux, méthodologie, mort, paléographie, philologie, philosophie, politique, religion, santé, savoir, site archéologique, société, vie économique, vie financière et vie quotidienne. La partie Céramologie appartient d'après la logique du thésaurus à la partie Art, mais elle a été remontée au premier niveau car trop de références la concernaient. De même, la Mort est une partie à part entière car les fouilles de sépultures sont une grande activité des archéologues et il y a donc besoin de beaucoup de termes pour la qualifier.

En outre, pour plus de clarté et éviter toute ambiguïté, voici un tableau³⁹ mettant en relation chacun des hyperonymes avec ce qu'ils représentent, leur composition :

Architecture	Art de construire les bâtiments. Toutes installations (objets, dispositifs, bâtiments, etc.) servant à un usage déterminé.
Art	Activité dont la finalité est de caractère esthétique, désintéressée, non utilitaire.
Céramologie	Étude de la céramique. Regroupement de tout ce qui concerne la céramique : art, style, forme, technique.
Défense	Tout ce qui concerne la guerre, l'armée et l'armement.
Droit	Ensemble des règles qui régissent les rapports entre des membres d'une même société.
Faune	Ensemble des animaux d'une région, d'un milieu ou d'une époque déterminés.
Flore	Espèces végétales ayant une fonction, une utilité, particulière pour les hommes.
Géographie	Science qui étudie et décrit la Terre à sa surface, en tant qu'habitat de l'homme et de tous les organismes vivants (comment les sociétés jouent de l'espace pour s'organiser et se structurer).
Histoire-civilisation	Concerne tout ce qui traite en général d'une civilisation.
Iconographie	Tout ce qui concerne l'étude descriptive des différentes représentations figurées.
Loisirs	Tout ce qui concerne la distraction, ce qui procure du plaisir.
Matériaux	Toutes les matières premières utiles à la fabrication d'un objet, d'un bâtiment, etc.
Méthodologie	Tous les modes d'approches de l'Antiquité ainsi que l'organisation de la recherche.
Mort	Tout ce qui concerne la mort, la cessation complète et définitive de la vie d'un être humain, d'un animal. (de la façon de mourir aux décors funéraires)
Paléographie	Science des anciennes écritures et de leur support. Tout ce qui concerne la manière d'écrire (support et contenu)
Philologie	Étude d'une langue d'après les documents écrits qui nous la font connaître. Épigraphie [Tout ce qui permet l'étude d'une langue ancienne d'après des documents écrits.]
Philosophie	Tout ce qui concerne la réflexion critique sur les problèmes de l'action et de la connaissance humaine.
Politique	Tout ce qui est relatif au gouvernement de l'État, à la cité et à la vie administrative.
Religion	Toutes les constructions idéologiques qui deviennent un objet de croyance, dont certaines entraînent obéissance et respect.
Santé	Tout ce qui concerne l'hygiène et la médecine.
Savoir	Tout ce qui concerne le savoir et les techniques tels qu'ils étaient sous l'Antiquité. (Connaissances plus ou moins systématisées)

³⁹ Ce tableau a été réalisé dans le cadre d'une présentation de mon travail en cours à mon maître de stage, Monsieur Menon.

Site archéologique	Il s'agit de structures archéologiques non bâties et dont la destination est ambiguë ou multiple.
Société	Tout ce qui concerne l'homme en tant qu'individu et l'homme appartenant à un groupe. (sociologie)
Vie économique	Ensemble des actions et éléments de production et de consommation d'une communauté ou d'une société.
Vie financière	Ensemble des actions et éléments relatifs aux ressources pécuniaires, à l'argent.
Vie quotidienne	Tout ce qui traite avant tout de la culture matérielle d'une civilisation.

12. Tableau récapitulatif des 26 hyperonymes et de leur domaine d'action.




Le PACTOLS-Sujets correspond à un vocabulaire relativement stable depuis l'intégration de termes spécifiques pour la Préhistoire – la communication avec des chercheurs en Préhistoire est facilitée par la présence d'équipes spécialisées dans le domaine au même étage de la MAE que le service Documentation/FRANTIQ – même si actuellement, quelques modifications sont effectuées en raison de la réalisation d'une table de concordance des termes dans le projet DAPHNE⁴⁰. On peut également supposer que l'entrée dans le CCI de l'Institut National de Recherches Archéologiques Préventives (INRAP) va faire augmenter les descripteurs du côté des structures modernes et industrielles.

2) Les descripteurs dans PACTOLS-Sujets et la gestion des termes

Les termes du micro-thésaurus Sujets, ou descripteurs, sont des mots ou des locutions qui se réfèrent à des concepts. Il s'agit de termes autorisés. On les nomme également souvent mots-clés dans le cadre d'une indexation de documents ou bien dans le cadre d'une recherche documentaire.

Ils sont en français, sauf en cas d'ambiguïté. Par exemple, comme le montre les deux copies d'écran suivantes, le latin est utilisé pour distinguer les basiliques civiles et religieuses. Il peut aussi être utilisé dans le cas d'emprunt dans une autre langue comme *falcata*.

⁴⁰ Pour un rappel sur ce que signifie le projet DAPHNE, se référer à sa présentation dans l'introduction générale de ce mémoire ou visiter le site Internet issu du projet : <http://www.daphne.cnrs.fr>

Open Theso Web De l'aide ?   

Navigation Recherche

Terme: basilique

Options: Tous français

Cliquez sur un des boutons pour lancer la recherche

Hierarchique Alphabétique Permutée

architecture

- architecture (conception)
 - architecture religieuse
 - édifice religieux
 - lieu de culte
 - église
 - basilique**

Terme(s) générique(s)

- église

↓

basilique

Eglise chrétienne bâtie sur le plan de l'édifice civil romain (Lar). Pour les édifices civils, indexer basilica

NBR de notices: 392
Microthesaurus: Sujets

Crée le : 08/02/07 Modifié le : 08/02/07

Terme(s) Associé(s)

Employé pour

- basilique chrétienne

Autre(s) langue(s)

- basilica (religious) (eng)
- Basilika (ger)
- basilica cristiana (ita)

Communiquer une illustration pour ce terme avec les mentions de copyright

Navigation Recherche

Terme: basilica

Options: Sujets français

Cliquez sur un des boutons pour lancer la recherche

Hierarchique Alphabétique Permutée

architecture

- architecture (conception)
 - architecture publique
 - édifice public
 - basilica**

Terme(s) générique(s)

- édifice public

↓

basilica

Edifice public romain, allongé et divisé en trois nefs, qui servait de tribunal, puis de lieu de promenade et de commerce (LexHCR)

NBR de notices: 45
Microthesaurus: Sujets

Crée le : 08/02/07 Modifié le : 08/02/07


Terme(s) Associé(s)

Employé pour

- basilique romaine

Autre(s) langue(s)

- basilica (judicial) (eng)
- basilica romana (ita)



13. Copies d'écran montrant la distinction entre les basiliques civiles et religieuses par l'utilisation du latin

Le descripteur « basilique » fait référence aux basiliques religieuses chrétiennes.

Le descripteur « basilica » désigne l'édifice civil

Les descripteurs sont dans un langage naturel, ce qui signifie d'après le *vocabulaire de la doc* de l'ADBS [BOULOGNE : 2004] (l'Association des professionnels de l'information et de la

documentation) qu'ils « sont écrits dans une écriture relative à un système d'expression parlée ou écrite, dont les éléments et les structures sont communs à un groupe social, et dont les règles résultent de l'usage sans être nécessairement prescrites ni explicites ». Ainsi, on trouve le descripteur Technique de construction plutôt que Construction (technique de) car il s'agit de l'expression que l'on utilise dans le langage courant. Il n'y a ni abréviations, ni acronymes, sauf de rares exceptions qui concernent notamment des techniques comme SIG (Système d'information géographique) qui est le nom d'un logiciel ou encore AMS qui signifie *Accelerator Mass Spectroscopy*, soit datation par spectromètre de masse avec accélérateur.

Enfin, ils sont au singulier, sauf pour des sens particuliers tels que « bronze » pour le matériau et « bronzes » pour les statues, « sport » pour le type de loisir et « sports » comme regroupement des différentes disciplines sportives. Ou encore, ils sont au pluriel si les règles de la langue française l'exigent : thermes, latrines.

Afin d'aider à la compréhension et désambiguïser l'utilisation de certains termes, il existe des notes d'application (NA), *scope notes* en anglais et des notes historiques. Quand j'ai commencé mon stage, les PACTOLS comptaient 3308 NA⁴¹ car seuls les termes pouvant porter à confusion étaient documentés. Ce nombre a depuis augmenté afin de transformer le thésaurus PACTOLS-Sujets en ontologie. Ainsi, une de mes premières réalisations a été de documenter les définitions pour certaines parties telles que la Mort, la Faune ou encore la rubrique Méthodologie. En outre, au-delà de l'objectif ontologique, ces définitions rendent le thésaurus accessible à un plus large public. Car si pour un archéologue la signification du terme « feuille de laurier » est évidente, ce n'est pas toujours le cas pour un public non averti.

Enfin, l'illustration de certains sujets est également une possibilité d'enrichissement du thésaurus et cette option a commencé à être appliquée. Néanmoins, la question des droits d'auteurs sur les photos reste un souci majeur pour leur diffusion. Quant aux notes historiques, elles correspondent à la vie du concept dans le thésaurus. Elles renseignent sur la création, la transformation, le développement ou encore sur le regroupement de termes. Par manque de temps, ces notes historiques sont très peu renseignées.

Le micro-thésaurus PACTOLS-Sujets se compose d'environ 4500 descripteurs. Afin de gérer tous les termes y figurant, il existe donc certaines règles, méthodes d'écriture permettant de préciser le sens, la signification des termes et bien évidemment des relations.

⁴¹ La liste des ouvrages ayant servi à l'élaboration de ces notes d'application m'a été communiquée par Madame Lequeux et figure en annexe III, *Liste des ouvrages et sites Internet de référence, utilisés pour le création des PACTOLS et les définitions des Notes d'Application, élaborée par Madame Lequeux*.

L'antonymie – les opposés pour indiquer les rapports entre les deux notions – est signalée par un « slash » entre les deux entités : masculin/féminin ou vice/vertu.

L'homonymie par homographie est désambiguïsée par l'utilisation de parenthèses contenant des termes permettant de renseigner le sens du concept. Par exemple, elles sont utilisées pour Musée : « musée (lieu de conservation) » et « Musée (poète) » ; ou encore pour Mortier : « mortier (construction) » et « mortier (récipient) ». L'homonymie pose souvent problème avec les moteurs de recherche, la transformation du thésaurus en ontologie devrait permettre de le régler.

En ce qui concerne le choix des relations entre les différents concepts, elles ont été réalisées d'après les relations essentielles définies dans les normes ISO 2788 [ISO 2788:1986] pour les thésaurus monolingues, et ISO 5964 [ISO 5964: 1985] pour les thésaurus multilingues. On distingue trois types principaux de relations présents dans le PACTOLS-Sujets : les relations hiérarchiques, les relations associatives et les relations d'équivalence.

L'organisation du micro-thésaurus PACTOLS-Sujets est basée sur des relations de nature hiérarchique. Ces relations servent à exprimer les rapports de supériorité et de subordination entre les concepts exprimés par les descripteurs. Elles constituent la structure essentielle du thésaurus autour de laquelle s'organisent l'indexation et la recherche documentaire. Théoriquement, ces relations dites hiérarchiques se subdivisent en relations génériques et relations partitives. La relation générique permet d'organiser autour d'un descripteur la classe dont il fait partie et se fait en termes génériques et termes spécifiques :

ARME OFFENSIVE Terme générique ARME (*Broader Term*)




ARME OFFENSIVE Terme spécifique ARBALETE (*Narrower Term*)

Elle se note de la manière suivante :

TG Terme générique (BT)

TS Terme spécifique (NT)

C'est également une relation qui peut parfois se qualifier de nature parentale. On parle de relation entre père et fils. Les relations génériques peuvent être à caractère soit mono-hiérarchique – un terme ne peut avoir qu'un seul père – soit poly-hiérarchique – un terme peut avoir plusieurs pères. Dans le thésaurus qui nous intéresse, les relations peuvent être poly-hiérarchiques comme le montre cette copie d'écran :

Open Theso Web De l'aide ?   

Navigation Recherche

Terme : objet religieux

Options : Sujets français

Cliquez sur un des boutons pour lancer la recherche

Hierarchique Alphabétique Permutée

vie quotidienne
 meuble
 objet
 objet religieux
 religion
 objet religieux

Terme(s) générique(s)
 • objet
 • religion
 ↓
objet religieux
 NBR de notices : 1574
 Microthesaurus: Sujets
 Créé le : 08/02/07 Modifié le : 08/02/07

Terme(s) Associé(s)
 Employé pour
 • objet de culte
 • objet magique
 • objet sacré
 Autre(s) langue(s)
 • religious object (eng)
 • Kultgerät (ger)
 • oggetto religioso (ita)

14. Copie d'écran montrant la poly-hiérarchie du micro-thésaurus PACTLOS-Sujets – objet religieux

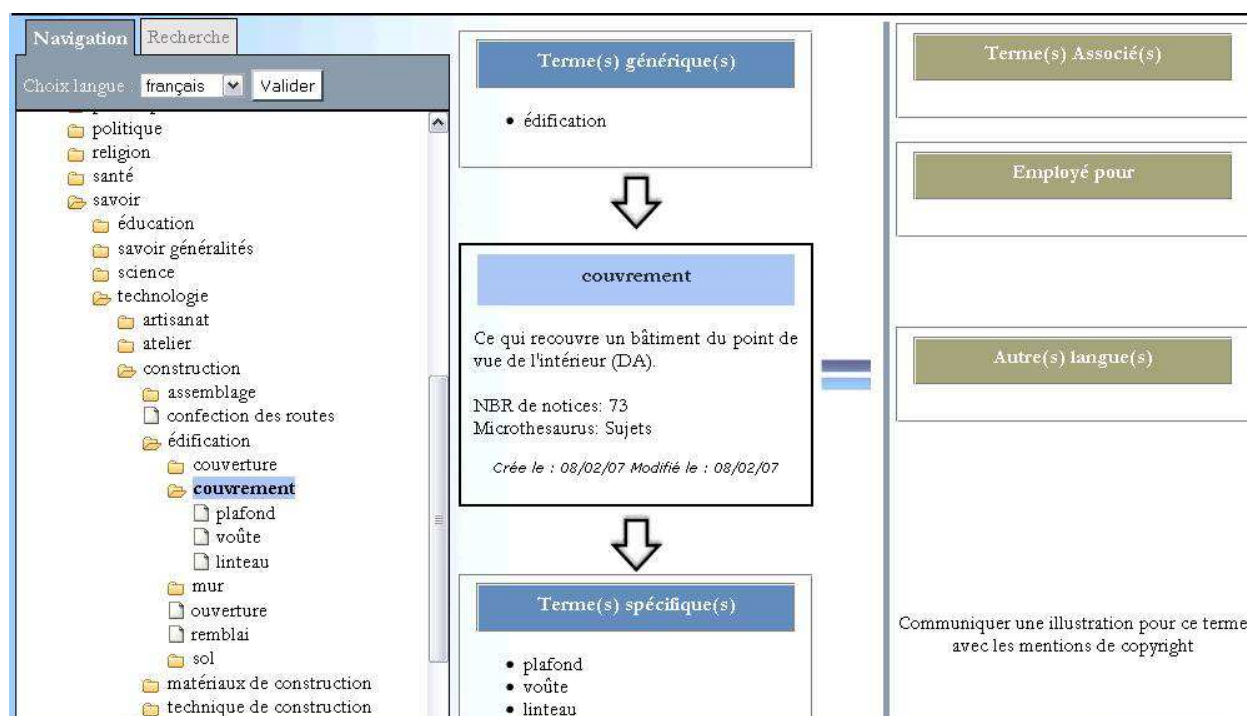
On remarque sur cette copie d'écran que le concept *Objet religieux* fait référence à deux termes génériques : *Religion* et *Objet*.

Le deuxième type de relation hiérarchiques, la relation partitive ou relation tout/partie est utilisée pour découper un concept en ses constituants, c'est-à-dire l'entité en ses parties. Dans le micro-thésaurus ce type de relation est confondu dans la relation générique. En effet, dans PACTOLS-Sujets on peut trouver sous un terme générique :

- **les contenants d'un contenu** : les Sports englobent l'acrobatie, l'athlétisme, la boxe, la lutte, etc.
- **les composants d'un générique** : Motif végétal comprend les hyponymes Lotus, Palmette et Rosace
- **les facettes d'une activité** : le terme générique Agriculture est subdivisé en Saison agricole, Technique agricole, Outil agricole, etc.

Enfin, le terme générique peut être un terme de regroupement, comme par exemple Bâtiment à fonctions multiples.

La copie d'écran suivante montre la hiérarchie présente dans le micro-thésaurus PACTOLS-Sujets.



15. Copie d'écran illustrant la hiérarchie des concepts – Couvrement

Néanmoins, il est intéressant de savoir que si les deux types de relations hiérarchiques sont utilisés dans un même thésaurus, autre que les PACTOLS qui ne possèdent pas cette particularité, les relations seront notées de la manière suivante afin de les différencier : **TG P** Terme générique partitif / **TS P** Terme spécifique partitif et **TG G** Terme générique en relation générique / **TS G** Terme spécifique en relation générique.

Les relations associatives servent à indiquer les analogies pouvant exister entre deux descripteurs, entre deux termes associés. Néanmoins, il convient de préciser que les liens unissant les deux descripteurs sont variables et relèvent, dans un thésaurus, du bon sens de l'utilisateur et c'est notamment sur ce point que l'ontologie peut s'avérer plus efficace, plus pertinente qu'un thésaurus car elle peut qualifier le type de relation sous-entendue.

La relation associative se note de la manière suivante :

TA Terme associé (RT *Related term*)

Dans le micro-thésaurus PACTOLS-Sujets, ce type de relation est toujours réciproque. Elle permet de limiter la poly-hiérarchie tout en établissant des liens entre les notions. Pour ne citer que deux exemples, la Voie – dans Organisation de l'espace – est liée avec Route commerciale – dans Commerce ; et Artisanat alimentaire – dans Technologie est lié à Alimentation – dans Vie quotidienne – et à Produit agricole dans Vie économique.

En outre, la présence de cette relation permet de suggérer une autre interrogation ou une autre indexation.

Enfin, les relations d'équivalence constituent la troisième grande catégorie de relations présentes dans PACTOLS-Sujets. Elles renvoient généralement des synonymes ou quasi-synonymes vers le descripteur. Ce sont les relations : EMPLOYER (*use*) et EMPLOYE POUR (*use for*) et elles se notent ainsi : EM employé et EP employer pour.

Ces relations dans le micro-thésaurus regroupent :

- **Les synonymes réels** : *limes* employer Frontière.
- **Des graphies différentes** : Ramsès 2 et non Ramsès II
- **Les quasi-synonymes** : *Insula* = pâté de maison + immeuble (*Insula* n'est pas un synonyme, il est remplacé par l'un et/ou l'autre), Inscription égyptienne = épigraphie + égyptien (langue) (Inscription égyptienne n'est pas un synonyme, il est remplacé par l'un et l'autre).
- **Quelques traductions** si elles ne sont pas traitées dans la partie multilingue comme le latin ou le catalan
- **Des termes obsolètes** comme le changement de nom. Par exemple, dans les Logiciel documentaires sont regroupés les anciens logiciels très utilisés dans le domaine archéologique (Sofia, Sox, Texto, Mandoline, etc.)

3) Représentation du thésaurus PACTOLS et son utilité

Nous savons maintenant comment est structuré le micro-thésaurus PACTOLS-Sujets, nous connaissons le type de relations qui le compose ainsi que sa thématique. Nous avons donc observé le thésaurus du côté « administrateur » en quelque sorte, voyons maintenant ce même thésaurus du côté « utilisateur » en se demandant qui l'utilise, à quelles fins et de quelles manières. Pour se faire, le développement se fera en grande partie sur l'intégralité des thésaurus PACTOLS car ils sont, dans un premier temps tout au moins, indissociables dans leur représentation et sont utilisés de la même manière.

Les PACTOLS sont utilisés par deux catégories de personnes, d'une part par des gestionnaires dont Madame Lequeux est le principal et d'autre part, par des utilisateurs qui se divisent eux même en deux catégories : les indexeurs et personnels des bibliothèques, et les étudiants et chercheurs – « le public ».

Madame Lequeux, en tant que gestionnaire principal des PACTOLS, est chargée de faire évoluer les thésaurus. C'est elle qui fait les ajouts, suppressions et modifications de termes et relations. Ces manipulations surviennent généralement après des remarques des utilisateurs indexeurs et bibliothécaires ou selon les nouveautés dans la thématique traitée par les PACTOLS. Pendant mon stage, j'ai également participé à l'évolution des PACTOLS-Sujets, en effectuant les manipulations de base : ajouts, suppressions et modifications ; mais aussi et surtout en enrichissant les notes d'application, les NA. Ainsi, j'ai ajouté des définitions manquantes dans certaines rubriques : Faune, Méthodologie, Mort, Défense et dans d'autres selon les besoins de l'ontologie. Dans le souci d'enrichir de manière pertinente la partie Sujets des PACTOLS car je suis à l'origine historienne et non archéologue, j'ai procédé de la manière suivante pour définir les termes : dans un premier temps j'ai réuni les outils documentaires⁴² nécessaires pour définir les termes selon le domaine choisi, j'ai recherché les définitions et je les ai retranscrites dans un documents Word afin de les soumettre à l'avis d'un expert – Madame Lequeux – après ses éventuelles corrections et validations, j'ai finalement entré les définitions dans le logiciel de gestion du thésaurus. Le fait de définir plus de 450 termes m'a permis de prendre connaissance à la fois du thésaurus et de son logiciel, mais également du domaine de connaissances concerné : l'archéologie et les sciences de l'Antiquité. Je me suis, en d'autres termes, approprié l'objet de l'étude.

La seconde catégorie de personnes utilisant les PACTOLS sont les personnes dites « utilisateurs ». Il y a d'une part le personnel des bibliothèques et d'autre part les chercheurs, étudiants et autres personnes utilisant le thésaurus pour effectuer des recherches documentaires.

Le personnel des bibliothèques se sert des PACTOLS pour indexer les documents présents dans leur fonds documentaire. L'indexation est une opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans un document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse. Cette transcription se fait grâce aux PACTOLS.

L'indexation doit répondre à plusieurs questions qui trouvent leur réponse dans chacun des micro-thésaurus des PACTOLS :

⁴² La liste des ouvrages et des sites Internet que j'ai utilisés figure en annexe IV, *Liste complémentaire des ouvrages et sites Internet de référence, utilisés pour compléter les Notes d'Application et les définitions encyclopédiques de l'ontologie.*

- **Quoi ?** *Sujets*
- **Où ?** *Lieux et/ou Toponymes*
- **Quand ?** *Chronologie et/ou Cultures*
- **Qui ?** *Anthroponymes ou Peuples*
- **Comment ?** *Sujets ou Œuvres*

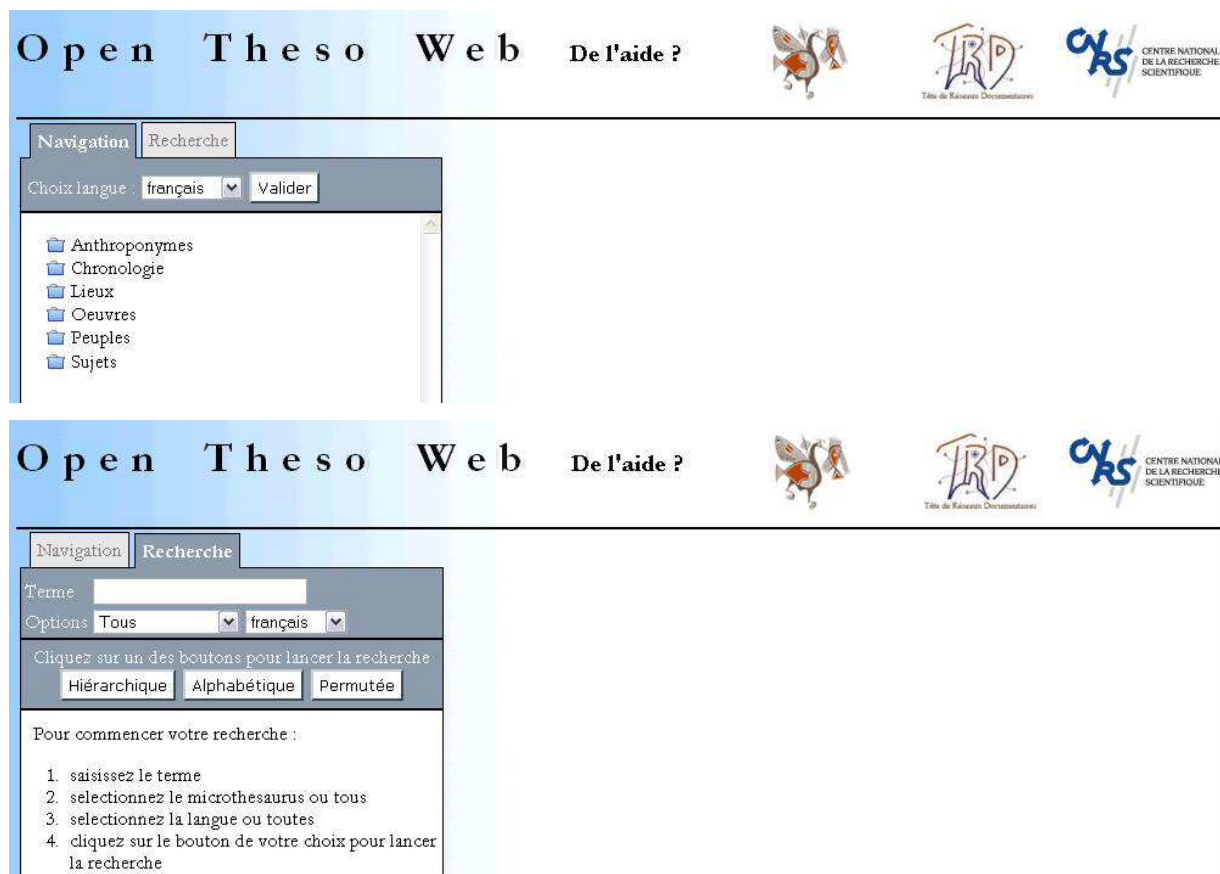
L'indexation a pour objectif de permettre une recherche efficace des informations contenues dans un fonds documentaire ; recherche qui est effectuée par le deuxième type d'utilisateurs du thésaurus. Ces derniers utilisent le processus d'indexation non pas en entrée, comme les indexeurs, mais en sortie. Ils se servent des PACTOLS et de leur langage contrôlé pour rechercher des documents à l'aide notamment d'opérateurs booléens et de troncatures.

Ainsi, les PACTOLS sont connus par un large éventail de catégorie de personnes et pour des usages différents. Pour permettre cela, depuis 2006-2007, les thésaurus PACTOLS ne sont plus gérés par le logiciel THESO-TEXTO, mais par une application libre en PHP créée par l'informaticien du réseau FRANTIQ, Miled Rousset, avec l'aide de stagiaires. Cette application est le fruit du projet OpenTheso : gestionnaire de thésaurus multilingue, commencé en 2005 d'après un sujet de stage pour un étudiant en Master IID-SIA à Lyon 3. La première version du gestionnaire de thésaurus a été réalisée durant ce stage mais ne gère pas le multilinguisme faute de temps – quatre mois. Une deuxième phase a commencé en 2006 et a été entièrement réalisée par l'informaticien de FRANTIQ, qui a eu pour but de pallier l'absence de multilinguisme. Aujourd'hui, le logiciel OpenTheso est stable même si des améliorations sont toujours en cours et est exploitable, et exploité, par le réseau FRANTIQ. Ce dernier, d'ailleurs, teste en même temps qu'il l'utilise cette nouvelle application pour gérer les PACTOLS. Ce logiciel libre, est un vrai logiciel de gestion et de recherche documentaire (LGRD). Il est destiné avant tout aux gestionnaires et autorise les ajouts, corrections, suppressions et autres traductions ainsi que l'import d'images pouvant illustrer les concepts.

Cet OpenTheso, comme annoncé dans l'introduction, est maintenant couplé avec l'OpenThesoWeb, l'Opac utilisateur. Ainsi, les PACTOLS sont accessibles librement et gratuitement sur Internet⁴³ et peuvent être consultés aussi bien par les bibliothécaires pour réaliser leurs indexations que par les personnes désirant consulter simplement le thésaurus pour des fins diverses. Comme le montrent les copies d'écran qui suivent, à travers cet Opac, la

⁴³ <http://193.48.137.41:8080/OpenThesoWeb/>

consultation peut s'effectuer de deux manières : soit par une navigation au cœur des micro-thésaurus, soit par une recherche avec le choix entre plusieurs modes d'affichage que sont les listes hiérarchiques, les listes alphabétiques et les listes permutées.



16. Copies d'écran montrant les différents modes de consultation des thésaurus PACTOLS

Enfin, les PACTOLS, par choix, sont également autonomes de manière à pouvoir être utilisés dans différentes applications ; ce qui est le cas pour l'indexation et la recherche dans la base Archives de la MAE, et pour le produit « AdlFI » qui indexe les notices de fouilles. En outre, les listes alphabétiques et/ou hiérarchiques des micro-thésaurus, sur documents Word ou txt sont téléchargeables depuis le site de FRANTIQU : <http://frantiq.mom.fr>.

II. Choix méthodologiques

Comme nous l'avons vu tout au long de ce qui précède, la réalisation d'une ontologie n'est pas une tâche aisée. Dans un projet idéal, cela nécessiterait du personnel qualifié, du temps

et des ressources financières, ne serait-ce que pour rémunérer le personnel mobilisé. Mais mon stage, et c'est un des points qui a contribué à lui donner tout son intérêt pédagogique et formateur, ne réunissait pas ses trois critères et cela a influencé les choix méthodologiques que j'ai fait pour concevoir l'ontologie d'après le micro-thésaurus PACTOLS-Sujets.

1) Critères de sélection

Les trois critères évoqués à l'instant – le temps, les hommes et l'argent – s'ils n'étaient pas réunis – ont été néanmoins les critères primordiaux dans mes décisions. En effet, il m'a fallu, après mes premières recherches sur l'ingénierie ontologique et l'appropriation du thésaurus, faire des choix méthodologiques afin de pouvoir commencer la conception de l'ontologie.

De manière abrupte mais très concrète, le temps est nécessaire car la tâche est longue à réaliser d'une part intellectuellement, et d'autre part, manuellement, enfin on parle plutôt maintenant d'informatique mais l'idée reste la même : il faut du temps pour apprendre à utiliser un logiciel et ensuite pour le « remplir » ; les hommes, c'est-à-dire le personnel mobilisé, sont essentiels : idéalement, il faudrait une équipe de trois personnes/compétences au minimum – une personne spécialisée en ingénierie ontologique pour guider, renseigner, qui, en d'autres termes servirait de référent, une autre pouvant être considérée comme relativement novice dans le domaine ontologique mais experte en gestion de l'information, qui exécuterait certes les tâches de base mais qui apporterait également un regard critique sur la structure de l'ontologie et une dernière personne qui serait spécialisée dans le domaine de connaissances concerné par l'ontologie – enfin, l'argent selon sa quantité peut être utile et utilisé pour différentes choses : s'il est à profusion il peut servir à s'offrir les services d'une entreprise spécialisée dans l'ingénierie ontologique, s'il est compté mais présent, il pourra éventuellement servir pour l'achat d'un logiciel spécialisé, et enfin, si les ressources financières sont au minimum, il servira à la rémunération du personnel mobilisé pour l'exécution de la tâche.

Si l'on considère les trois critères dans leur ensemble et qu'on les applique au contexte de ma mission cela donne la représentation suivante :

- **le personnel** prévu pour le projet était « idéal » : un référent en ingénierie ontologique avec qui a été envisagé à l'origine du projet Adonis un partenariat : Madame Nathalie

Aussenac-Gilles représentant l'IRIT – IC3 UMR 5505 de Toulouse 3⁴⁴, une personne spécialisée dans la gestion de l'information : moi-même, et un expert en archéologie et dans les sciences de l'Antiquité : Madame Lequeux. On peut même ajouter la présence d'un service informatique au cœur de la Maison René-Ginouvès et celle, à distance, d'un informaticien : Miled Rousset. Néanmoins, dans la pratique, je n'ai pas pu bénéficier de tous les soutiens prévus au départ.

- **Le critère temps** peut se concevoir de deux manières, d'une part en considérant la durée du stage : trois mois, et d'autre part, d'après la disponibilité des membres, partenaires, susceptibles de participer à la réalisation du projet : disponibilité variable, allant d'une présence et d'un soutien sans faille de la part de Madame Lequeux, à une disponibilité des plus restreintes, voire quasi inexistante de la part de Madame Aussenac-Gilles très occupée dans d'autres projets au moment où j'ai effectué mon stage.
- En ce qui concerne **le financement** il a été réservé entièrement à la rémunération du personnel car il était très faible.

En outre, à ces trois critères essentiels, un dernier est venu s'ajouter : apporter de la plus-value au micro-thésaurus PACTOLS-Sujets. En effet, pour que le projet et le stage soient constructifs, il ne suffisait pas, très schématiquement, de transposer tel quel le thésaurus dans un langage permettant la manipulation automatique des données par une machine et leur visibilité sur le web, mais d'y apporter des améliorations. En d'autres termes, il s'agissait de transformer le thésaurus en ontologie, donc de s'appuyer sur un formalisme fort, et non de se contenter de transférer la structure du thésaurus dans un langage permettant l'accessibilité du vocabulaire contrôlé dans l'environnement du web sémantique.

C'est donc en fonction de ces quatre critères que j'ai été amenée à choisir une méthode pour concevoir l'ontologie. Mais avant de parvenir à faire ce choix qui va être présenté en détail dans la suite du mémoire, il ne faut pas négliger les solutions qui ont été laissées de côté non pas parce qu'elles ne menaient pas à la réalisation de la mission, mais parce qu'elles ne correspondaient pas aux critères imposés.

⁴⁴ L'IRIT, Institut de Recherche en Informatique de Toulouse est une Unité Mixte de Recherche, UMR 5505, commune au CNRS, à l'Institut National Polytechnique de Toulouse (INPT), à l'Université Paul Sabatier (UPS) et à l'Université des Sciences Sociales de Toulouse 1 (UT1).

L'IC3 est une des équipes de l'UMR 5505. IC3 signifie Ingénierie de la Connaissance, de la Cognition et de la Coopération. Cette équipe, menée par Nathalie Aussenac-Gilles a comme axe central l'ingénierie des modèles de connaissances et/pour des systèmes coopératifs.

2) Solutions écartées

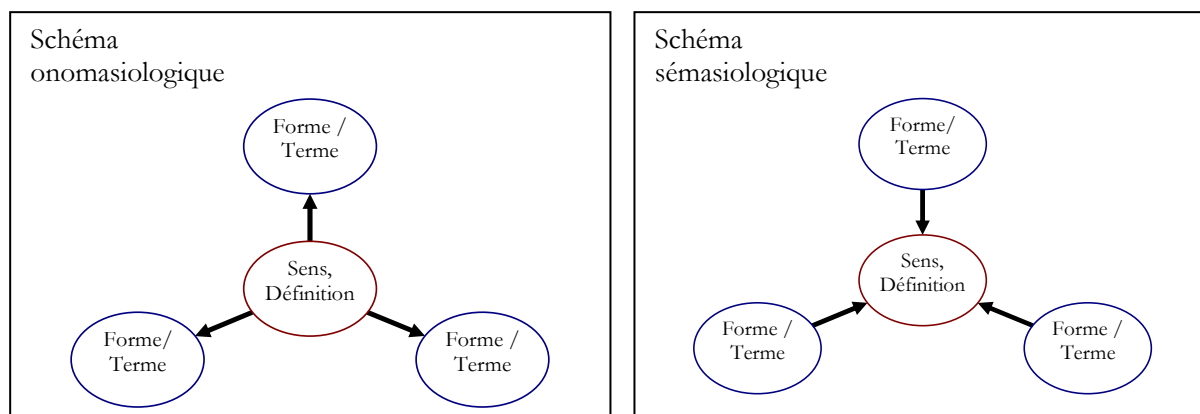
Afin de présenter de manière structurée les solutions qui n'ont pas été retenues pour réaliser la transformation du thésaurus en ontologie, elles suivront l'ordre de ma propre progression intellectuelle. Le classement effectué sera donc le reflet de mes recherches, de mes rencontres, de mes espérances, de mes certitudes mais aussi de mes tâtonnements et doutes. Pour chacune des grandes étapes de mon raisonnement qui correspondent aux quatre solutions écartées, une description générale de la méthodologie envisagée sera accompagnée des raisons de son abandon.

Immédiatement, compte tenu des moyens financiers inexistantes pour la réalisation de la mission, mise à part ma rémunération en tant que stagiaire, une première éventualité s'est dissipée : la solution dite payante. En outre, cette solution enlevait tout le côté formateur en ingénierie ontologique dont j'ai pu bénéficier au cours de ce stage et il n'est pas certain non plus que cette solution convienne dans le cadre du projet ADONIS. Néanmoins, devant l'ampleur de la tâche cette solution de facilité m'a traversé l'esprit. L'idée principale de cette solution que j'ai appelée « payante » consiste à s'offrir les services plus ou moins complets d'une entreprise spécialisée sinon dans l'ingénierie ontologique du moins dans la recherche d'information et les technologies du web. Une des plus connues actuellement dans l'univers de la gestion de l'information, et celle qui a été une des premières à s'intéresser aux ontologies est Mondeca⁴⁵. L'option payante ayant été écartée immédiatement, pour des raisons de planning, je n'ai pas effectué de recherches précises sur cette entreprise ni cherché à entrer en contact avec leurs équipes directement afin de savoir si notre projet les intéressait. Cela dit, d'après les informations disponibles sur le site [mondeca.com](http://www.mondeca.com), cette entreprise dispose d'une plate-forme logicielle qui pourrait convenir à nos besoins : l'ITM, Intelligent Topic Manager. Enfin, avec cette solution envisagée de prime abord et très peu réfléchie car Mondeca ne propose pas explicitement de transformer un thésaurus en ontologie, c'est surtout une expérience dans le domaine des ontologies qui était attendue, un moyen de trouver un appui d'expert dans le domaine.

La seconde solution, qui m'a d'ailleurs retenue le plus longtemps, est celle de la terminologie. Cette manière de concevoir l'ontologie m'a été suggérée par un des membres du TC 37, Comité Technique 37 de l'ISO, consacré à l'élaboration des normes concernant la terminologie et autres ressources langagières et ressources de contenus. La méthode envisagée

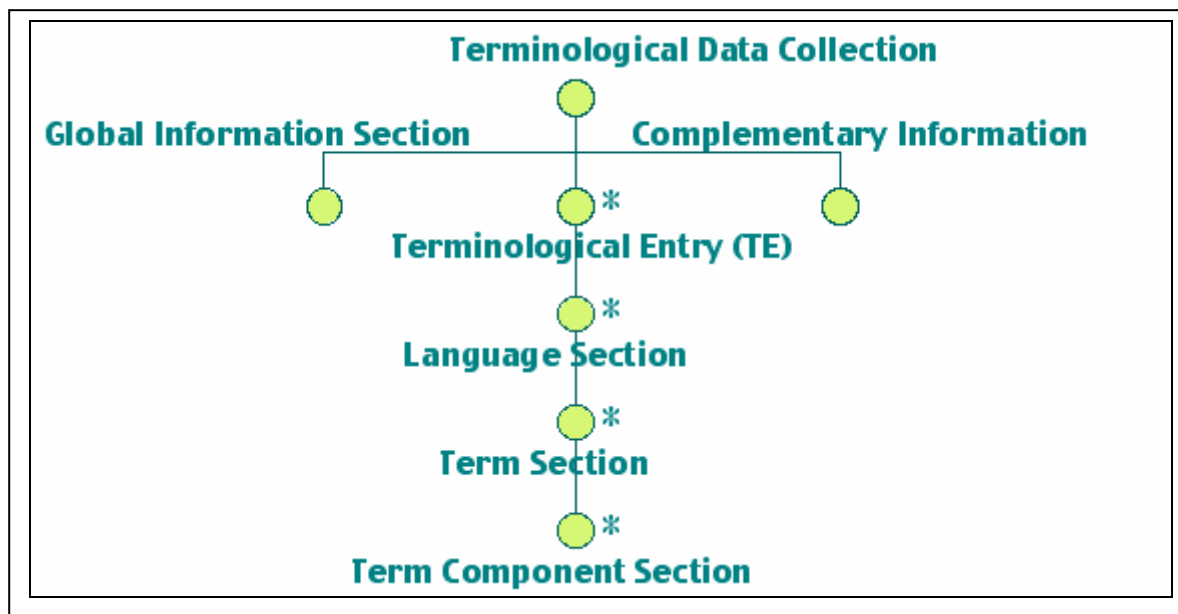
⁴⁵ <http://www.mondeca.com>

consistait à transformer dans un premier temps le micro-thésaurus PACTOLS-Sujets en base terminologique puis, dans un second temps de se servir de cette base terminologique comme ressource pour l'ontologie. En procédant de cette façon, l'ontologie reposait sur un fondement terminologique normalisé, et non standardisé comme l'est OWL qui provient du W3C, et garantissait entre autres choses une interopérabilité entre les systèmes. En outre, la terminologie comme le prévoit la norme ISO 704 « analyse et organise les concepts sur la base des notions identifiées dans un système onomasiologique ». C'est-à-dire que l'étude sémantique des dénominations se fait en partant du concept et recherche ensuite les signes linguistiques qui lui correspondent, alors que dans un système sémasiologique – plus courant dans les ontologies – on part du signe pour aller vers l'idée.



17. Représentation des deux schémas : onomasiologique et sémasiologique

Ainsi, la stratégie onomasiologique propose de partir du sens, de la définition, pour ensuite en décliner les différentes formes. Pour ce faire, concrètement, il existe un schéma unique de base appelé TMF, Terminological Markup Framework, correspondant à la norme ISO 16642. C'est un cadre de mise en œuvre normalisé pour la terminotique – discipline qui s'intéresse aux outils informatiques pour la collecte, la production, le traitement l'échange et la diffusion des données terminologiques – lui-même lié au TML, Terminological Markup Language, un langage XML spécialisé pour l'expression des données terminologiques.



18. Méta-modèle d'organisation d'une base de données terminologique ⁴⁶

Ce méta-modèle est un modèle abstrait de représentation de terminologies multilingues : une base de données terminologique doit comporter diverses informations telles qu'un titre, une date de création, de modification ainsi que quelques commentaires d'explication. Les données sont ensuite structurées en entrées terminologiques, par langue, par terme, et enfin par les relations qu'elles entretiennent avec d'autres termes. Le but de ce schéma n'est pas d'uniformiser, mais de rendre interopérable, compatible et réutilisable plusieurs applications de bases de données terminologique comme pourrait l'être l'application du thésaurus dans ce schéma en ontologie. En effet, une fois ce schéma de base appliqué, c'est au concepteur de choisir les relations hiérarchiques et de les créer tout en sachant que ces relations s'exerceront non pas entre les termes, mais entre les entrées terminologiques qui correspondent aux concepts.

Néanmoins, si théoriquement cette méthodologie semble tout à fait cohérente et rationnelle, deux inconvénients majeurs sont apparus, en relation avec les quatre critères de sélection évoqués précédemment. En effet, dans la pratique, la réalisation concrète de cette méthode a posé problème d'une part en ce qui concerne mon niveau de compétence en informatique et en ingénierie des connaissances et d'autre part au niveau du temps de réalisation. Au fur et à mesure que je me documentais sur cette méthode envisagée, le temps avançait et je ne trouvais toujours pas de logiciel adapté à mes besoins, ayant abandonné très rapidement l'idée d'apprendre le langage TML et de créer ma propre application, par manque de compétence en la matière et pour des raisons de temps limité. En outre, des doutes s'accumulaient quant à la pertinence de la méthode pour le cas particulier de ma mission : il m'était demandé de réaliser une ontologie, or,

⁴⁶ Document extrait du site : <http://www.loria.fr/projets/TMF/>, TMF Webpage.

rien dans la littérature ne parlait de cette option pour concevoir une ontologie. C'est alors que j'ai eu l'occasion de rencontrer Madame Aussenac-Gilles, présente à la MAE pour un autre projet. Je lui ai fait part de mes recherches et de l'impasse dans laquelle je me trouvais au sujet de la méthode terminologique que j'envisageais pour le moment plus qu'une autre – en effet, il convient de préciser que c'est par souci de clarté que les quatre solutions sont décrites une à une mais il est évident que la progression ne s'est pas effectuée de manière si abrupte et bornée. Diverses recherches étaient effectuées en parallèle et l'organisation des solutions montre seulement de quelle manière s'est déroulée la progression de mon raisonnement pour parvenir à me fixer sur une méthodologie. Ainsi, c'est sans me suggérer d'abandonner la piste terminologique, qu'elle trouvait néanmoins tortueuse pour le but à atteindre, que Nathalie Aussenac-Gilles m'a orientée vers d'autres approches en m'invitant à me renseigner davantage sur le langage SKOS et en me conseillant fortement d'utiliser dans un premier temps un certain logiciel appelé DOE qui m'aiderait à structurer l'ontologie d'après les concepts présents dans le thésaurus. Enfin, une journée d'études consacrée à l'état et aux nouvelles perspectives du web sémantique et social a confirmé mes doutes quant à la pertinence de cette solution d'après les critères imposés mais également par rapport à ce qui se faisait à ce moment en matière d'ingénierie ontologique. Cette méthode n'était pas à l'ordre du jour pour créer une ontologie et surtout, la démarche était trop longue et trop complexe pour mon niveau de compétence.

Après ce deuxième abandon, majeur, il reste donc deux manières de concevoir l'ontologie d'après le micro-thésaurus PACTOLS-Sujets que je n'ai pas retenues : l'une concernant l'utilisation du langage SKOS, et l'autre utilisant directement l'éditeur d'ontologie Protégé⁴⁷. Je ne m'attarderai pas sur le fonctionnement de SKOS ni sur celui de Protégé car ils font déjà chacun l'objet à un endroit ou un autre dans ce mémoire d'une description⁴⁸. J'expliquerai donc simplement en quoi aurait pu consister la méthode, et pourquoi elle n'a pas été appliquée.

En ce qui concerne l'utilisation de SKOS, *Simple Knowledge Organization System*, la solution aurait consisté à transposer le thésaurus dans ce langage. C'est-à-dire à encoder presque tel quel le thésaurus. Cela aurait d'ailleurs pu se faire à l'aide d'un logiciel tel que ThManager (outil en Open Source pour concevoir et visualiser les vocabulaires en SKOS RDF). Ce n'est donc pas, cette fois, la mise en application qui a été la raison principale de l'abandon de cette piste, mais plutôt sa relative simplicité, car transposer le thésaurus dans ce langage, même si quelques modifications auraient sans doute été nécessaires, n'apportait pas toutes les améliorations espérées par la

⁴⁷ Logiciel libre disponible à l'adresse suivante : <http://protege.stanford.edu/>

⁴⁸ SKOS est défini dans le chapitre II, II. 2 et le fonctionnement de Protégé est expliqué succinctement dans le chapitre II, III, 1 et plus en détail dans ce chapitre III, III, 2

conception d'une ontologie. En effet, SKOS n'a clairement pas, aujourd'hui, pour ambition de modéliser une ontologie, mais de rendre un vocabulaire contrôlé comme le thésaurus mais aussi les taxinomies ou autres classifications accessibles dans l'environnement du web sémantique sans se soucier de la lourde tâche de formalisation induite par une ontologie. En d'autre terme, l'utilisation de SKOS n'apporterait pas la plus-value escomptée par la réalisation de la mission et elle nécessiterait par conséquent dans un second temps, l'export du thésaurus en SKOS dans un éditeur d'ontologies comme Protégé qui accepte le langage SKOS. Nous revenons donc encore une fois à ce logiciel, Protégé, alors pourquoi avoir écarté la dernière solution qui aurait consisté à entrer « tout simplement » le micro-thésaurus PACTOLS-Sujets dans cet éditeur d'ontologies ? La raison est simple, étant débutante dans le domaine de l'ingénierie ontologique, il n'était pas naturel pour moi de « défaire » la structure du thésaurus afin d'en faire des taxinomies. Le passage immédiat dans Protégé, même avec une première manipulation (sur une ancienne version) effectuée au cours de ma formation universitaire, me semblait quasiment impossible. « Par où commencer ? Quelle structure, quel squelette, donner à l'ontologie ? », telles étaient les questions récurrentes. Finalement, j'étais à la recherche d'une rationalisation de la conception de l'ontologie. Comment partir sur de bonnes bases et être certaine de tenir la structure jusqu'à la fin ? C'est dans l'utilisation du logiciel DOE (*Differential Ontology Editor*), créé d'après l'engagement sémantique et l'engagement ontologique initiés par Bruno Bachimont et décrit dans le deuxième chapitre de ce mémoire, que j'ai enfin trouvé des réponses à ces questions comme nous allons rapidement pouvoir le découvrir puisqu'il s'agit de la solution que j'ai choisi d'appliquer.

Mais avant d'y parvenir, revenons quelques instants sur la logique sous-jacente à ces quatre solutions rejetées. Voyons comme il n'est pas inintéressant, pédagogiquement, d'analyser le cheminement parcouru.

En effet, à la vue de toutes ses solutions laissées de côté, mais envisagées chacune pendant un laps de temps plus ou moins long, on est en droit certes de penser qu'il aurait été préférable, compte tenu du temps relativement court dont je disposais pour réaliser ma mission, que je trouve plus rapidement la méthodologie adéquate. Mais on peut également s'interroger sur les raisons de ce parcours. Pourquoi être passée par ces étapes alors que j'avais rencontré dans mes lectures et ma formation universitaire sinon la description du logiciel DOE, au moins celle de Protégé ? Avec le recul, il me semble que ce cheminement suit une progression logique et démontre une certaine maturité dans mes réflexions et mes choix. Pour être plus claire, si l'on reprend l'une après l'autre les différentes étapes, on remarque que dans la première, j'ai choisi la facilité, la rapidité et la sécurité d'être encadrée par des ontologistes ; dans la seconde, il y a une

volonté de s'inscrire dans du long terme, de présenter un projet rigoureux mais que je n'étais pas en mesure de réaliser seule et en si peu de temps, en outre, dans cette solution j'avais un appui fort de la part d'une personne au niveau théorique, mais pas dans la pratique ; la troisième est le témoignage d'une prise de conscience que le temps était compté mais son abandon extrêmement rapide démontre également le souci d'apporter une plus-value au thésaurus et par là-même au projet et à ma mission ; enfin, le dernier abandon, provisoire puisque l'utilisation de Protégé est envisagée dans un second temps, illustre l'ambition de réaliser un travail reposant sur des bases solides car en optant pour une démarche certes plus longue, cela a au moins eu le mérite de tenter de rationaliser en priorité la conception de l'ontologie. Par ailleurs, cela correspondait davantage à mes compétences dans le domaine et, cette rationalisation, comme nous allons le découvrir, permettra à une autre personne de reprendre le travail commencé si besoin est.

3) Solution retenue : utilisation du logiciel DOE

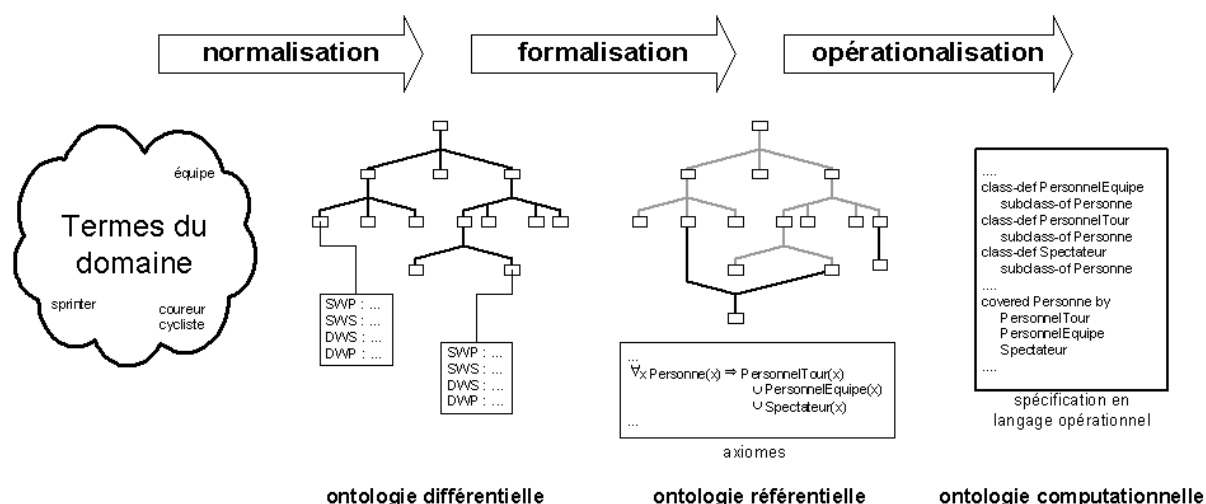
Cette partie s'inspire de l'article publié par les créateurs du logiciel : Raphaël Troncy et Antoine Isaac : DOE : une mise en œuvre d'une méthode de structuration différentielle pour les ontologies [TRONCY et ISAAC 2002].

DOE est un éditeur d'ontologie, que l'on a d'ailleurs évoqué dans le deuxième chapitre de ce mémoire consacré à la conception des ontologies.

Il s'agit d'un outil d'aide à la structuration d'ontologies développé à l'Institut National de l'Audiovisuel (l'INA) suite notamment au projet OPALES (Outils pour des Portails Audiovisuels Éducatifs et Scientifiques). Il a pour ambition de contribuer à résoudre le problème posé par la création des taxinomies, fondements des ontologies. Comment et selon quels critères organiser les concepts (et les relations) propres à un domaine selon la relation de subsumption nécessaire à la création de toute taxinomie et par là même d'ontologie ? Car, selon les créateurs de DOE, même s'il existe de nombreux travaux et productions en matière de conceptualisation d'ontologies, très peu s'intéressent à la construction des hiérarchies produites et se concentrent davantage sur des questions comme celles du cycle de vie des ontologies construites, de l'ordonnancement des étapes auquel l'ontologiste devra se soumettre pour valider son travail (définition des questions auxquelles devra répondre le système, collecte de documents pertinents pour l'acquisition des connaissances, évaluation des résultats...). À travers la création de ce logiciel, ils ont ainsi voulu guider le concepteur d'ontologie dans la phase de conceptualisation

proprement dite, celle où les concepts de l'ontologie sont dégagés, définis par un certain nombre de propriétés et organisés entre eux. Pour ce faire, ils se sont appuyés sur la méthodologie proposée par Bruno Bachimont qui lui-même s'est inspiré des travaux de François Rastier sur la sémantique différentielle – méthode que nous avons d'ailleurs évoquée auparavant dans ce mémoire⁴⁹ – qui vise à contraindre l'utilisateur à un engagement sémantique en introduisant une normalisation sémantique des termes manipulés dans l'ontologie. Ainsi, les deux créateurs ont voulu redonner la place qui se doit au langage en l'utilisant comme moyen principal d'explicitation des concepts. Enfin, pour terminer sur la présentation des principaux objectifs de cet éditeur d'ontologie qui vient s'ajouter à une liste déjà bien fournie, on peut dire que DOE se propose de remplir un vide méthodologique présent dans la plupart des principaux éditeurs évoqués précédemment⁵⁰. Il tente de rationaliser la conception des ontologies en implémentant une méthodologie particulière de structuration qui règlemente en quelque sorte la création de l'ontologie.

Cette méthodologie propose trois étapes : la normalisation, la formalisation et l'opérationnalisation comme l'illustre la figure suivante extraite de [TRONCY et ISAAC 2002].



19. Les trois étapes de la méthode de construction d'ontologies

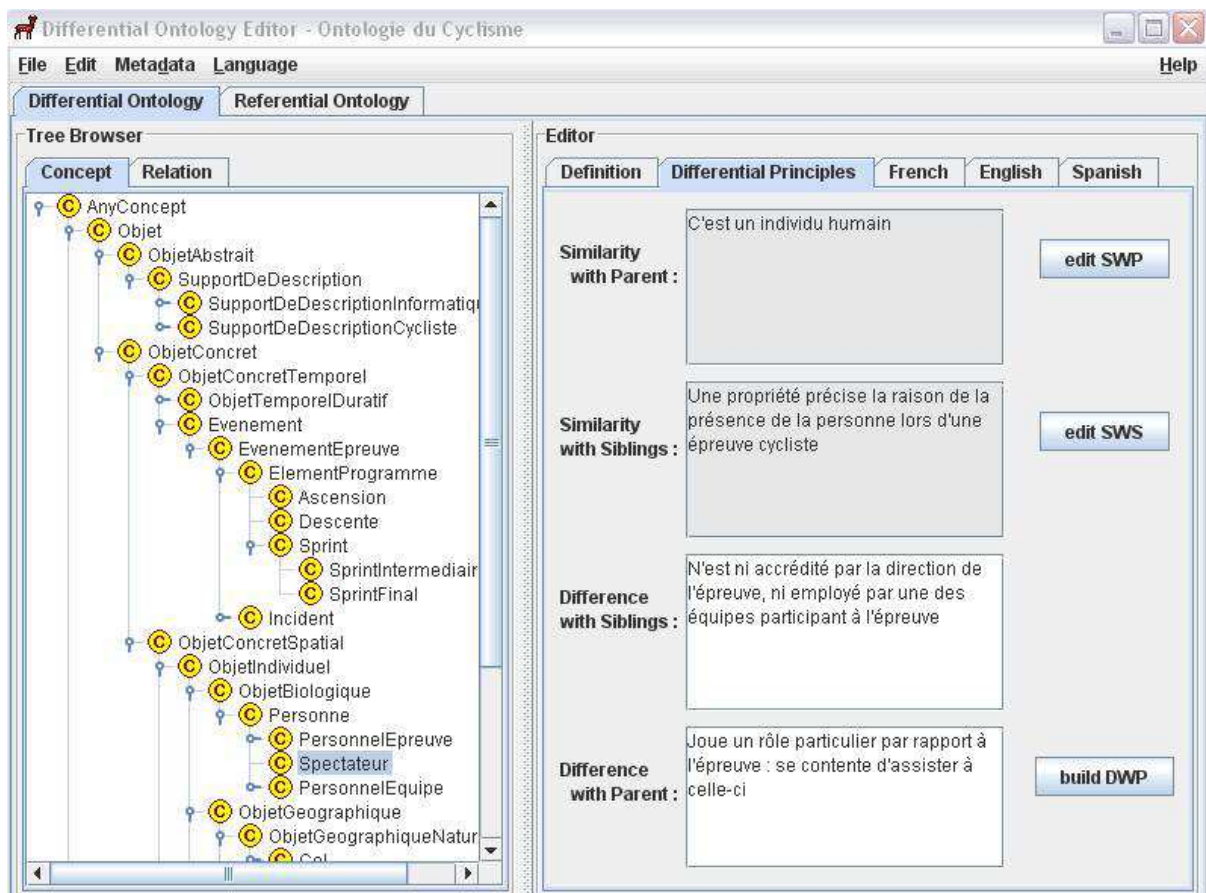
Ces trois étapes sont donc prises en compte par le logiciel DOE et donnent lieu à la création de trois ontologies : différentielle, référentielle et computationnelle, la dernière étant le fruit du seul travail du logiciel.

Concrètement, se servant de la sémantique différentielle comme d'une base, le logiciel propose de rentrer les concepts que l'on veut voir figurer dans son ontologie pas à pas en veillant à

⁴⁹ Chapitre II, I,2

⁵⁰ Chapitre II, III, 5

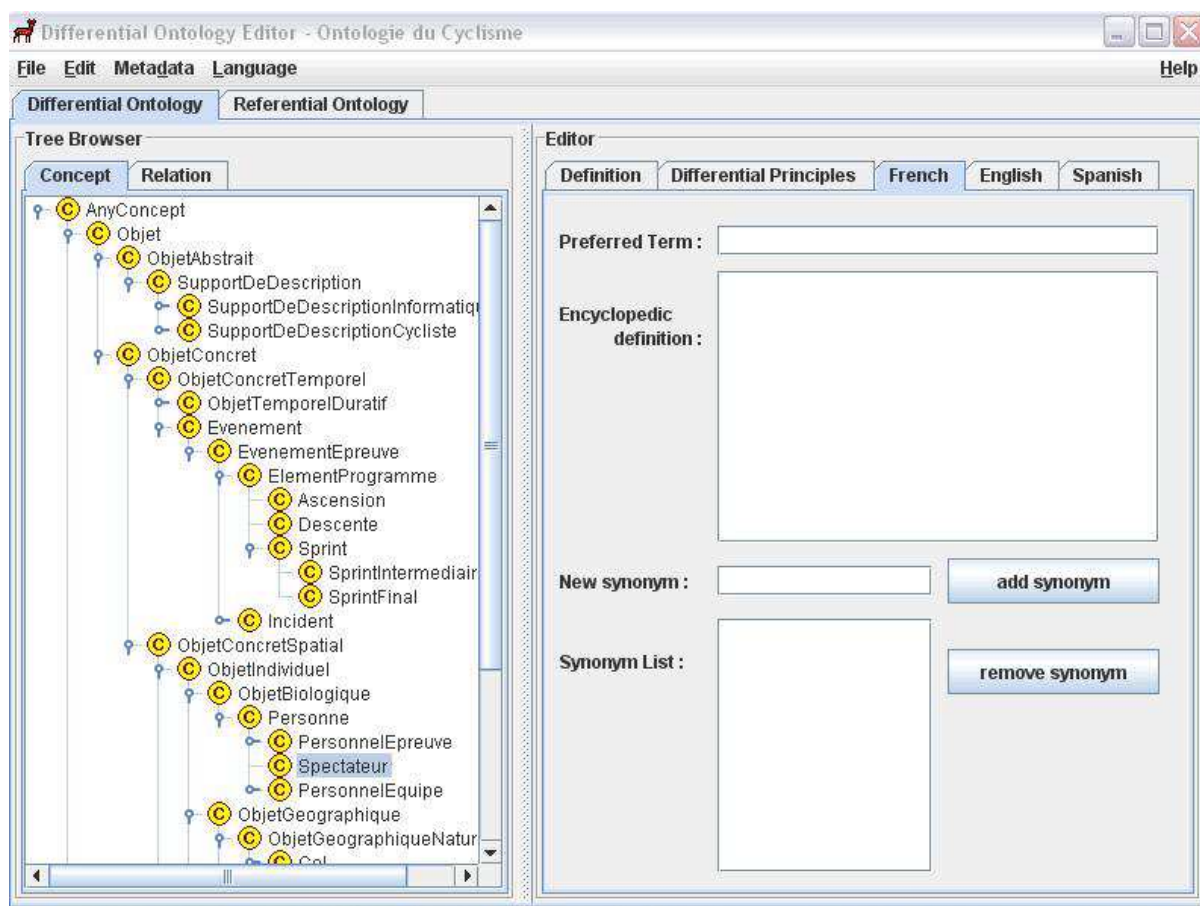
désambigüiser le sens de chacun des termes. Cette désambiguïsation se fait au moyen de la définition de chaque terme dans son rapport aux autres. Le terme se définit à la fois par rapport à son hyperonyme et par rapport à ses co-hyponymes (sa fratrie), par l'explicitation des axes de communauté sémantique qu'il partage avec eux et par la précision des traits sémantiques qui l'en distinguent. Ainsi, pour chacun des termes ajoutés, on doit spécifier quatre « principes différentiels » : la similarité sémantique qui le lie à son père (*similarity with parent*, ou SWP), le principe sémantique qu'il partage avec ses frères (*similarity with siblings*, ou SWS) et sa position particulière sur cet axe sémantique, qui le différencie de sa fratrie (*difference with siblings*, ou DWS) et enfin la spécificité qui le différencie de son père (*difference with parent*, ou DWP). Comme on peut le voir sur la copie d'écran qui suit, peu à peu, en appliquant cette règle, se construit un arbre de concepts.



20. Copie d'écran montrant la structure de l'ontologie du cyclisme, réalisée par R. Troncy et A. Isaac, et les principes différentiels de la notion Spectateur.

On remarque d'après l'exemple issu de l'ontologie du cyclisme créée par Raphaël Troncy et Antoine Isaac qu'il est bien possible de relier chaque terme à un autre par la relation « est un », c'est-à-dire de subsomption, spécificité caractéristique des taxinomies. En outre, cette copie d'écran nous donne l'occasion de découvrir d'autres particularités du logiciel. Dans l'onglet

« définition », à titre informatif, est générée une définition différentielle des notions en affichant le chemin menant de la racine de l'arbre à la notion considérée et en récapitulant les valeurs des différences avec les pères rencontrés sur ce chemin. Par ailleurs, il est possible d'attacher à chacune des notions, des aspects propres à une langue donnée. Deux éléments terminologiques peuvent être saisis dans l'éditeur : le Terme préféré, celui qui est utilisé pour désigner le plus souvent la notion (et qui a donc de fortes chances d'apparaître également en tant que libellé de la notion), et une liste des différents Synonymes de celle-ci. L'éditeur gère aussi l'ajout de nouvelles langues et permet à l'utilisateur de renseigner les champs terminologiques en français, anglais, espagnol, italien, allemand et néerlandais. Parmi les attributs de la notion propre à une langue, les créateurs de DOE ont également choisi de faire figurer une Définition encyclopédique – ces possibilités complémentaires du logiciel sont visibles sur la copie d'écran qui suit.

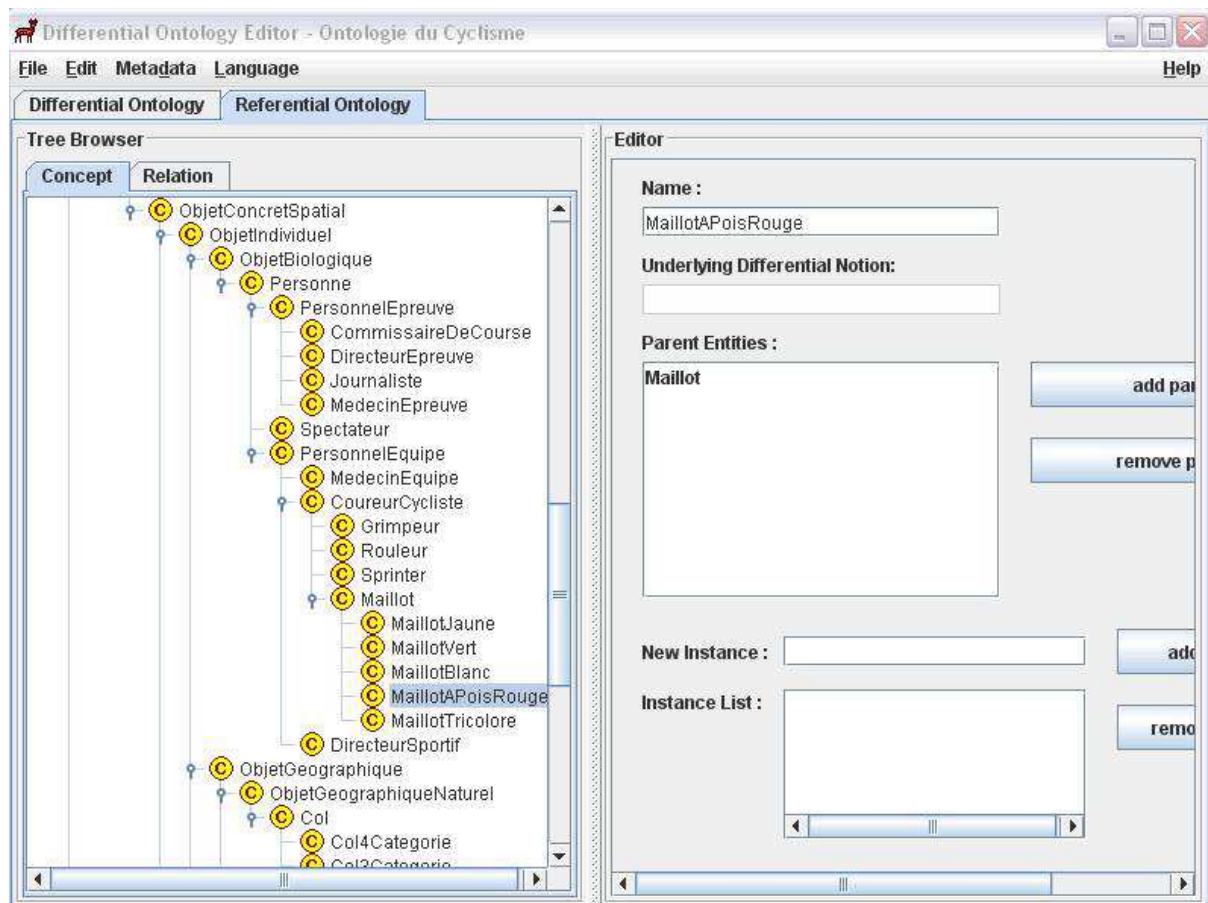


21. Copie d'écran montrant les possibilités complémentaires de DOE.

Cette définition n'est pas issue d'un processus de modélisation linguistique, mais se veut l'équivalent d'un commentaire, d'un complément informel de la définition élaborée à l'aide des principes différentiels. En revanche, il convient de préciser que cette gestion de plusieurs langues, pour un certain nombre d'informations attachées aux notions, n'a pas pour ambition de rendre

l'éditeur multilingue. Ce problème est en effet très complexe à gérer ; des notions verbalisables dans une langue pourraient ne pas l'être immédiatement dans une autre. L'éditeur se contente donc de donner la possibilité de mettre en correspondance des termes et des définitions informelles. De plus, on constate qu'un certain nombre d'informations centrales (les principes différentiels, notamment) ne peut être explicité que dans une seule langue, qui prendra alors le statut de langue de référence pour l'ontologie. Ainsi, l'on comprend que toutes les langues ne sont pas traitées de la même manière.

En ce qui concerne la seconde étape prise en considération par le logiciel, parallèlement à la constitution de l'ontologie différentielle, il se crée une ontologie référentielle. Cette dernière est concrètement « le produit fini » de l'ontologie au niveau des connaissances. On doit pouvoir attacher aux notions de l'ontologie différentielle – étape qui a permis de fixer un certain nombre de notions en normalisant leur sens – une signification formelle permettant de fixer la référence des notions, et qui pourra être traduite de manière automatique dans un langage compréhensible par l'ordinateur. L'ontologie référentielle a donc comme fondement la taxinomie créée grâce aux 4 principes différentiels. Aucune des informations apportées dans l'ontologie ne peut remettre en cause la hiérarchie de subsomption : l'arbre de concept établi dans la première étape est donc repris intégralement dans la seconde. L'utilisateur peut à ce niveau d'avancement ajouter des concepts qui ne présentent plus les informations sémantiques des notions différentielles. Ainsi, dans l'ontologie du cyclisme qui est donnée en exemple dans DOE, *Maillot* n'est présent que dans l'ontologie référentielle. Cela signifie qu'il ne répondait pas aux 4 principes différentiels mais que par extension, les créateurs de l'ontologie ont jugé nécessaire de le faire figurer aux côtés des déclinaisons des types de coureurs cyclistes : grimpeur, rouleur et sprinter.



22. Copie d'écran montrant la présence et la situation du concept *Maillot* dans l'ontologie référentielle du cyclisme

Ces nouveaux objets peuvent être ajoutés en tant que spécialisation d'une entité issue de l'ontologie différentielle, mais aussi d'une entité introduite lors de l'étape référentielle – pour reprendre l'exemple de *Maillot*, dans l'ontologie référentielle, ce dernier est décliné selon ses couleurs : jaune, vert, à pois rouges, blanc, tricolore. L'héritage multiple étant possible, l'éditeur tient à jour, pour chaque entité, la liste de ses parents et permet d'en ajouter ou d'en retirer. En revanche, la possibilité de définir des concepts par conditions nécessaires et suffisantes avec des combinaisons booléennes, ou encore par l'expression de contraintes relationnelles, n'est pas proposée dans DOE. Cela veut dire qu'il est impossible de faire figurer de nouveaux concepts dans l'ontologie référentielle en se basant sur des déductions logiques telle que la règle de transformation concernant l'associativité issue de l'algèbre de Boole⁵¹ :

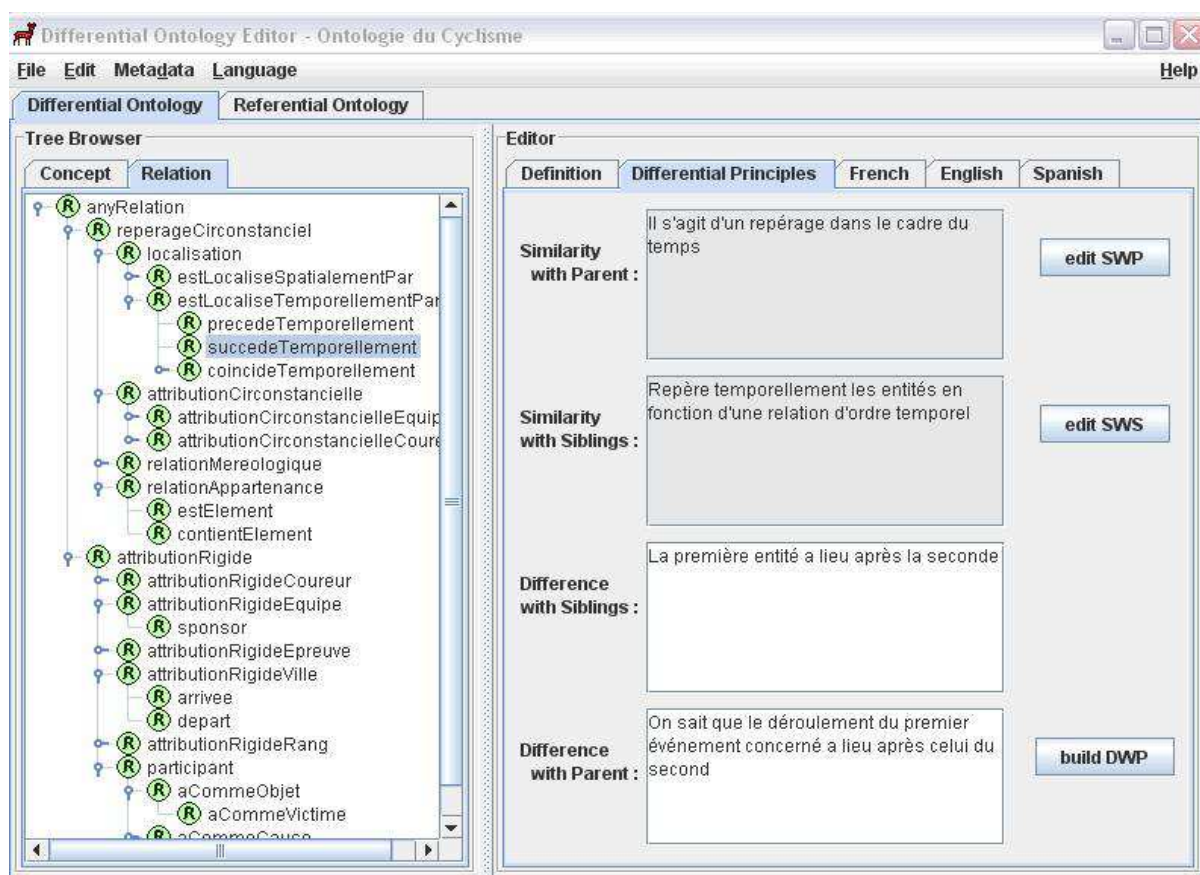
⁵¹ L'algèbre de Boole est la partie des mathématiques, de la logique et de l'électronique qui s'intéresse aux opérations et aux fonctions sur les variables logiques. Le nom provient de George Boole, un mathématicien britannique qui, durant le milieu de XIXe siècle, restructura complètement la logique en un système formel. Plus spécifiquement, l'algèbre booléenne permet d'utiliser des techniques algébriques pour traiter les expressions à deux valeurs de la logique des propositions.» Contributeurs à Wikipédia. Algèbre de Boole (logique) [en ligne]. Wikipédia,

$$A + (B + C) \leftrightarrow (A + B) + C$$

$$A \cdot (B \cdot C) \leftrightarrow (A \cdot B) \cdot C$$

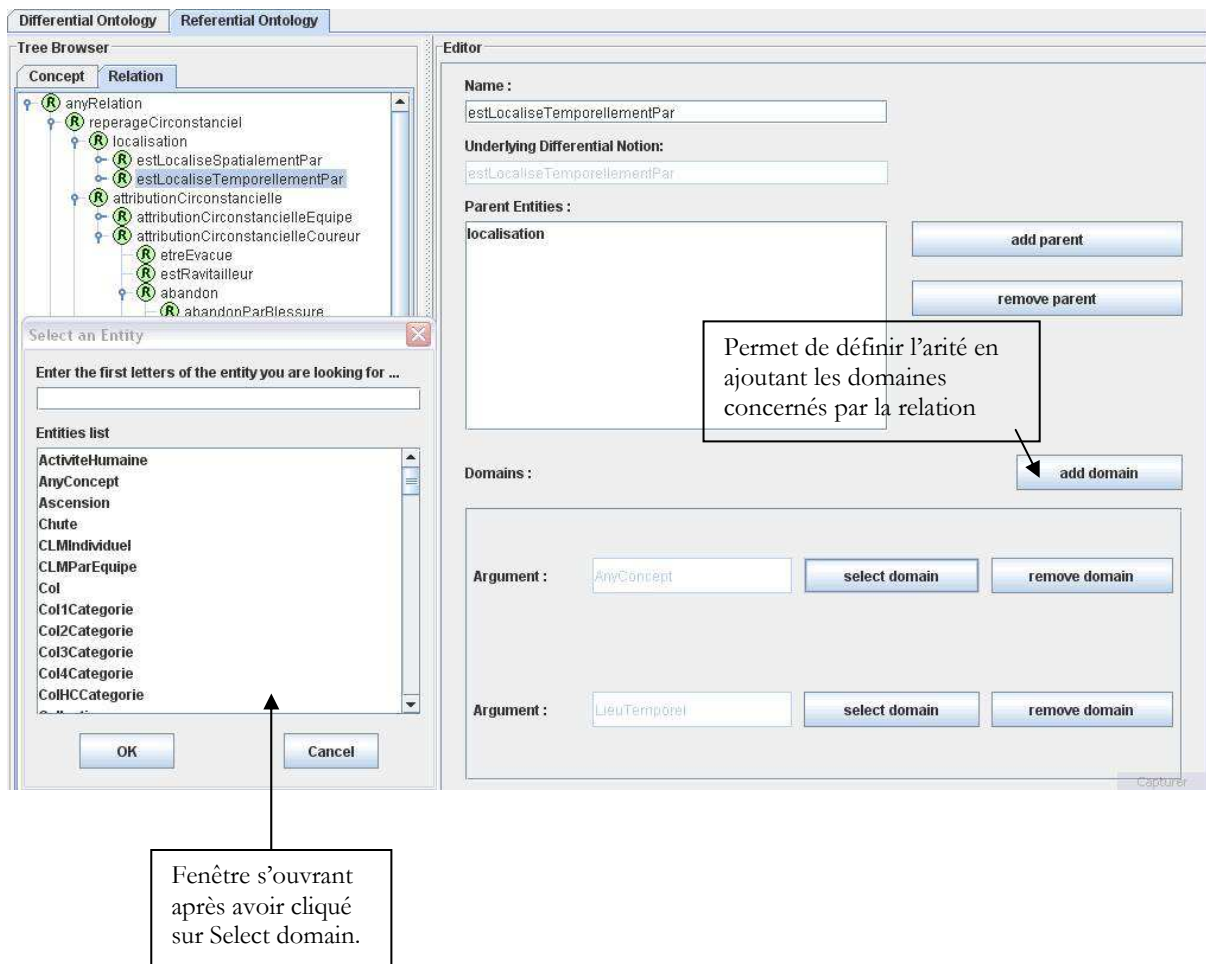
Cette fonctionnalité, présente dans Protégé, équivaldrait à un pas de plus vers les règles d'inférence, règles qui contribuent notamment à améliorer la recherche documentaire, mais ce n'est pas l'objectif de DOE qui préfère laisser ce processus aux mains d'autres éditeurs d'ontologies.

Enfin, l'éditeur ne serait pas complet s'il se contentait de traiter uniquement les concepts selon la relation de subsomption sous-entendue dans les taxinomies. Nous pouvons voir sur la copie d'écran ci-dessous, qu'il faut, parallèlement à la création des deux ontologies, différentielles et référentielles pour les concepts, faire de même avec les relations, selon le même principe de base qu'est le principe différentiel.



23. Copie d'écran montrant la hiérarchie des relations selon le principe différentiel

Selon Raphaël Troncy et Antoine Isaac, au niveau de l'ontologie référentielle, la description de la référence des relations est spécifique : pour une relation donnée, il faut pouvoir spécifier son arité – c'est-à-dire son nombre d'arguments, d'entrées – ainsi que ses domaines. L'éditeur propose d'entrer ces informations. Le choix des domaines se fait parmi la liste des concepts déjà saisis dans l'ontologie. À ce stade, l'éditeur prend en compte les relations de spécialisation entre relations : l'arité, quand elle a été définie, doit être propagée correctement le long de la hiérarchie de subsomption. De même, les domaines d'une relation doivent hériter des domaines de la relation dont elle hérite. En cas d'héritage multiple, les domaines de la relation fille doivent être des sous-concepts des domaines de toutes les relations parentes. Sur la copie d'écran suivante, on peut d'ailleurs remarquer les deux arguments : Any Concept et Lieu Temporel qui qualifient la relation « Est Localisé Temporellement Par ».



24. Copie d'écran de l'interface de DOE permettant de saisir les spécificités des relations

En ce qui concerne la troisième étape, celle de l'opérationnalisation, il va être question de chercher à munir les concepts présents dans l'ontologie référentielle d'une signification en termes d'opérations informatiques. De cette manière, on spécifie le comportement des objets

informatiques présents dans le système qui utilise l'ontologie, pour aboutir à ce qui est appelé une Ontologie computationnelle. Les concepts computationnels se définissent par les inférences, les calculs que pourra effectuer un système à partir des données des individus qu'ils couvrent dans le monde. Pour ce faire, le système utilise un langage opérationnel de représentation des connaissances qui fait appel à des capacités d'inférence précises répondant aux besoins exprimés lors de la spécification du système. Pour un langage de représentation reposant sur des graphes conceptuels, il s'agira d'opération de manipulation de graphes, tandis que pour un langage fondé sur le paradigme des logiques de description, il s'agira plutôt de tests de subsomption permettant la classification des individus introduits dans la base de connaissances. Dans DOE, ce passage à une ontologie computationnelle se fait par un export de l'ontologie référentielle dans un certain nombre de langages opérationnels. Ainsi, il est théoriquement possible d'exporter l'ontologie référentielle créée dans DOE, dans un éditeur d'ontologies qui prend en charge la suite de la conception d'une ontologie.

Certes, cette présentation du logiciel est relativement succincte par rapport à celle qui en est faite par ses créateurs mais elle n'en est pas moins rigoureuse. Néanmoins, malgré ou à cause de la brièveté de cette présentation, ce logiciel peut sembler complexe voire peu compréhensible, en dépit du souci de donner de nombreux exemples, mais sa mise en application et son utilisation pas à pas grâce à la réalisation de la transformation du thésaurus en ontologie vont permettre d'éclaircir ces propos.

III. Réalisations et perspectives

Cette partie, puisqu'il s'agit de la dernière, vient clôturer ce mémoire et se veut être comme l'aboutissement, la conséquence, des propos précédents. L'exposé des réalisations et des perspectives liées à la transformation du micro-thésaurus PACTOLS-Sujets en ontologie est entre autres choses l'occasion de mettre en relation les savoirs acquis lors des deux premiers chapitres, l'objet de l'étude : le micro-thésaurus PACTOLS-Sujets et l'utilisation du logiciel DOE proposant la mise en application de la méthode de Bruno Bachimont. Très simplement, mais de manière efficace, nous verrons donc dans un premier temps le travail qui a été accompli dans le but de

créer une ontologie à partir du thésaurus, et dans un second temps, nous nous projetterons dans l'avenir en donnant les différents objectifs qui restent à atteindre.

1) Réalisations

Les réalisations sont étroitement, sinon uniquement, liées au processus de transformation du micro-thésaurus PACTOLS-Sujets en ontologie. Ce processus de transformation doit suivre deux phases : la modélisation de l'ontologie à l'aide du logiciel DOE, puis sa représentation et son raffinement dans un éditeur dit complet – Protégé. Cette deuxième phase permettra l'implantation du modèle créé dans DOE, le choix d'un langage de représentation et le raffinement du modèle en fonction de ce que l'on souhaite faire de l'ontologie.

À la fin de ma mission, la première phase a été accomplie pour ce qui concerne une partie du thésaurus mais avant d'en décrire le modèle, voyons quelles ont été les différentes étapes de construction et quelle est la structure, le squelette, qui a été retenu pour concevoir l'ontologie, c'est-à-dire pour collecter, assembler et classer les concepts et les relations.

Appropriation et destructuration du thésaurus

L'un des premiers défis à relever pour concevoir une ontologie est généralement celui de réunir les concepts nécessaires à la représentation du domaine de connaissance. Comme nous avons pu le voir précédemment⁵², pour y parvenir, Bruno Bachimont et de nombreux autres spécialistes en ingénierie des connaissances proposent d'extraire les termes à partir d'un corpus spécialisé au moyen de méthodologies proposées en traitement automatique des langues (TAL). *A priori* il semble inutile d'utiliser cette méthode puisque nous disposons déjà des concepts devant figurer dans l'ontologie. Cependant, étant donné la taille du micro-thésaurus, cela pourrait être un excellent moyen pour classer les termes selon les relations qu'ils entretiennent entre eux et vérifier la pertinence des termes présents dans le micro-thésaurus PACTOLS-Sujets. En effet, nous pourrions comparer le résultat de l'extraction à la composition du thésaurus et ainsi voir s'ils sont représentatifs du corpus – corpus sensé être le reflet du domaine concerné par l'ontologie. Cependant, ces méthodes requièrent un minimum de temps pour, d'une part réunir le corpus et d'autre part, analyser le résultat. Quelle que soit la méthode d'extraction envisagée le temps

⁵² Chapitre II, I. 2

nécessaire pour effectuer le processus dans son entier était trop long. En revanche, il était essentiel de prendre connaissance du thésaurus. Quels termes le composent ? Quels sont leurs thématiques, leurs genres grammatical, lexical, etc. ? Comment sont-ils classés ? Quel est le « mouvement », la logique du thésaurus ? En bref, il me fallait m'approprier le thésaurus avant de parvenir à le transformer en ontologie.

La solution dite automatique ayant été écartée faute de temps, il ne restait donc plus que la méthode « traditionnelle » pour connaître l'objet de l'étude : sa manipulation à travers son utilisation et son étude, son analyse.

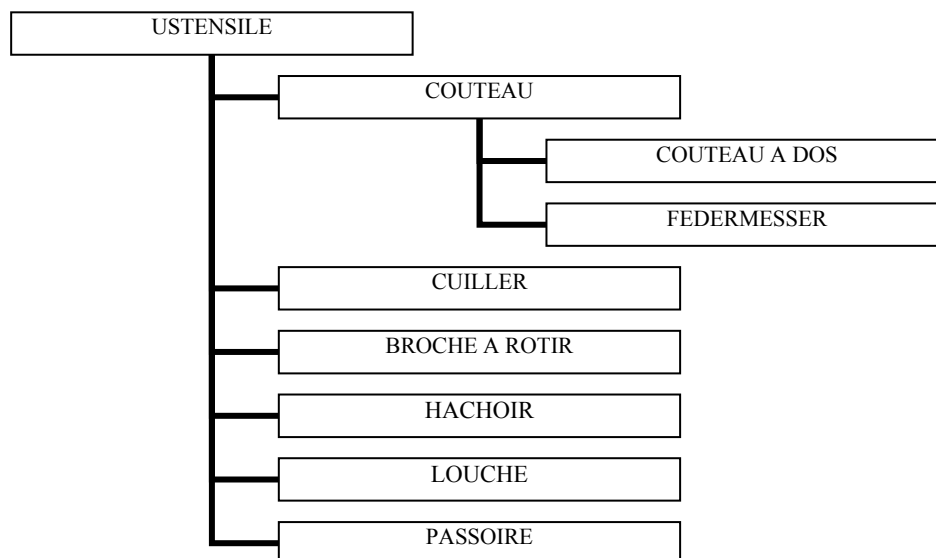
Les occasions n'ont pas manqué pour utiliser le thésaurus PACTOLS. En effet, le service dans lequel j'ai effectué mon stage se trouve être le cœur du réseau FRANTIQ, le lieu de centralisation où toutes les interrogations, tous les problèmes remontent et où les maintenances de la base du CCI et du thésaurus sont effectuées. J'ai donc participé, comme le stipulait l'intitulé de mon stage *Participation aux différentes facettes d'un centre de documentation*, aux tâches quotidiennes d'un centre de documentation. De cette manière, en « nettoyant » la base du CCI gérée sous Koha⁵³, en enrichissant le micro-thésaurus PACTOLS-Sujets en Notes d'Application et parfois en gérant ses termes (suppression, création, déplacement) dans le logiciel OpenTheso, en effectuant les recherches bibliographiques demandées par les chercheurs et en effectuant des interrogations sur la base de données DAPHNE alors en période de tests, j'ai pris peu à peu connaissance de la composition et des évolutions du thésaurus et de son utilité en tant que vocabulaire contrôlé permettant la recherche et l'indexation de références bibliographiques.

Par ailleurs, outre cette appropriation par manipulation fonctionnelle, j'ai procédé à son analyse par l'étude de sa composition selon deux méthodes : le classement des termes dans un tableau et l'ordonnancement de certains dans des graphiques. Le classement des termes a eu pour but de me contraindre à catégoriser le vocabulaire contenu dans le thésaurus. Comme nous l'avons vu précédemment, il convient de créer des taxinomies où la relation de base est une relation de subsomption dans laquelle, pour simplifier, les termes doivent être reliés par l'expression « est un » ou « *is a* ». Il faut donc trouver de grandes catégories de départ dans lesquelles chaque concept peut être susceptible de trouver sa place en créant de plus en plus de catégories à l'intérieur de la première.

⁵³ Koha est un SIGB (Système intégré de gestion de bibliothèque) sous licence libre et ouverte.

Prenons un exemple avec Ustensile⁵⁴ :

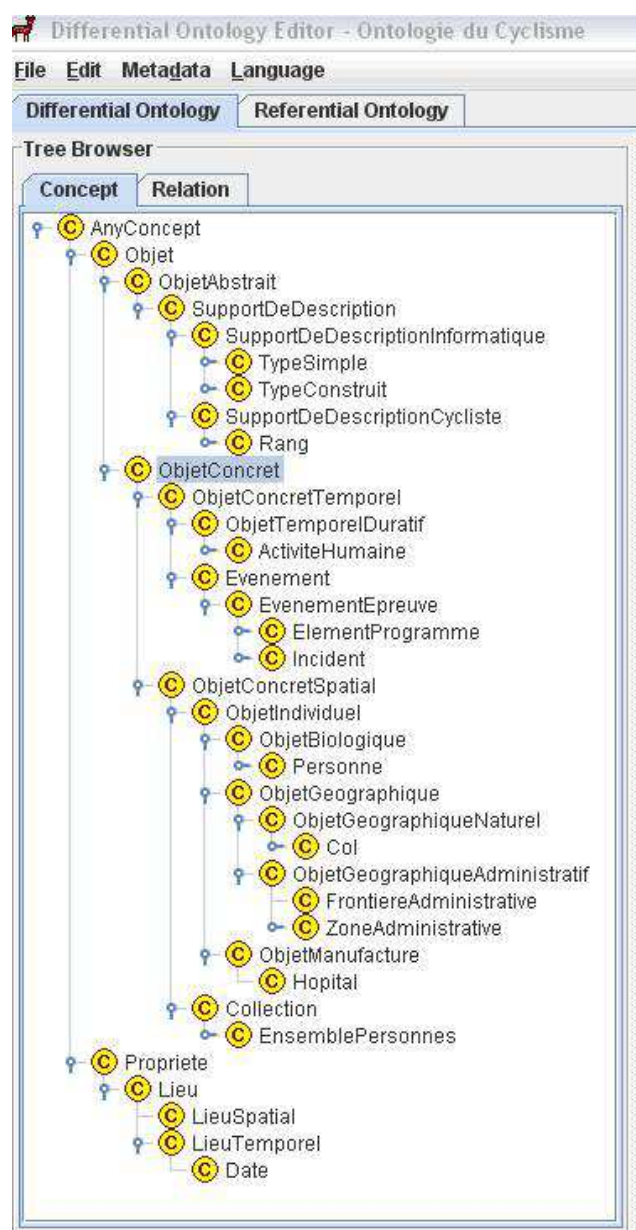
Ustensile est le nom de la première catégorie. À l'intérieur de celle-ci il faut ranger des ustensiles pour que la relation de subsomption fonctionne. On peut y mettre Couteau, Cuiller, Broche à rôtir, Hachoir, Louche et Passoire. On sait que dans le thésaurus il y a deux types de couteau : Couteau à dos et Federmesser, ils iront donc dans la rubrique Couteau car on peut dire : un Couteau à dos « est un » Couteau et un Federmesser « est un » Couteau. Nous pouvons faire de même en remontant la chaîne : un Couteau « est un » Ustensile.



25. Représentation schématique de la taxinomie d'Ustensile

On remarque avec cet exemple que les choses bloquent au niveau d'Ustensile, on comprend qu'il faut aller plus haut et trouver d'autres catégories pour englober Ustensile et, nous supposons, qu'il faudra faire de même – c'est-à-dire, trouver d'autres catégories – de manière à arriver au sommet de la taxinomie, soit le concept le plus général : Any Concept, et ainsi représenter tout le micro-thésaurus PACTOLS-Sujets. En outre, il est essentiel de définir chaque concept qui entre dans la taxinomie de manière à lever toute ambiguïté relative aux concepts utilisés. Ainsi il faut préciser ce qu'est un ustensile pour savoir ce qui peut être mis à l'intérieur de cette rubrique et ainsi de suite. C'est ce que propose de faire DOE, comme nous allons le voir rapidement. DOE propose également à travers l'ontologie du cyclisme disponible dans le logiciel, une hiérarchie des concepts de haut niveau, c'est-à-dire des concepts structurant l'ontologie, les premiers à partir desquels tous les autres sont catégorisés.

⁵⁴ Exemple à base de termes du thésaurus



26. Structure fondamentale de l'ontologie du cyclisme

C'est donc en m'inspirant de cette hiérarchie proposée que j'ai envisagé 4 rubriques pour classer le vocabulaire du thésaurus : Objets, Propriétés, Activités, Je ne sais pas⁵⁵.

OBJETS	PROPRIETES	ACTIVITES	Je ne sais pas
<i>Personnes, faune, flore, outils, abstraits, géo</i>	<i>Couleur, lieu temporel, physiques, matière (ce qui constitue l'objet)</i>	Humaines, techniques,	
Abattoir sacré	Noir	Accouchement	Trésor public : = argent ? = lieu ?

⁵⁵ Cette dernière rubrique est personnelle et ne figure pas dans l'ontologie du cyclisme. Elle a pour but éventuel de réajuster les catégories en fonction du vocabulaire qui la compose.

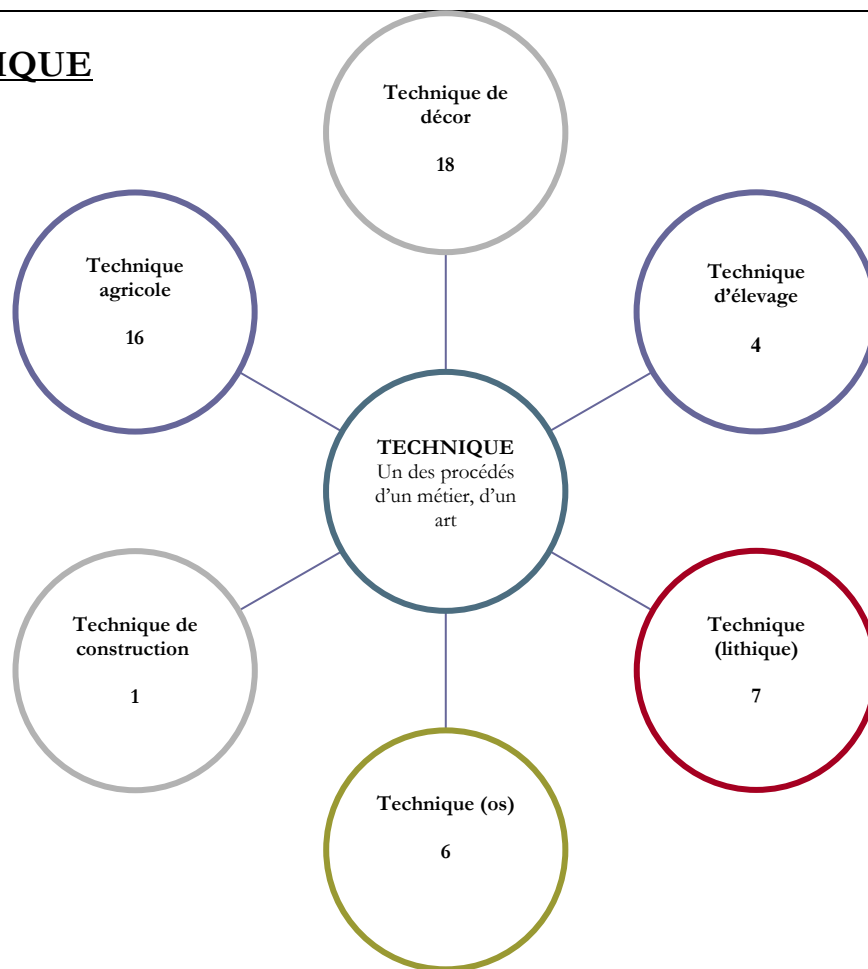
Abbatiale (église principale d'une abbaye) rob.	Rouge	Accroissement	Propriété privée : = terre ? Notion ?
Abri sous roche (EP Abri, Abri rupestre)	Blanc	Acculturation / faits de civilisation	Amas de débitage
Abeille	Fine	Accusation	Ambactitude (activités) DEF ?
Abreuvoir	Accadien	Vente d'esclave achat/vente	Abstraction divinisée
Abbaye	Accusatif / cas_grammaire	Acoustique (science)	Alimentation
Abside (partie arrondie en hémicycle de certaines églises)	Adolescence (temporelle)	ACP (analyse en composantes principales)	*Analyse documentaire, voir si ce n'est pas un outil documentaire. Est dans activités.
Absolu (notion de philosophie) 0 notice	Agate (matériaux, pierre)	Acquisition de ressource naturelle	
Académie ancienne (courant philosophique) Académie moyenne Académie nouvelle Académie	Age d'or	Acquisition des données	Amphictyonie (assemblée) SUJET abstrait ou réunion de personnes : collectif ?
	Age pour travailler	Acrobatie	
	Ambre	Impact humain (EP action de l'homme)	Ampoule à eulogie (objet, céramique) = outil ou pas ?
	ADN	Activités commerciales	*Anatomie en tant que discipline ou ensemble du corps ? est dans Activités

27. Extrait du tableau général de classement des termes du thésaurus par catégorisation.

Je n'ai pas classé tout le vocabulaire du thésaurus, je ne l'ai fait que pour 1250 termes. Je me suis arrêtée après avoir bien compris d'une part quels types de termes composaient le thésaurus et d'autre part, après avoir bien défini ce que devait contenir les grandes rubriques.

Par ailleurs, l'ordonnancement de certains de ces termes dans des graphiques est venu compléter ma compréhension du thésaurus. J'ai représenté sous forme de schémas certaines des taxinomies déjà présentes dans le micro-thésaurus PACTOLS-Sujets.

TECHNIQUE



Technique (os) : chaîne opératoire os ; débitage de l'os ; façonnage de l'os ; forage (os) ; tournage (os) ; traitement thermique.

Technique de construction : levage.

Technique (lithique) : chaîne opératoire ; débitage, ébauche, remontage, taille, technique (Mettre méthode) Levallois.

Technique d'élevage : domestication ; dressage, pâturage, transhumance.

Technique de décoration : damasquinure ; décor excisé ; décor incisé ; décor peint ; dorure, filigrane ; granulation ; graphitage ; griffade ; impastation ; impression ; incrustation ; perforation ; piquetage ; polissage ; repoussé ; champlévé ; ciselure.

Technique agricole : amendement ; assolement ; culture en terrasse ; culture extensive ; culture fruitière ; culture intensive ; défrichement ; piquetage ; domestication des plantes ; érussage ; essartage ; horticulture ; irrigation ; jachère ; riziculture ; sylviculture.

28. Représentation de l'ordonnement graphique du concept Technique

Le nombre inscrit sous le nom de la technique indique le nombre de techniques correspondant. Ces dernières sont ensuite énumérées sous le graphe.

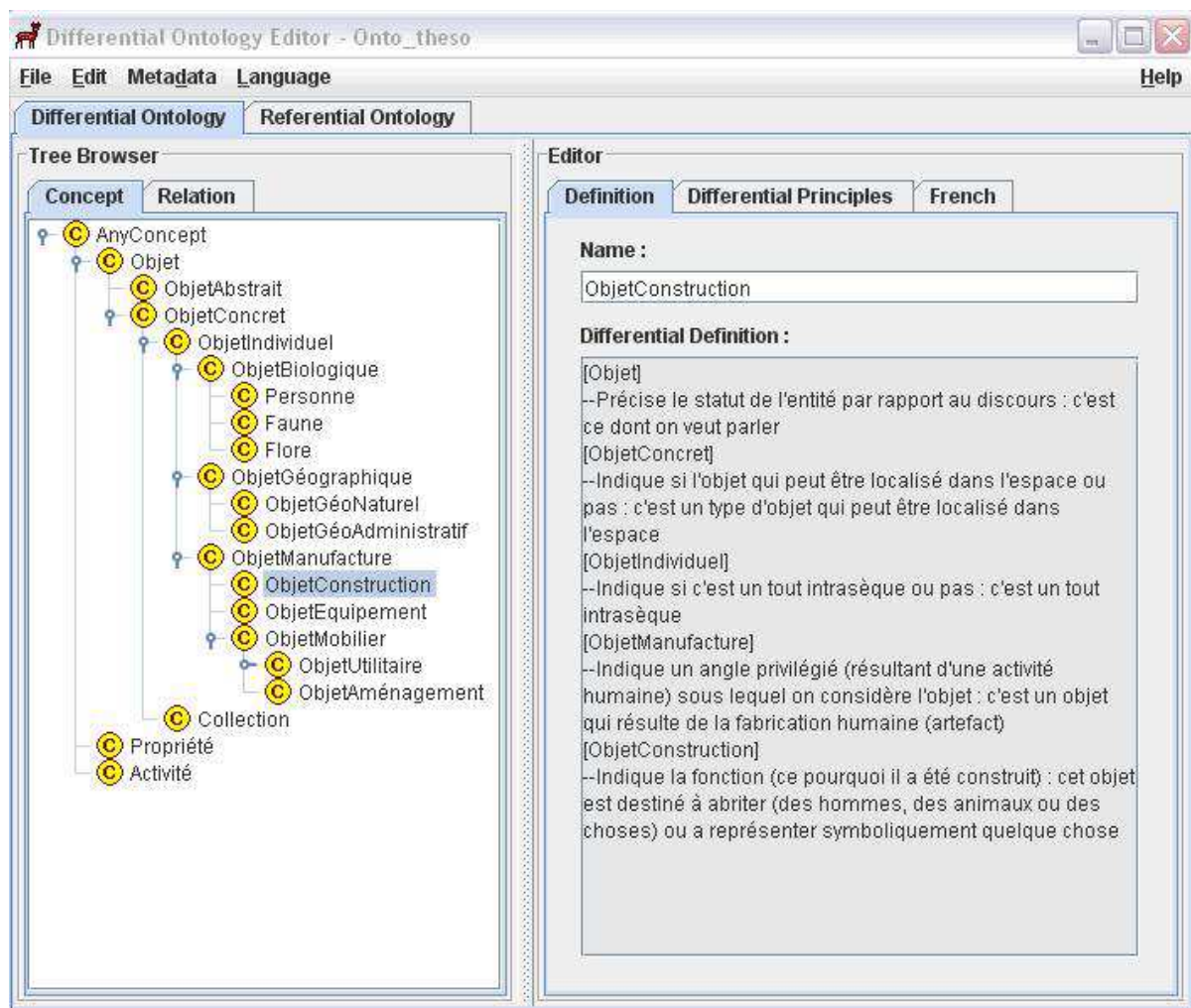
Ainsi, par les différents procédés qui nous venons d'évoquer, je suis parvenue à m'approprier le micro-thésaurus PACTOLS-Sujets. Il reste maintenant à voir comment il a été

saisi dans le logiciel DOE, objectif de la première étape de création de l'ontologie s'il faut le rappeler.

Le logiciel invite donc à créer des taxinomies pour les concepts et les relations que nous voulons voir figurer dans l'ontologie selon la relation essentielle de subsomption et selon le principe de l'ontologie différentielle où l'on doit définir chaque concept et relation en fonction de ses similarités et différences avec ses frères et son père. Nous pouvons dire dès à présent que la structuration de la hiérarchie des relations est celle qui m'a pris le moins de temps et sur laquelle j'ai le moins réfléchi, sachant d'avance que le logiciel DOE n'offrait que des fonctionnalités relativement limitées quant à leur représentation et que la seconde étape – le passage du modèle créé dans DOE dans Protégé – me permettrait de développer davantage ce point essentiel de la création de l'ontologie. En revanche, le choix de la hiérarchie des concepts s'est révélé être un véritable « casse-tête ». En effet, les possibilités de représentation sont nombreuses et l'enjeu est des plus importants puisque de la structuration initiale dépend toute l'ontologie. Trois grands principes de conception, en relation avec l'ontologie de haut niveau du cyclisme proposée dans DOE⁵⁶ et la structure actuelle du micro-thésaurus PACTOLS-Sujets, ont fait l'objet d'interrogations de ma part :

- **dans un premier temps**, voulant reprendre la structure de l'ontologie du cyclisme mais ne parvenant pas à me faire à l'idée – peu commune – qu'une activité était un Objet Concret Temporel, j'ai envisagé la possibilité de partir de trois grandes rubriques : Objet, Propriété et Activité

⁵⁶ Figure 26, *Structure fondamentale de l'ontologie du cyclisme*, p. 92



29. Copie d'écran du logiciel DOE représentant la hiérarchie des concepts selon trois grandes rubriques : Objet, Propriété, Activité

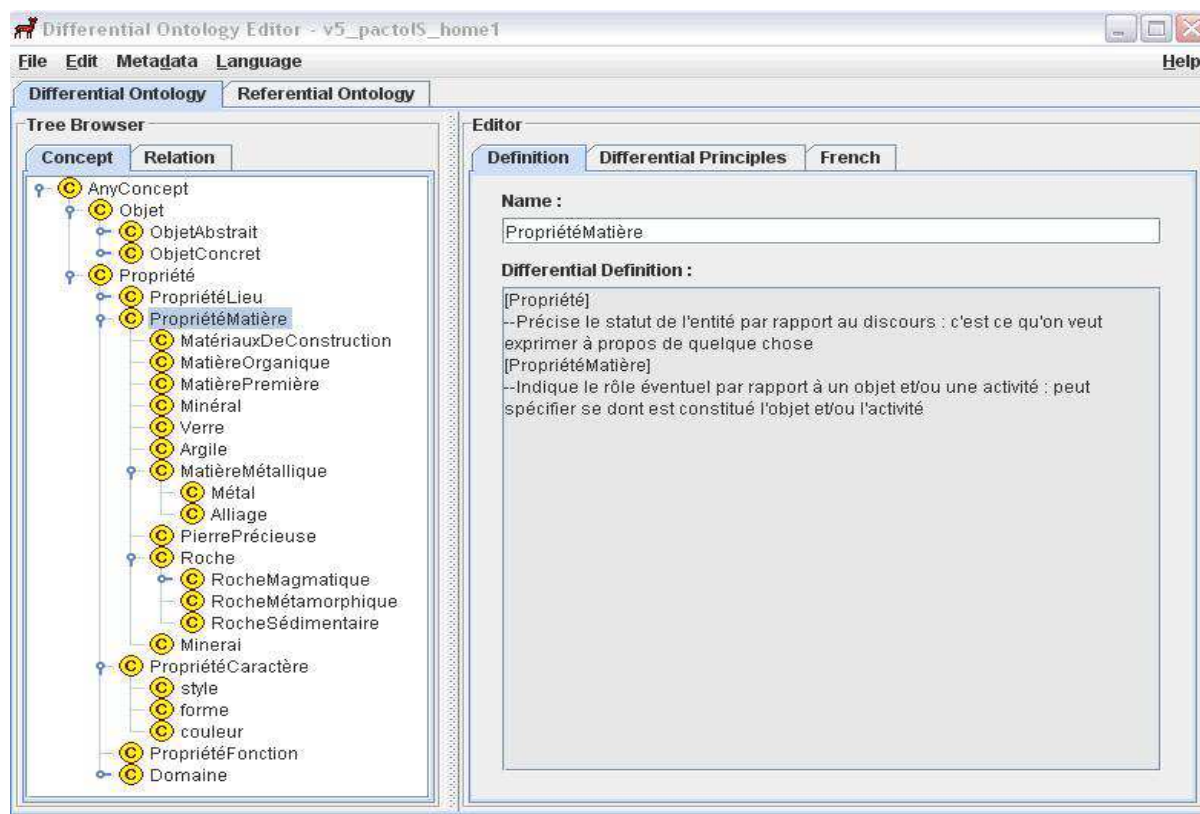
Cette solution a été plus ou moins abandonnée, ou du moins modifiée, après avoir vraiment compris le principe de structuration initial proposé par les concepteurs de l'ontologie du cyclisme. Les activités ont finalement été incluses dans Objet.

- Ensuite, l'éventualité de mettre les activités dans Objet ne me satisfaisant toujours pas, mais sentant que les choses ne fonctionnaient pas ainsi, j'ai émis l'hypothèse de garder la structure du thésaurus en modifiant légèrement la structure de base de manière à réduire le nombre de concepts de premier niveau. Il fallait aussi trancher sur une place prépondérante pour les branches du thésaurus qui avaient une poly-hiérarchie, comme par exemple les Bâtiments militaires qui dépendaient de deux hyperonymes : Défense et Architecture.

Architecture	Intégration des divers bâtiments et d'une partie de MORT
Art	Intégration de CERAMOLOGIE dans Arts décoratifs
Céramologie	L'intégrer à ART
Défense	Intégrer les bâtiments militaires dans ARCHITECTURE
Droit	Intégration d'une partie de MORT
Flore	L'intégrer dans MATERIAUX
Matériaux	Intégration de FLORE
Mort	L'intégrer dans les différentes parties : ARCHITECTURE, RELIGION, DROIT et PHILOSOPHIE
Philosophie	Intégration d'une partie de MORT
Religion	Intégration d'une partie de MORT

30. Tableau résumant les changements qui pourraient être effectués dans la structure du micro-thésaurus PACTOLS-Sujets dans le cadre de sa transformation en ontologie.

- Enfin, une autre possibilité, intermédiaire pourrions-nous dire, a été envisagée. Elle aurait consisté à faire un mélange entre les deux premières solutions en maintenant la rubrique Propriété extraite de la première solution, afin d'alléger les autres rubriques issues du thésaurus remanié. La partie Matériaux aurait été intégrée entièrement dans Propriété.



31. Copie d'écran du logiciel DOE représentant la solution intermédiaire envisagée pour transformer le thésaurus en ontologie.

Ainsi, c'est avec ces trois propositions de structuration que je suis allée à un rendez-vous de suivi de stage avec Monsieur Menon pour lui demander conseil. Après avoir essayé de refaire un classement des concepts du thésaurus selon la méthode proposée par le logiciel DOE, j'ai perçu que cette approche, que nous pourrions qualifier de philosophique mais qui est également caractéristique de certains langages informatiques, ne faisait pas l'unanimité au sein du service, et notamment auprès de Madame Lequeux, responsable du réseau FRANTIQ et créatrice des PACTOLS. Le maintien de la structure du thésaurus semblait être primordial pour FRANTIQ et Madame Lequeux étant donné le succès actuel des PACTOLS au sein de la communauté des utilisateurs-indexeurs qui le préfère d'ailleurs au thésaurus du ministère de la culture : Patriarche. En outre, pour Madame Lequeux, initiatrice du projet, cela relève également d'une question de pragmatisme *versus* abstraction linguistique et de doutes sur l'utilisation future d'un outil trop loin des usagers – documentalistes et chercheurs. Néanmoins, malgré les difficultés sous-jacentes à la déstructuration complète du micro-thésaurus, j'étais consciente que garder sa structure n'était pas le but de l'exercice de transformation attendu. C'est donc forte du soutien de Monsieur Menon qui m'a confortée dans mon entreprise de remaniement total de la hiérarchie du PACTOLS-

Sujets, que mon choix de modélisation s'est fixé sur la répartition du vocabulaire selon deux grandes catégories : Objet et Propriété.

Description de la modélisation de l'ontologie obtenue à l'aide du logiciel DOE

Cette modélisation a été, en toute logique, transposée dans le logiciel DOE et a donné lieu à un fichier XML qui va maintenant faire l'objet d'une description. Ce fichier est l'aboutissement d'un mois et demi de travail environ et de très nombreux essais de structuration. Même si pour des raisons de temps limité il ne contient pas la totalité du micro-thésaurus PACTOLS-Sujets, mais uniquement certaines branches et concepts isolés de ce dernier – soit environ 780 termes – il est représentatif de la transformation du micro-thésaurus PACTOLS-Sujets en ontologie, objectif de ma mission de stage.

DOE, rappelons-le brièvement, est un éditeur d'ontologies qui vise à rationaliser leur conception et plus particulièrement à organiser les concepts entre eux et à expliciter le sens attribué aux termes mobilisés en proposant une méthodologie fondée sur la sémantique différentielle. Dans cette optique, il donne notamment lieu à la création de deux ontologies : l'une différentielle et l'autre référentielle.

Pour décrire la modélisation obtenue et le travail effectué dans DOE, nous aborderons la description du fichier obtenu sous des angles différents. Dans un premier temps nous prendrons connaissance de la structuration de l'ontologie de haut niveau, puis nous descendrons peu à peu au cœur de l'ontologie de domaine. Par ailleurs, nous exposerons clairement les ajouts qui ont été effectués lors du passage de l'ontologie différentielle à l'ontologie référentielle. Enfin, nous nous intéresserons à la taxinomie des relations qui, comme annoncé auparavant, a fait l'objet d'une attention relativement limitée étant entendu que le logiciel DOE n'en permettait pas une exploitation minutieuse.

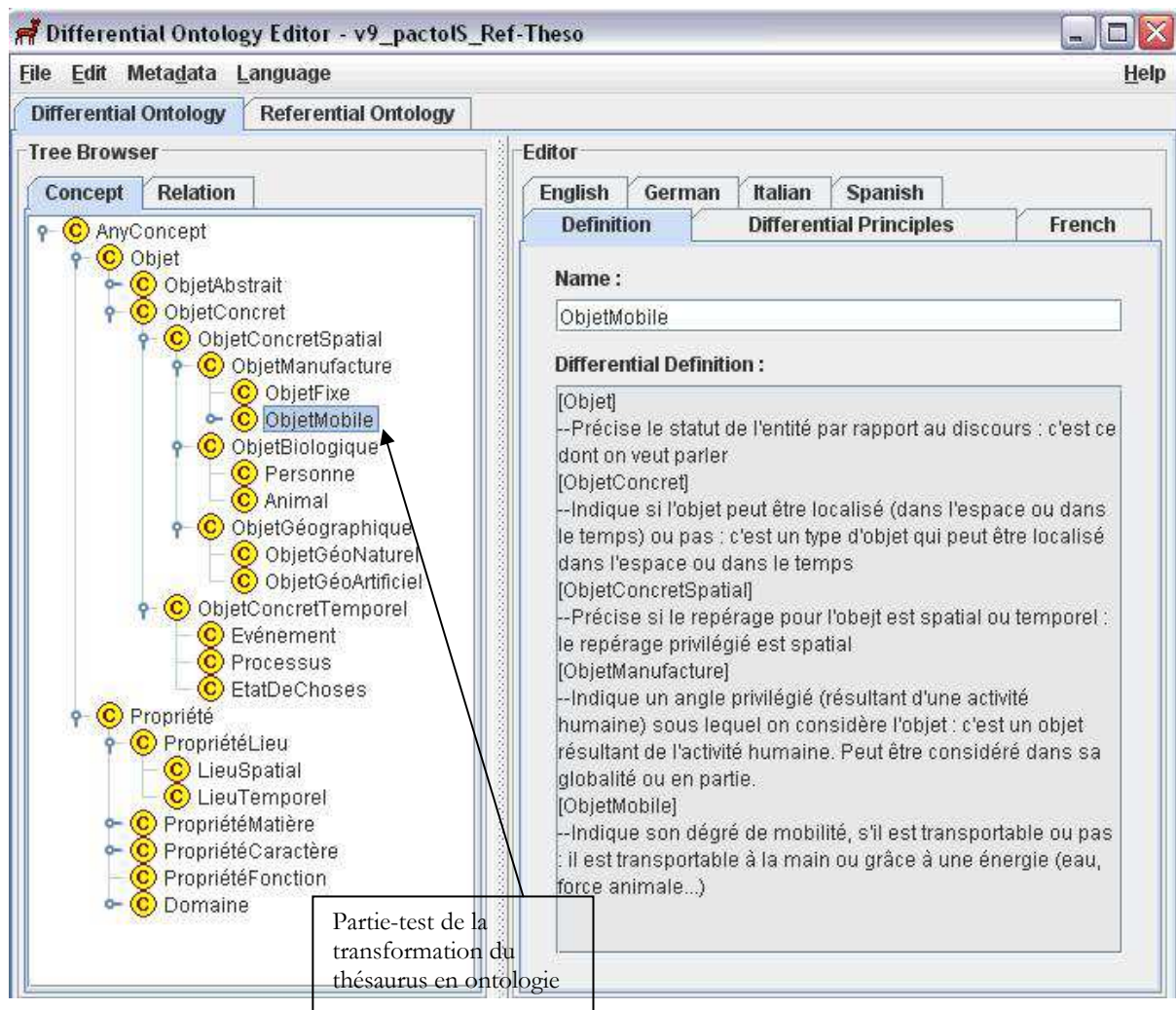
Ontologie différentielle

Si le micro-thésaurus PACTOLS-Sujets s'est révélé être un excellent réservoir de termes pour créer l'ontologie, et si des concepts majeurs d'assez haut niveau sont ressortis pour former des sous-arbres dans la taxinomie comme nous allons le voir rapidement, il s'est posé le problème de la création de l'ontologie de haut niveau. Cette dernière est le sommet de l'ontologie et, pour

respecter la méthodologie mise en œuvre dans DOE, elle se doit de répondre aux principes différentiels. Par conséquent, les concepts la constituant sont sensés se trouver sur le même axe sémantique de manière à être, pour le premier niveau, tous frères et fils du même concept le plus général : Any concept. Or comment faire appartenir à la même fratrie des notions telles que : Aliment, Outil, Site archéologique ou encore Méthodologie et Éducation ? En d'autres termes, comment relier l'ontologie de domaine issue du thésaurus au concept le plus général, Any Concept ? La solution se trouve dans l'introduction de nouveaux concepts permettant d'effectuer cette jonction entre le sommet de l'ontologie et les concepts propres à l'archéologie et aux sciences de l'Antiquité. Mais quels concepts choisir ? La réponse est d'autant plus difficile qu'elle relève d'un champ de connaissances très vaste et complexe : celui de la philosophie. En effet, on revient finalement aux fondements de l'Ontologie relevant de la *Science de l'être en tant qu'être* et de la *théorie ou conception du réel* puisque nous devons créer une ontologie générale ayant comme sujet l'étude des catégories des choses qui existent dans le monde, soient les concepts de haute abstraction tels que les entités, les événements, les états, les processus, les actions, le temps, l'espace, les relations ou encore les propriétés. Cela dit, il existe des recours pour éviter de se perdre dans les méandres de la métaphysique ou d'autres branches de la philosophie. Ainsi, quelques tentatives comme UpperCyc et l'initiative du SUO⁵⁷ sous l'égide de l'IEEE⁵⁸ ont pour but de standardiser les concepts d'une ontologie de haut niveau afin que celle-ci puisse resservir à tous ceux qui en ont besoin. D'ailleurs ce sont ces produits qui ont inspiré les créateurs de l'ontologie du cyclisme pour compléter leur ontologie de domaine et sur laquelle je me suis moi-même appuyée pour obtenir le résultat suivant :

⁵⁷ SUO signifie Standard Upper Ontology. <http://suo.ieee.org/>. Une ontologie appelée SUMO (Suggested Upper Merged Ontology) se propose de servir de structure de haut niveau pour toute sorte d'ontologie.

⁵⁸ IEEE signifiait à l'origine Institute of Electrical and Electronics Engineers, mais son champ d'action s'est aujourd'hui étendu à toutes sortes de disciplines et s'intéresse à l'avancée des technologies en général.



32. Copie d'écran de l'ontologie de haut niveau structurant l'ontologie de domaine issue du micro-thésaurus PACTOLS-Sujets

Cette copie d'écran nous permet de constater, comme indiqué dans les propos précédents, que le vocabulaire se répartit selon deux grandes notions : les Objets et les Propriétés. C'est de celles-ci, filles du concept le plus général : Any Concept, que toute l'ontologie va découler. En effet, peu à peu, chaque notion mère se définit par ses enfants pour parvenir aux environs du quatrième niveau au début de l'ontologie de domaine. Ce quatrième niveau débute notamment à la notion Objet Mobile qui est la partie de l'ontologie que j'ai choisie de développer afin de tester le processus de transformation du thésaurus en ontologie.

Par ailleurs, nous pouvons remarquer que certaines branches comme Objet Concret Temporel ou Objet Abstrait sont peu ou pas développées. La raison est simple : par souci d'économiser le temps qui m'était imparti, j'ai opté pour la création d'une partie-test et laissé de côté les autres branches, ne développant que celles qui étaient nécessaires à la structuration de l'ontologie de haut niveau et permettant une cohérence minimale au travail effectué.

Le résultat obtenu peut évidemment être critiqué, mais c'est le cas pour toute ontologie car leur création relève de la pensée humaine qui, même si elle se doit d'être la plus neutre possible, est au final le reflet de la pensée du créateur. Cela dit, il est évident qu'un philosophe aurait été plus à même de modéliser cette ontologie générale. Néanmoins, les concepts retenus grâce à la méthodologie sous-jacente au logiciel DOE ont au moins le mérite de pouvoir être expliqués par l'énonciation des principes différentiels, comme nous pouvons le voir. Les figures suivantes représentent les définitions différentielles qui ont permis de descendre jusqu'à la notion d'Objet Mobile⁵⁹.

[Objet]

--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler

[Propriété]

--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

33. Définitions différentielles des deux premiers concepts de haut niveau structurant l'ontologie

L'ontologie sera donc divisée en objets, concept extrêmement large, et en propriétés qui pourront en quelque sorte qualifier les objets.

[Objet]

--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler

[ObjetConcret]

--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps

[ObjetConcretSpatial]

--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial

[ObjetManufacture]

--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet résultant de l'activité humaine. Peut être considéré dans sa globalité ou en partie.

[ObjetMobile]

--Indique son degré de mobilité, s'il est transportable ou pas : il est transportable à la main ou grâce à une énergie (eau, force animale...).

34. Définition différentielle d'Objet Mobile

Ces définitions, qu'elles soient destinées à définir les concepts de haut niveau ou les concepts spécifiques au domaine de connaissance, assurent le maintien d'une même cohérence, logique, tout au long de la construction de l'ontologie.

⁵⁹ Dans le souci de ne pas surcharger le corps du mémoire, ne sont représentées ici que les définitions structurantes et celle d'Objet Mobile. Pour prendre connaissance des autres définitions différentielles sous-jacentes à la copie d'écran, aller à l'annexe VI, *Arbre des concepts de l'ontologie différentielle sur l'archéologie et les sciences de l'Antiquité?*

Ainsi et comme nous allons pouvoir le constater, pour descendre plus bas dans l'ontologie et réellement utiliser les termes du micro-thésaurus PACTOLS-Sujets, il faut que les concepts candidats à la notion d'Objet Mobile aient comme première caractéristique d'être transportables à la main, ou grâce à une énergie telle que l'eau ou la force animale.

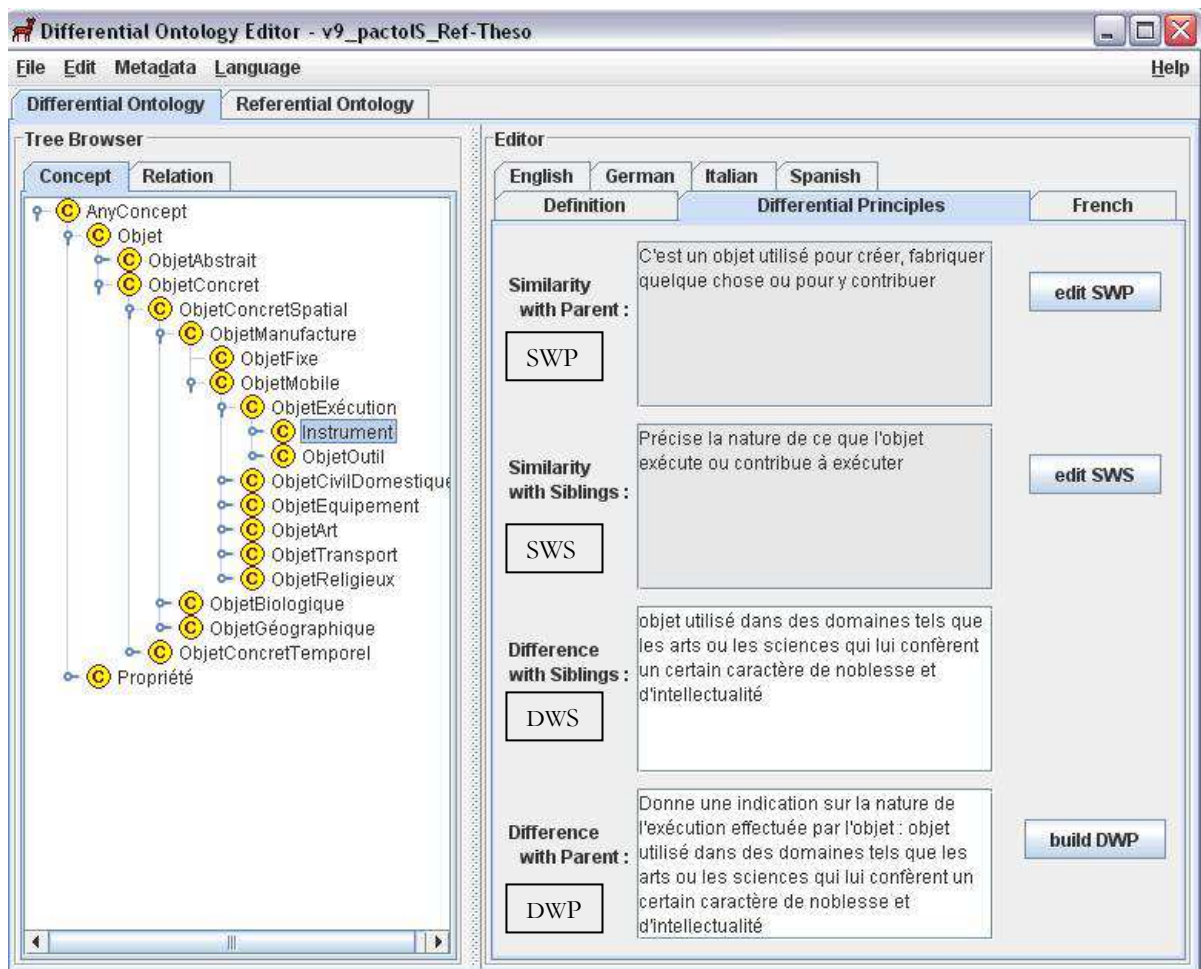
La description de la branche Objet Mobile nous conduit peu à peu dans l'ontologie de domaine. Objet Mobile donne naissance à 6 enfants que sont Objet Exécution, Objet Civil Domestique, Objet Équipement, Objet Art, Objet Transport et Objet Religieux. Eux-mêmes donnent naissance à d'autres concepts et ainsi de suite jusqu'au niveau où l'on souhaite aller. En effet, il faut savoir que dans l'ontologie différentielle, ici décrite, ne doivent figurer que les concepts de base, on parle de Classe⁶⁰, et non pas les extensions, les instances de concepts – valeurs des classes – niveau minimum qui apparaîtra dans l'ontologie référentielle si on le souhaite ou dans la prochaine étape de la construction – le passage dans Protégé. Dans l'ontologie qui nous concerne, le choix a été fait de ne faire figurer que des classes, de manière à pouvoir compléter ultérieurement sans restriction l'ontologie.

Pour enrichir la branche Objet Mobile, je me suis contrainte à remplir pour chaque nouveau terme l'onglet *Differential Principles* du logiciel DOE afin d'en rationaliser la conception. Ainsi, pour valider l'entrée du concept Instrument sous la notion Objet Exécution⁶¹, il m'a fallu remplir les quatre pavés figurant à droite de l'interface du logiciel.

⁶⁰ Par exemple, si l'on considère Fruit comme une classe et Pomme comme un individu de la classe, Golden et Elstar seront des instances.

⁶¹ *Définition différentielle* : [ObjetExécution]

--est considéré d'après ce qu'en fait l'être humain : objet utilisé pour créer, fabriquer quelque chose de concret ou d'abstrait, ou pour y contribuer

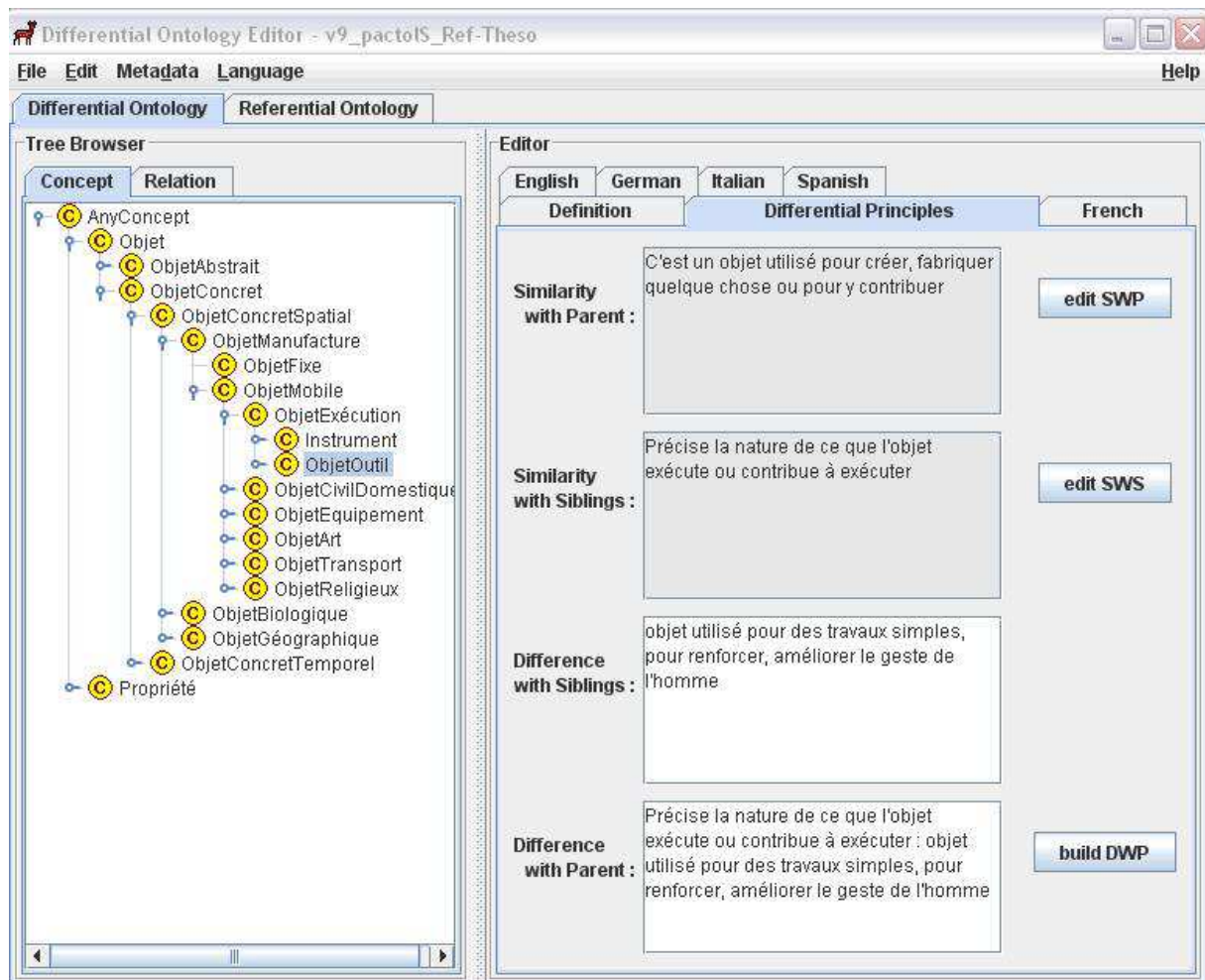


35. Copie d'écran mettant en évidence le principe différentiel attaché à la notion d'instrument.

On observe que j'ai explicité, dans le premier pavé, SWP, pourquoi Instrument hérite des propriétés de la notion qui le subsume, c'est-à-dire de son père, Objet Exécution : *c'est un objet utilisé pour créer, fabriquer quelque chose ou pour y contribuer* ; dans le second pavé, SWS, la propriété – admettant plusieurs valeurs exclusives – qui a permis de caractériser les notions d'une même fratrie, justifiant ainsi le principe de différence avec les frères (DWS) : *précise la nature de ce que l'objet exécute ou contribue à exécuter* ; dans le troisième pavé, DWS, la propriété qui permet de distinguer la notion d'Instrument de sa notion-sœur, Objet Outil : *objet utilisé dans des domaines tels que les arts ou les sciences qui lui confèrent un certain caractère de noblesse et d'intellectualité* ; et enfin, dans le quatrième pavé, DWP, la différence qui permet de distinguer le père du fils, l'Objet Exécution d'Instrument : *donne une indication sur la nature de l'exécution effectuée par l'objet : objet utilisé dans des domaines tels que les arts ou les sciences qui lui confèrent un certain caractère de noblesse et d'intellectualité*.

Ce processus de normalisation est très long, néanmoins, suite à la création de l'ontologie du cyclisme, Raphaël Troncy et Antoine Isaac se sont aperçus de certains faits qui ont permis

d'en automatiser une partie. Pour aider à la compréhension des propos qui vont suivre voici la copie d'écran de la notion sœur d'Instrument, à savoir Objet Outil



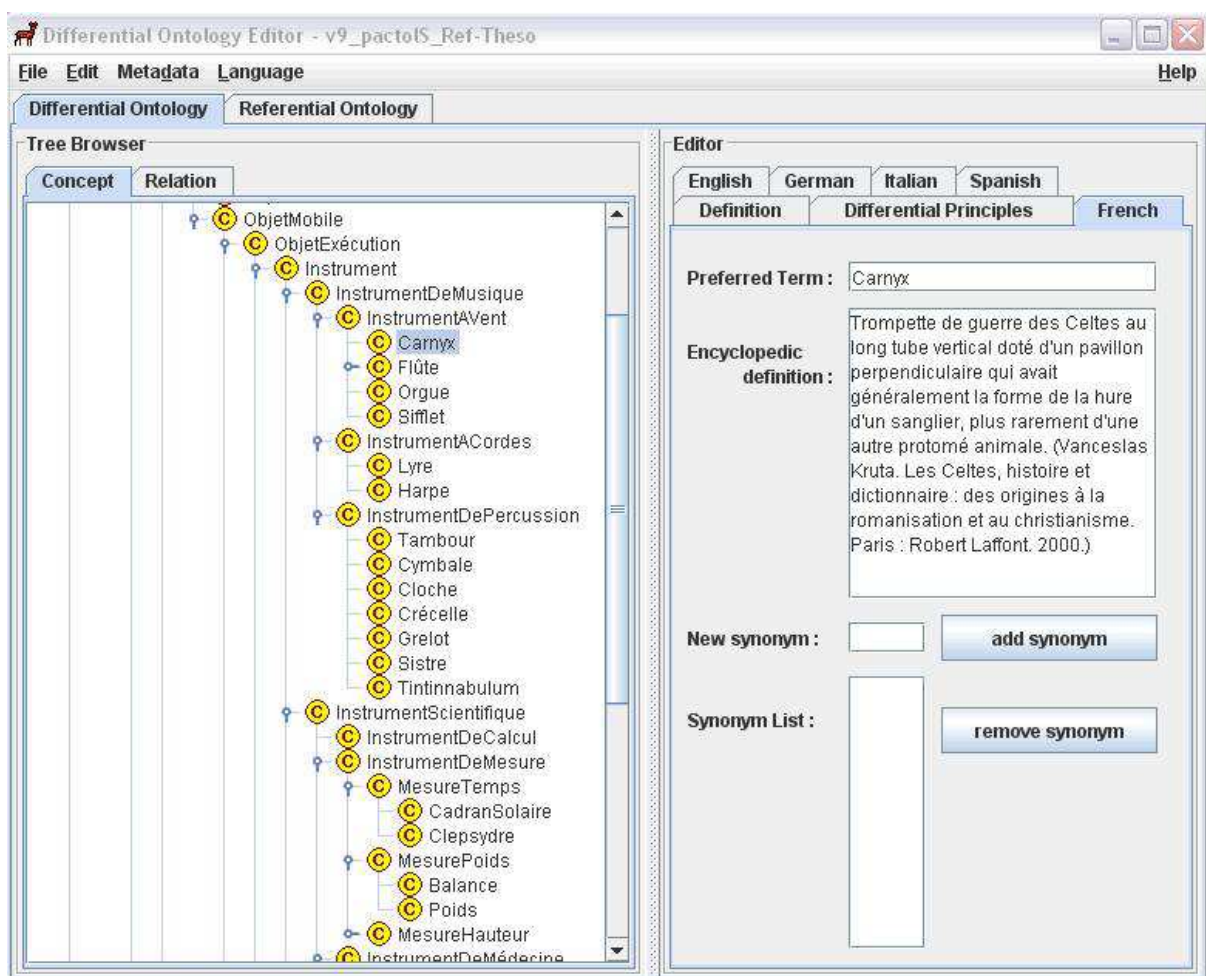
36. Copie d'écran mettant en évidence le principe différentiel attaché à la notion d'Objet Outil, sœur de la notion Instrument

- En premier lieu, il y a souvent partage de SWP et SWS par l'ensemble des concepts membres d'une même fratrie. En effet, toutes les notions d'une même fratrie sont filles d'un même père, et toutes admettent une propriété discriminante déterminée à partir d'un axe sémantique commun : cela est nécessaire à la cohérence de la structure locale de l'ontologie. L'éditeur remplit donc automatiquement ces deux champs pour toute une fratrie dès lors qu'une des notions possède l'information. Néanmoins, toutes possibilités de modifier ces attributs ne sont pas supprimées et nous pouvons reformuler un axe sémantique commun par l'intermédiaire du bouton Edit. Les éventuels changements issus de cette manipulation sont alors répercutés dans la fratrie à laquelle la notion appartient.
- Ensuite, le principe de différence avec le père (DWP) apparaît souvent comme la somme de la spécification de l'axe commun aux frères (SWS) et de la différence avec les frères

(DWS). En effet, on se donne d'abord le moyen de créer la différence, que l'on concrétise ensuite pour achever la définition du concept. Le logiciel propose donc de construire lui-même la différence avec le père (DWP) en concaténant simplement les énoncés des principes DWS et SWS. Cela dit, si cela ne convient pas, il est possible à tout moment de modifier ce que propose l'éditeur en le remplaçant par notre propre création et ce en utilisant le bouton Build DWP.

En accomplissant ce travail de normalisation sémantique – que ce soit pour l'ontologie de haut niveau ou différentielle – on passe alors d'un candidat-terme à une notion dont la signification, ancrée dans le domaine, est invariable et peut donc fonctionner comme une primitive exprimant une connaissance. Ce passage d'un candidat-terme à une notion a été réalisé 785 fois aussi bien pour l'ontologie de haut niveau – beaucoup plus petite – que pour l'ontologie de domaine composée des termes issus du thésaurus.

Pour la plupart des notions, les informations complémentaires telles que traductions et définitions encyclopédiques ont été renseignées (figure ci-dessous).



37. Copie d'écran montrant certaines informations complémentaires apportées à la notion Carnyx

Comme on peut le voir sur la copie d'écran, à la fin de la définition encyclopédique, la référence bibliographique dont est extraite la définition est précisée. Cela a été fait pour chaque définition et la définition en elle-même a fait l'objet d'une validation de la part de Madame Lequeux. Les listes des ouvrages utilisés sont les mêmes que celles utilisées pour l'enrichissement des Notes d'Application du micro-thésaurus PACTOLS-Sujets, elles figurent en annexe⁶². Enfin, il est impossible de faire apparaître l'intégralité de la modélisation de l'ontologie différentielle dans le corps du mémoire car cela demanderait un espace trop important. La hiérarchie est donc visible en annexe⁶³ à travers des copies d'écran.

Ontologie référentielle

Selon la méthodologie employée, les notions figurant dans l'ontologie différentielle et constituant un arbre ontologique sont de nature linguistique. Il reste donc à introduire des concepts obéissants à une sémantique formelle et extensionnelle pour que ceux-ci servent en tant que primitives dans un langage formel de représentation de connaissances. Ces concepts sont ceux figurant dans l'ontologie différentielle d'une part, et d'autre part, ceux qui, schématiquement, ne s'insèrent pas dans les critères de différenciation parce qu'ils ne peuvent être le frère ou le fils de personne. En outre, les notions de l'ontologie différentielle, qui deviennent des concepts en passant dans l'ontologie référentielle, peuvent être précisées, c'est-à-dire que ces concepts peuvent être le fils de plusieurs parents si cela est nécessaire – on parle alors d'héritage multiple.

Ainsi, l'ontologie référentielle que j'ai créée, sur la base non modifiable de l'ontologie différentielle, comprend 10 concepts de plus et trois héritages multiples :

CONCEPT AJOUTE	PARENT
Document Pédagogique	Document
Modèle Réduit	Sculpture
Trésor	Objet Objet
Trésor Militaire	Trésor
Trésor Monétaire	Trésor

⁶² Annexes III, IV, *Liste des ouvrages de références pour les définitions des NA élaborée par Mme Lequeux et Liste complémentaire des ouvrages et sites Internet de référence utilisés pour compléter les NA et les définitions encyclopédiques de l'ontologie*

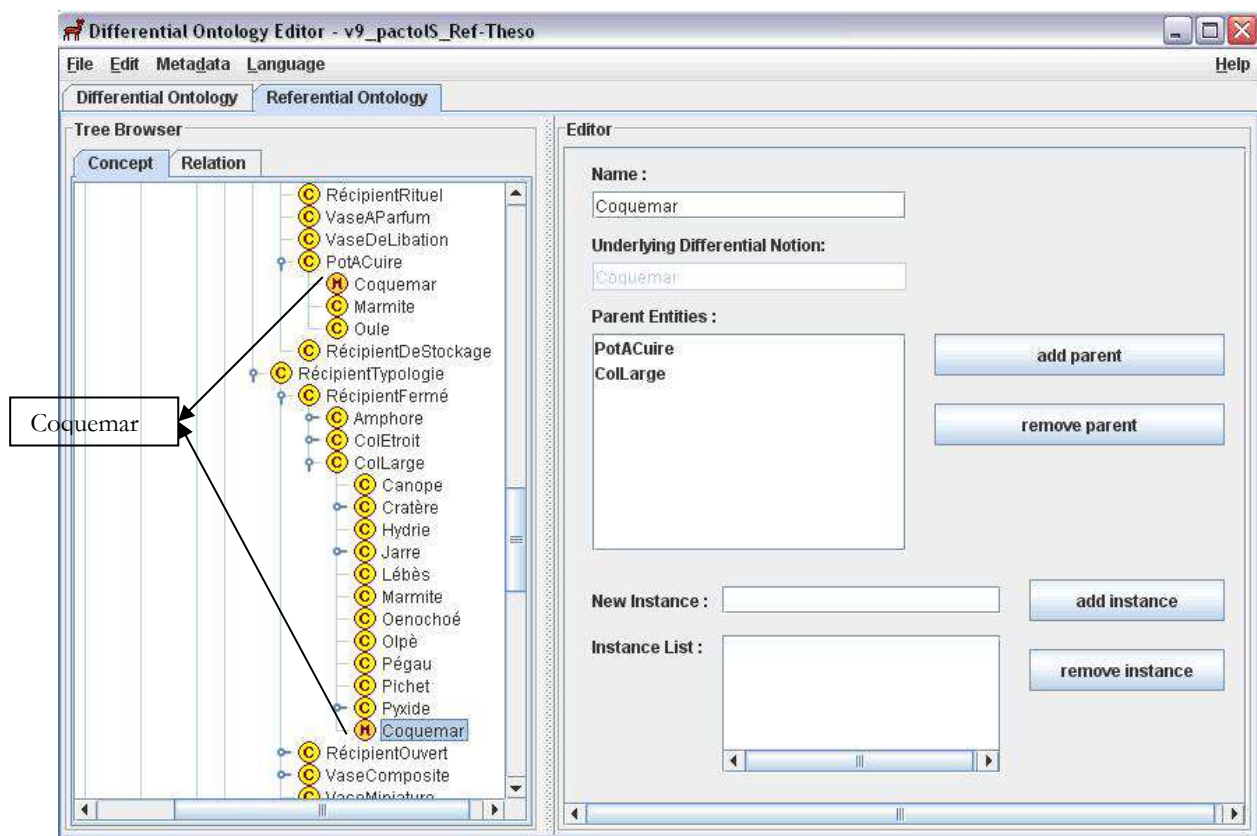
⁶³ Annexe VI, *Arbre des concepts de l'ontologie différentielle sur l'archéologie et les sciences de l'Antiquité*

Trésor de Bijoux	Trésor
Objet en Bois	Objet
Objet Métallique	Objet
Ébauche	Objet Industrie
Matière Première	Propriété Matière
HERITAGE MULTIPLE	PARENTS
Coquemar ⁶⁴	Pot à Cuire
	Col Large
Cuiller	Ustensile
	Récipient Plat-Creux
Cachet d'Oculiste	Instrument de Médecine
	Sceau

38. Tableau révélant les modifications apportées à l'ontologie référentielles

Les concepts ajoutés sont à noter soigneusement à la main car le logiciel, même s'il fournit des statistiques, n'édite pas de liste pour les nouveaux entrants. Nous ne voyons que les écarts de chiffres entre les deux ontologies. En revanche, comme l'illustre la copie d'écran, l'héritage multiple est plus visible car il est signalé par un M dans la hiérarchie.

⁶⁴Coquemar signifie selon le Trésor de la langue française informatisée : Bouilloire de terre ou de métal, munie d'une anse et parfois d'un couvercle, d'un bec et de (trois) pieds.



39. Copie d'écran illustrant le cas particulier de l'héritage multiple avec l'exemple de Coquemar

L'ontologie référentielle est l'ultime étape de restructuration des termes du thésaurus à des fins ontologiques. Elle est composée de concepts qui vont agir comme des primitives formelles qui ne sont plus définies par les principes différentiels, mais qui, en étant reliés aux notions différentielles par leur libellé acquièrent un sens interprétatif. Il s'agit du « produit fini » de la construction de l'ontologie au niveau des connaissances.

Point de vocabulaire : CONCEPT et NOTION

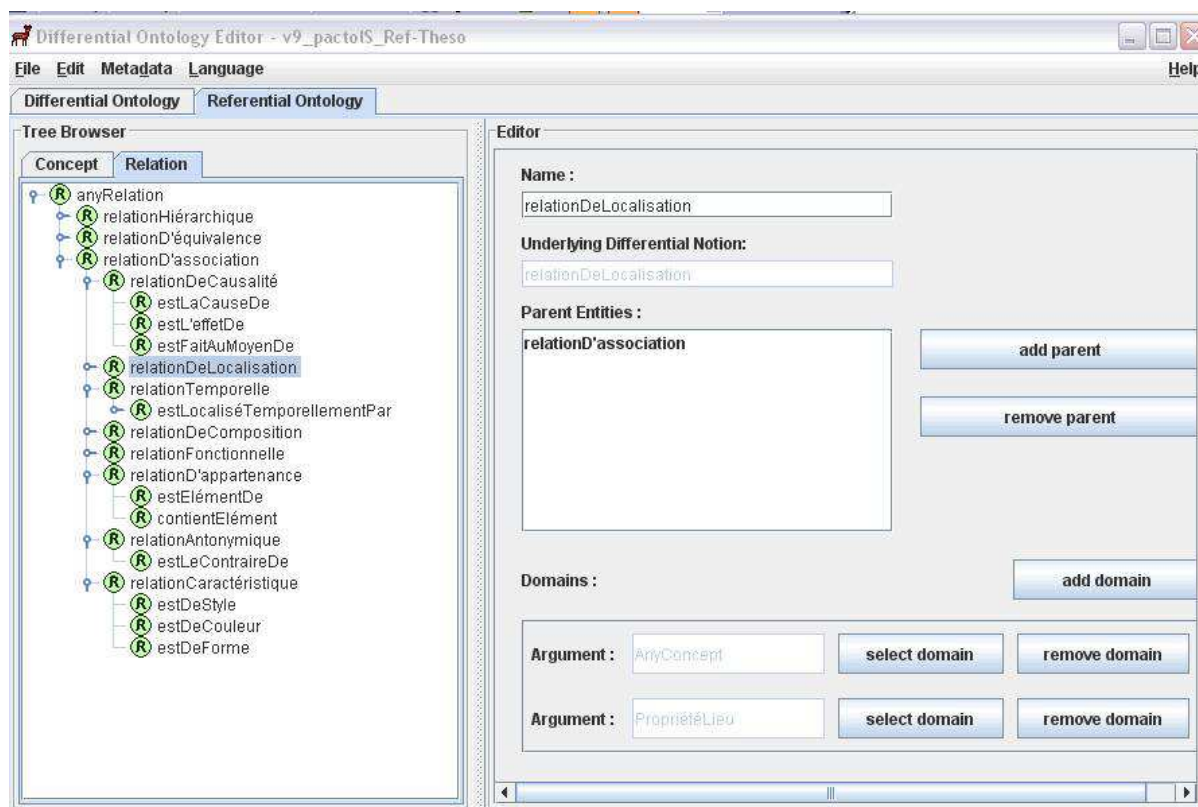
Même si parfois les termes Concept et Notion ont été utilisé indifféremment pour éviter les répétitions, il faut savoir qu'une notion désigne ici un candidat-terme entré dans l'ontologie différentielle, et que les concepts sont les termes présents dans l'ontologie référentielle. Les notions deviennent donc des concepts.

Relations

Enfin, pour terminer la modélisation du micro-thésaurus PACTOLS-Sujets en ontologie à l'aide du logiciel DOE, il nous reste à nous intéresser aux relations. Comme évoqué

précédemment, ces dernières n'ont pas fait l'objet d'autant d'intérêt que la structuration des taxinomies de concepts. Par ailleurs, peu d'informations étaient accessibles sur la gestion des relations dans DOE et les mails écrits aux personnes susceptibles de me renseigner sur le sujet⁶⁵ étaient toujours sans réponses adéquates à la fin de mon stage.

Néanmoins, comme le montre la copie d'écran, j'ai produit, de la même façon que pour les concepts, une hiérarchie des relations :



40. Copie d'écran de la hiérarchie des relations rattachée à l'ontologie référentielle

Après discussions avec Madame Lequeux, je me suis aperçue que l'attente envers la structuration des relations n'était pas des plus complexes. En effet, l'amélioration escomptée dans la création d'une ontologie en ce qui concerne les relations, réside principalement dans la qualification des objets : dire par exemple que tel objet est fait en tel matière ou que telle représentation iconographique orne telle objet. C'est-à-dire, finalement, préciser la nature des relations sous-tendue par les relations dites associatives des thésaurus.

⁶⁵ Les personnes concernées sont : Madame Aussenac-Gilles, partenaire du projet de transformation du thésaurus en ontologie ainsi que Raphaël Troncy et Antoine Isaac, créateurs de DOE. J'ai été en contact avec ces trois personnes, mais elles manquaient de temps pour me renseigner à la période durant laquelle j'effectuais ma mission.

L'arbre des relations est donc le reflet des relations du thésaurus, si ce n'est l'ajout des spécifications des relations associatives. En outre, comme visible sur la copie d'écran ci-dessus on peut voir que j'ai spécifié l'arité des relations en en donnant les domaines. Cependant, je ne sais pas si cela produit un quelconque effet – même si je suppose que oui – car rien ne transparait à l'écran.

Par ailleurs, avec le recul et les différentes informations que je suis parvenue à réunir depuis la fin de mon stage – notamment de la part de Nathalie Aussenac-Gilles – cette hiérarchie est soumise à critique. En effet, les deux premières catégories sont, par exemple, inutiles puisqu'elles sont pour la première sous-entendue dans la création de la hiérarchie des concepts où la relation de subsomption est la base ; et pour la seconde, les relations d'équivalences, elles sont précisées lors de la construction de l'ontologie au moment où l'on saisit les notions.

Ainsi, malgré quelques imperfections qui restent à tout moment modifiables, c'est un fichier cohérent et valable pour la totalité d'une branche de l'ontologie – même si les termes du micro-thésaurus PACTOLS-Sujets, du fait de sa déstructuration, se trouvent disséminés ici et là. J'ai pu fournir à la fin de mon stage et il a fait l'objet d'une présentation en public⁶⁶ le 03 septembre 2008.

En effet, comme je l'ai déjà mentionné, la mission qui m'a été confiée faisait partie d'une réponse à un appel à projets innovants lancé par le TGE ADONIS. Dans ce cadre, Madame Lequeux a été invitée à une Université d'été ayant pour objectif notamment de montrer l'avancement des différents projets retenus lors de cet appel. À cette occasion, elle m'a demandé de présenter le projet de transformation du thésaurus PACTOLS en ontologie, soit le résultat de mon travail. L'exposé a été relativement bref car nous ne disposions, Madame Lequeux et moi, que de 8 minutes pour présenter deux projets. Me concernant, l'exercice a été très formateur car cela m'a permis, d'une part de faire le point sur trois mois de stage en en extrayant le principal et d'autre part, cela a été l'occasion de communiquer le fruit de réflexions et d'expériences tout en le soumettant à la critique d'un public. Enfin, cette présentation et les échanges qui ont suivi ont également permis d'évoquer les différentes suites qui pouvaient être données au projet d'ontologie et plus largement au thésaurus, à plus ou moins court terme.

⁶⁶ Les diapositives réalisées pour cette présentation sont visibles à l'annexe VII, *Diapositives de la présentation des projets FRANTIQ pour l'Université d'été du TGE ADONIS*

2) Perspectives

Les trois mois n'ont pas suffi pour achever le projet de transformation du micro-thésaurus PACTOLS-Sujets en ontologie. Comme nous avons pu le constater dans les propos précédents, la première phase – la déstructuration du thésaurus et la modélisation de l'ontologie – a été effectuée pour quelques 800 termes/concepts. Il reste donc, selon le processus de transformation prévu initialement, la deuxième phase à réaliser : la représentation et le raffinement du modèle obtenu lors de mon stage dans un éditeur dit complet – Protégé. Par ailleurs, suite à l'Université d'été organisée par le TGE ADONIS et à la certitude de pouvoir bénéficier des conseils de Madame Aussenac-Gilles, le passage à la deuxième phase semble s'annoncer sous de meilleurs auspices et d'autres perspectives peuvent aujourd'hui être envisagées en ce qui concerne le devenir du projet et plus largement du thésaurus PACTOLS.

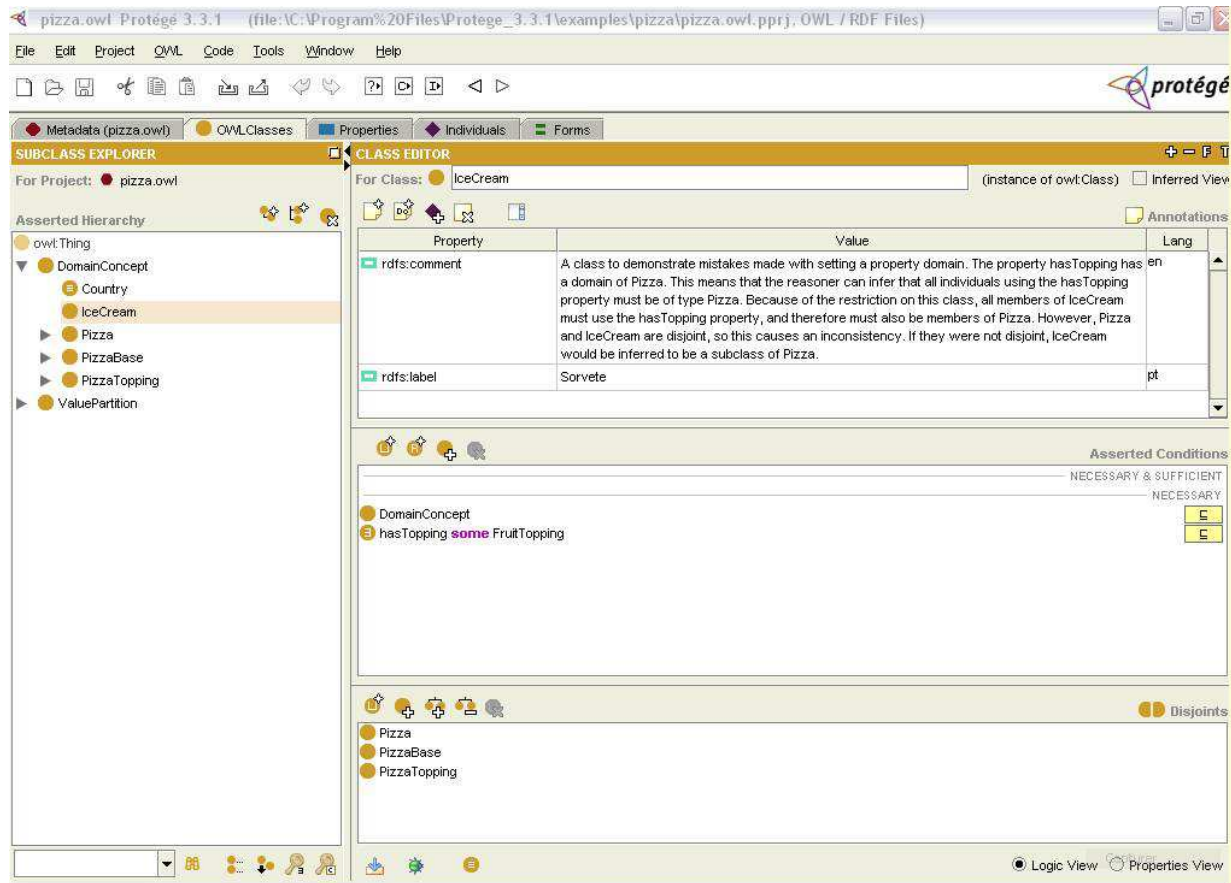
Réalisation de la deuxième phase, la représentation et le raffinement du modèle dans Protégé.

Cette phase est théoriquement consécutive à la phase de modélisation effectuée dans DOE pour parvenir à la création d'une ontologie à partir du micro-thésaurus PACOLS-Sujets. Faute de temps, elle n'a pas été réalisée mais voici quelques informations relatives au fonctionnement et aux capacités de Protégé et le résultat d'un test d'import du fichier obtenu dans DOE dans l'éditeur. Par ailleurs, mon choix s'est porté sur le logiciel Protégé car il s'agit d'un des plus connus, des plus performants et innovants dans le domaine de l'ingénierie ontologique, parce qu'il bénéficie entre autres choses d'une communauté d'utilisateurs très active et dynamique. Enfin, j'ai eu l'occasion d'effectuer quelques tests sur une ancienne version lors de ma formation universitaire et ce n'est donc pas pour moi un outil complètement inconnu.

Comme nous l'avons mentionné dans le chapitre II, Protégé est un éditeur d'ontologies distribué en *Open Source* et gratuit qui a été mis au point au sein du laboratoire *Stanford Medical Informatics* dans le cadre du programme *National Center for Biomedical Ontology*. Protégé propose une interface graphique composée de formulaires qui permettent de créer son ontologie indépendamment de toute contrainte liée à l'utilisation d'un quelconque langage de représentation. En effet, c'est lui-même qui prend automatiquement en charge l'opérationnalisation de l'ontologie en la transposant dans divers langages, comme le propose

d'ailleurs également DOE. Il existe plusieurs versions en cours de Protégé ainsi qu'une extension. J'ai opté pour la version stabilisée 3.3.1 et l'extension OWL pour représenter l'ontologie dans ce langage.

De manière basique, l'éditeur propose une interface composée de 4 onglets permettant la création de l'ontologie : *OWLClasses*, *Properties*, *Individuals* et *Forms* ; et d'un autre : *Metadata*, contenant les métadonnées concernant l'ontologie.



41. Copie d'écran de l'interface de base proposé par Protégé

Les 4 onglets font respectivement référence à l'éditeur de classes, de propriétés, d'instances et aux formulaires de saisies des instances. Dans le langage OWL, rappelons qu'une Classe est un groupe d'individus partageant les mêmes caractéristiques ; qu'une propriété permet de définir des faits ou des relations entre ces classes ; et qu'une instance est un individu d'une classe qui peut prendre les caractéristiques définies par les propriétés.

Mais l'un des avantages de Protégé est de pouvoir attacher à cette base un grand nombre de plugins venant notamment enrichir la représentation de l'ontologie et permettant sa validation

ainsi que son utilisation. En outre, Protégé autorise l'import d'ontologie en OWL et c'est ce qui m'a permis d'effectuer « rapidement » un test à partir du travail effectué dans DOE.

Théoriquement le passage d'un éditeur à l'autre – de DOE à Protégé – ne doit poser aucun problème étant donné que DOE propose un export en OWL et que c'est ce qui est accepté par Protégé. Cependant, je ne suis pas parvenue à réaliser cette opération. Devant ce problème, j'ai pris contact avec plusieurs personnes susceptibles de me venir en aide, et j'ai notamment envoyé un mail à l'un des créateurs de DOE, Monsieur Troncy. Voici une partie de sa réponse :

« [...] En théorie, vous pouvez sauvegarder votre travail et l'exporter dans le format RDFS ou OWL (syntax RDF/XML). Ce fichier *.rdfs ou *.owl est normalement lisible dans Protégé, où vous pourrez compléter l'étape de construction de l'ontologie référentielle (ou encore formelle)

Je dis, "en théorie", car il se peut que vous rencontriez des problèmes

- i) dans l'export de l'ontologie dans le format rdfs ou owl,
- ii) ou que Protégé ait des problèmes à lire le fichier exporté. »

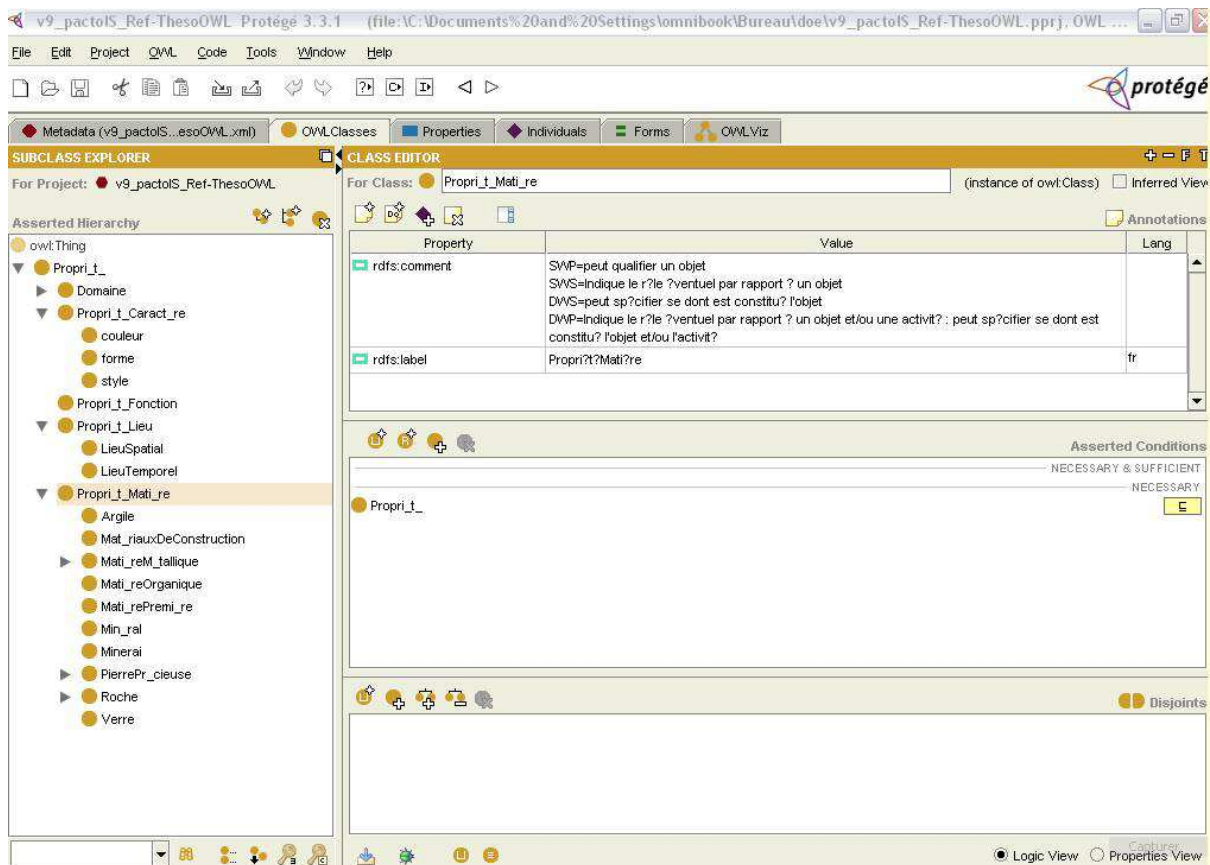
Si vous avez donc des problèmes, n'hésitez pas à nous contacter en envoyant votre ontologie sauvee dans DOE et nous essaierons de voir ce que l'on peut faire. [...] »

Très enthousiaste, j'ai accepté cette proposition et envoyé mon fichier, mais je n'ai toujours pas obtenu de réponse. Néanmoins, le côté positif c'est que j'ai appris d'où pouvait venir le bogue et c'est consciente de cela que j'ai cherché à nouveau comment sortir de cette impasse⁶⁷. La résolution partielle est venue de Monsieur Menon, mon directeur de mémoire et professeur en Ingénierie des connaissances. La manipulation n'a pas été des plus évidentes mais en voici brièvement la description :

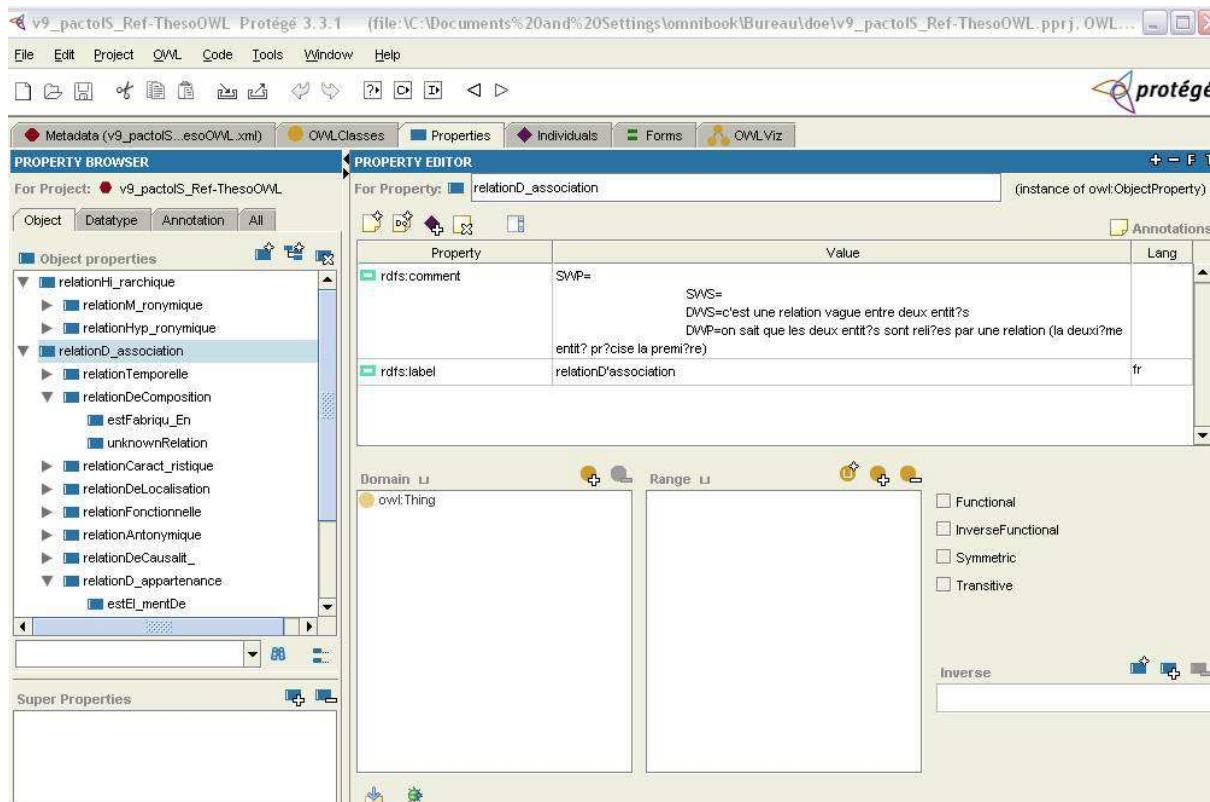
- transfert du fichier XML obtenu dans DOE dans la version 1.5 à la version 2.0 sortie très récemment car nous avons pensé que la nouvelle version était peut être plus apte à gérer les exports : non.
- Application d'une feuille de style permettant la transformation « manuelle » du fichier XML en OWL – feuille disponible dans DOE.
- Importation du fichier OWL dans l'éditeur d'ontologies Protégé.

Néanmoins, malgré la lourdeur et l'incertitude de la démarche nous sommes parvenus à obtenir un résultat dans Protégé :

⁶⁷ Cette recherche de solution s'est étalée sur une vingtaine de jours, elle ne s'est achevée que quelques jours après la fin de mon stage.

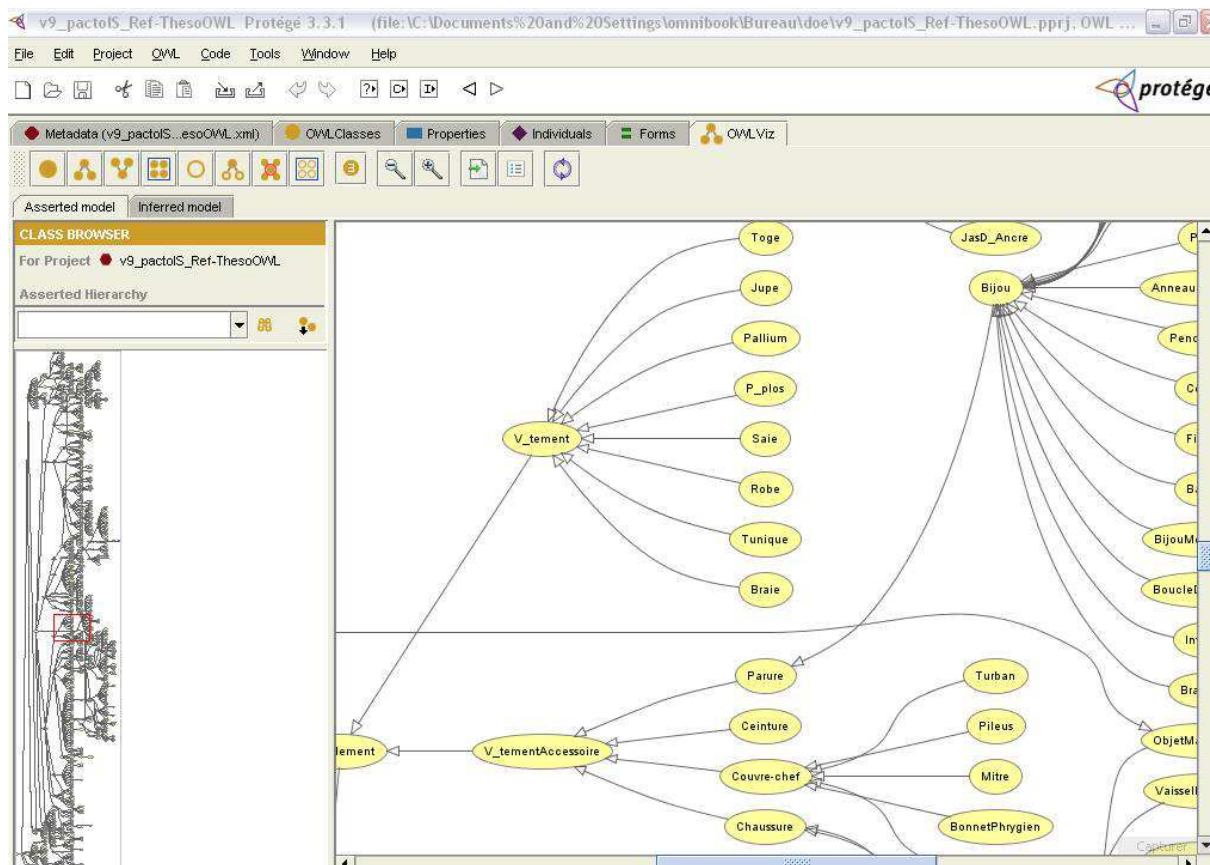


42. Copie d'écran montrant l'import dans Protégé de la modélisation de l'ontologie obtenue dans DOE pour les Classes



43. Copie d'écran montrant l'import dans Protégé de la modélisation de l'ontologie obtenue dans DOE pour les Propriétés

Le résultat ainsi obtenu est difficile à qualifier et à interpréter. En effet, de prime abord seule la seconde partie de la taxinomie des concepts créée dans DOE semble être passée dans l'éditeur alors que celle des relations est visible en totalité. Outre les problèmes d'accents, on retrouve donc parfaitement la branche des Propriétés composée elle-même des branches Domaine, Propriété Caractère, Propriété Fonction, Propriété Lieu, et Propriété Matière et tous leurs enfants, ainsi que toutes les relations. Les relations de DOE correspondent aux *Properties* de Protégé. Pour chacun des cas – *OWLClasses* et *properties* – on peut noter également que les principes différentiels préalablement écrits dans DOE apparaissent sous la forme de commentaires, la langue est spécifiée ainsi que les *Labels*. Il y donc des points positifs dans ce transfert. Mais une question demeure : où est le plus gros de l'ontologie, c'est-à-dire la partie concernant les Objets ? La première réponse à cette interrogation a été de dire qu'il y avait eu un bogue et qu'elle n'était pas passée lors de l'export. Mais récemment, alors que j'observais et testais à nouveau le résultat obtenu je me suis aperçue à travers l'utilisation d'un plugin de visualisation graphique (OWL Viz) que la totalité de la taxinomie des concepts était représentée, soient les quelques 800 concepts saisis dans DOE. On peut d'ailleurs voir sur le côté gauche de la copie d'écran suivante la représentation graphique de la totalité de l'ontologie, et à droite, un extrait de cette représentation. Les flèches présentes sur l'extrait symbolisent la relation de subsomption. On peut par exemple dire qu'une robe est un vêtement, qu'un bijou est une parure et qu'une parure est un accessoire de vêtement.



44. Copie d'écran prouvant la présence de la totalité de la modélisation effectuée dans DOE, dans Protégé

N'ayant pas eu le temps d'étudier en détail le fonctionnement de ce logiciel je ne peux à l'heure actuelle en dire davantage sur le résultat de l'export de la modélisation dans Protégé. En revanche, toutes ces questions sont une excellente entrée en matière pour reprendre le projet et poursuivre la création de l'ontologie selon le processus défini initialement. En effet, le logiciel doit notamment permettre de spécifier les relations envisagées dans DOE et permettre de valider et de tester l'ontologie, c'est-à-dire clore la phase de construction, et ce tout en utilisant le langage OWL. Ceci nous amène à aborder les autres suites qui peuvent être données au projet. Que faire avec cette ontologie, et est-il pertinent, après ce premier bilan, de continuer dans cette direction ?

Autres perspectives

Nous venons d'évoquer la suite directe, logique, à donner au travail que j'ai commencé lors de mon stage ; ce qui devrait se faire si les données contextuelles restent identiques. Mais nous venons de le voir, les circonstances ont évoluées, des perspectives nouvelles mais aussi des

perspectives qui peuvent s'inscrire dans la continuité sont possibles. Et la pierre angulaire de ces différentes possibilités est le TGE ADONIS, car de ce TGE dépend le financement du projet.

Ainsi, ce qui est ressorti principalement de l'Université d'été est la finalité que nous voulions donner à ce projet : une ontologie, oui, mais pour quoi faire ?

Madame Lequeux avait certes une idée sur le sujet dès le commencement du projet : de l'indexation automatique, mais sa concrétisation dépendait de la transformation du thésaurus PACTOLS en ontologie. Il était envisagé dans un premier temps de créer l'ontologie, puis ensuite, de voir ce qu'il était possible de faire avec. Or, la tâche de création s'est révélée très longue et le besoin de résultat se faisait plus fort.

Après l'Université d'été l'objectif reste le même : faire de l'indexation automatique, mais la façon d'y parvenir est susceptible de changer, voire d'être multiple. En effet, deux cas de figure sont envisageables : continuer sur l'idée de faire évoluer les PACTOLS vers une ontologie, ou bien développer des partenariats avec des personnes qui ont le même projet mais selon d'autres approches.

Dans un premier temps, il est donc tout à fait concevable de poursuivre la transformation du micro-thésaurus PACTOLS-Sujets en ontologie en utilisant le logiciel Protégé. D'après Nathalie Aussenac-Gilles, Protégé, dans sa version 4 Bêta propose un plugin permettant de mettre en place une indexation automatique. Par ailleurs, Madame Aussenac-Gilles nous a également suggéré de nous intéresser à SKOS afin de savoir si, dernièrement, le langage n'avait pas évolué de manière à développer davantage les relations entre concepts. En effet, si l'accent est maintenant mis plus sur le résultat que sur la manière d'y arriver, il peut être plus raisonnable de « simplement » transposer les PACTOLS en SKOS qui est un langage dédié aux thésaurus et compatible avec OWL. Cependant, cette solution peut se faire au détriment de la précision des relations recherchées dans la transformation en ontologie si SKOS n'a pas évolué. Toutes ces interrogations devront faire l'objet de recherches avant de poursuivre le projet et c'est à moi que ce travail va revenir car j'ai été chargée de poursuivre la tâche. Néanmoins, si durant mon stage l'accent avait été mis sur la conception de l'ontologie plus que sur ses finalités, cette fois cela sera l'inverse. D'ailleurs, on peut se demander si du point de vue du résultat, de la « rentabilité » du projet, le choix de faire d'abord une ontologie avant d'étudier en détail les possibilités qu'elle pouvait offrir a été judicieux. Cela dit, si l'on oublie le contexte de l'appel à projet, l'expérience est des plus enrichissante et intéressante. En effet, à court terme la décision peut être critiquable, mais à long terme, il me semble qu'elle peut s'avérer pertinente.

Dans un deuxième temps, il est aussi possible, suite aux rencontres faites pendant l'Université d'été du TGE Adonis d'envisager des partenariats avec d'autres projets, personnes ou organismes. Ainsi, deux projets présentés ont particulièrement retenu l'attention de Madame Lequeux. Le premier concerne une indexation semi-automatique à travers l'utilisation de folksonomies mise en place par FABULA.ORG. L'idée consiste à tagger les documents d'après un vocabulaire libre mais il pourrait être possible d'utiliser les thésaurus comme base de termes. Quant au second partenariat, il concerne le projet PIC : Plateforme intégrée de consultation et de manipulation de corpus textuels pour les SHS mené par Claude de Loupy de la société Syllab. Madame Lequeux est en contact avec ce monsieur afin de voir si les PACTOLS pourraient être intégrés à son projet.

Enfin, plus généralement, il est prévu, à plus ou moins grandes échéances, que les PACTOLS relatifs aux lieux fassent l'objet d'un géo-référencement.

Les perspectives ne sont donc pas limitées et ne manquent pas pour faire évoluer et vivre les PACTOLS. L'ontologie demeure un objectif réalisable, mais d'autres ont fait leur apparition et il est important d'en tenir compte. La suite à donner au projet se fera en fonction de trois des critères déjà mentionnés précédemment : le temps, l'argent, et les hommes.

CONCLUSION

Ainsi, ce mémoire touche à sa fin. À travers la mission qui m’a été confiée durant mon stage de fin de master II professionnel GID, au sein de l’UMS 844 Documentation et FRANTIQ du CNRS, nous avons pu suivre pas à pas la conception d’une ontologie à partir d’un thésaurus spécialisé dans les sciences de l’Antiquité et l’archéologie. Pour ce faire, le choix a été fait de procéder en deux temps, le premier théorique (chapitres I et II) et le second pratique (chapitre III).

Dans les deux premiers chapitres, nous nous sommes intéressés à ce qu’était une ontologie. Pour cela nous sommes partis des origines philosophiques du terme pour ensuite définir son sens en ingénierie des connaissances et en relation avec les langages documentaires puisque la mission les faisait se côtoyer. Ensuite, nous avons étudié la manière de concevoir et de réaliser une ontologie en ingénierie des connaissances en énumérant ses composants et en proposant une méthode de construction. Nous avons également abordé deux points plus techniques en présentant d’une part, le langage de représentation le plus répandu, standard du W3C, pour représenter des ontologies, OWL, et d’autre part, en exposant les différents logiciels permettant d’éditer des ontologies. Enfin, dans le troisième et dernier chapitre consacré à la mise en pratique de la théorie, nous avons découvert en détail l’objet de l’étude : le micro-thésaurus PACTOLS-Sujets, comment j’ai procédé à sa transformation, et quelles perspectives donner au projet.

Grâce à cette double approche, théorique et pratique, nous sommes maintenant en mesure, dans cette conclusion, de qualifier le travail effectué. Ainsi, en ce qui concerne la réalisation de la mission, il s’agit d’une adaptation personnelle de la méthode de B. Bachimont et des propos évoqués dans les deux premiers chapitres. En effet, contrairement aux cas les plus fréquemment rencontrés et donc théorisés, mon but n’était pas de créer une ontologie *ex-nihilo*, mais de la concevoir à partir d’un vocabulaire, d’une base terminologique donnée : le micro-thésaurus PACTOLS-Sujets.

L’ontologie ainsi créée est une ontologie de domaine, surmontée d’une ontologie de haut niveau. Elle est pour le moment à cheval entre deux natures. En effet, elle tend à devenir formelle grâce à la perspective de l’importer et la compléter dans Protégé, mais reste, dans DOE, semi-formelle. Enfin, elle est de granularité variable, en fonction de la précision du micro-thésaurus dans les différentes thématiques.

Cette expérience de transformation a été pour moi des plus positive et constructive. Même si elle ne concerne qu'une partie du thésaurus (les objets mobiles) la méthode utilisée est capable de contenir tout le thésaurus, « au prix d'un travail considérable de formalisation » [MENON 2007]. Le travail, sans véritable appui spécialisé certes, mais au sein d'une équipe dynamique et disponible m'a permis d'aller de l'avant, et ce sans me décourager, dans l'accomplissement de cette tâche minutieuse et de grande envergure. En outre, je me suis constituée peu à peu de solides connaissances, et je l'espère compétences, dans le domaine très particulier de l'ingénierie ontologique. Je me suis ainsi complètement familiarisée avec l'éditeur d'ontologies DOE, avec lequel j'ai pu modéliser l'ontologie de façon rationnelle. En revanche, si cet éditeur m'est apparu pendant longtemps comme un outil irréprochable, j'émetts aujourd'hui quelques soupçons quant à son intégrité. En effet, avec Madame Lequeux nous nous sommes aperçues lors de l'Université d'été du TGE ADONIS, qu'il était impossible d'utiliser ce logiciel sans connexion Internet. Miled Rousset, l'informaticien de FRANTIQ, a regardé de plus près ce qu'il se passait et nous a appris que DOE avait besoin d'une connexion pour entrer et/ou sortir des informations contenues dans l'éditeur. Les informations contenues dans l'éditeur se résumant au fruit de mon travail, nous sommes en droit de nous demander si notre travail n'était pas observé. Cela dit, nous n'avons pas assez approfondi la question pour se permettre de porter un jugement, il s'agit juste d'une constatation. Nous en avons parlé à Madame Nathalie Aussenac-Gilles qui devrait se renseigner prochainement auprès des personnes concernées.

Enfin, en proposant ce projet, Madame Lequeux souhaitait se rendre compte des possibilités du passage des PACTOLS en ontologie. L'exercice intellectuel lui a semblé très intéressant et pourrait apporter de bons résultats dans l'aide à la recherche. Cependant, pour le nouveau mode de fonctionnement de la recherche – sur programme de un ou trois ans (ANR, etc.) – ce passage en ontologie est trop chronophage. Comme nous l'avons évoqué, ce sont les objectifs du TGE ADONIS qui orientent le projet vers une utilisation pragmatique immédiate ; nouvelle perspective que je m'efforcerai d'adopter dans le cadre d'une seconde mission.

Table des matières

RESUME	3
SUMMARY	3
SOMMAIRE	5
INTRODUCTION	7
Contexte de la mission	8
Objet d'étude	11
CHAPITRE I : QU'EST-CE QU'UNE ONTOLOGIE ?	17
I. Un peu d'histoire	17
1) <i>Du principe philosophique : la science de l'être en tant qu'être, à la création du terme d'Ontologie</i>	18
2) <i>Évolutions à travers le temps</i>	19
II. L'ontologie en ingénierie des connaissances	21
1) <i>Définition</i>	21
2) <i>Les ontologies</i>	25
III. Ontologie et langage documentaire	28
1) <i>Les langages documentaires</i>	28
2) <i>Les ontologies et les langages documentaires</i>	32
CHAPITRE II : COMMENT CONSTRUIRE UNE ONTOLOGIE ?	37
I. Conception	38
1) <i>Constituants d'une ontologie</i>	39
Les concepts	39
Les propriétés	40
Les relations	40
2) <i>Engagement sémantique et engagement ontologique pour la conception et réalisation d'ontologies</i>	42
1 ^{re} étape : extraction de termes et analyse	43
2 ^e étape : la normalisation sémantique	43
3 ^e étape : l'engagement ontologique	45
4 ^e étape : l'opérationnalisation	45
II. Langages de représentation des connaissances	45

1) OWL	47
XML	48
RDF	48
OWL	49
2) SKOS	51
III. Éditeurs d'ontologies.....	53
1) Protégé.....	54
2) OilEd.....	55
3) OntoEdit.....	56
4) WebODE	57
5) DOE.....	57
CHAPITRE III : VERS LA TRANSFORMATION DU MICRO-THESAURUS PACTOLS-SUJETS EN ONTOLOGIE.....	60
I. Description du micro-thésaurus Sujets des PACTOLS.....	60
1) Structure et thématique de PACTOLS-Sujets	61
2) Les descripteurs dans PACTOLS-Sujets et la gestion des termes	63
3) Représentation du thésaurus PACTOLS et son utilité.....	69
II. Choix méthodologiques	72
1) Critères de sélection	73
2) Solutions écartées	75
3) Solution retenue : utilisation du logiciel DOE	80
III. Réalisations et perspectives.....	88
1) Réalisations.....	89
Appropriation et destructure du thésaurus.....	89
Description de la modélisation de l'ontologie obtenue à l'aide du logiciel DOE.....	99
2) Perspectives.....	112
Réalisation de la deuxième phase, la représentation et le raffinement du modèle dans Protégé.....	112
Autres perspectives.....	117
CONCLUSION	120
Table des matières	122
Table des illustrations	125
Table des acronymes	127

BIBLIOGRAPHIES	130
<i>Travaux cités</i>	131
<i>Indications bibliographiques</i>	134
Systèmes d'organisation des connaissances et ingénierie des connaissances	134
Ontologies et représentation des connaissances	134
Ingénierie ontologique.....	135
Ontologie en philosophie	137
Normes et standards.....	137
ANNEXES	139
Annexe I : <i>Organigramme de l'UMS 844 à la MAE</i>	140
Annexe II : <i>Carte des producteurs participants au réseau FRANTIQ</i>	141
Annexe III : <i>Liste des ouvrages et sites Internet de référence utilisés pour la création des PACTOLS et les définitions des Notes d'Application, élaborée par Madame Lequaux</i>	142
Annexe IV : <i>Liste complémentaire des ouvrages et sites Internet de référence utilisés pour compléter les Notes d'Application et les définitions encyclopédiques de l'ontologie</i>	144
Annexe V : <i>Définitions différentielles relatives à l'ontologie de haut niveau de utilisée pour l'ontologie créée à partir du micro-thésaurus PACTOLS-Sujets</i>	145
Annexe VI : <i>Arbre des concepts de l'ontologie différentielle sur l'archéologie et les sciences de l'Antiquité</i>	149
Annexe VII : <i>Diapositives de la présentation des projets FRANTIQ pour l'Université d'été du TGE ADONIS</i>	154

Table des illustrations

1. Construction d'une ontologie opérationnelle	27
2. Tableau récapitulatif du classement des langages documentaires	29
3. Généalogie « des langages documentaires »	31
4. Objets pouvant être qualifiés d'ontologies	33
5. Les différents niveaux de formalisme et d'engagement sémantique en représentation des connaissances	34
6. Le « gâteau » de Tim Berners-Lee extrait du support de l'exposé de Raphaël Troncy donné à l'occasion d'une journée d'étude sur le Web Sémantique en juillet 2008	47
7. Représentation d'un triplet RDF	49
8. Représentation de la relation associative dans SKOS	52
9. Syntaxe OWL	53
10. Copie d'écran de l'interface principale de l'éditeur d'ontologies Protégé	55
11. Copie d'écran de l'interface principale de l'éditeur d'ontologies DOE	59
12. Tableau récapitulatif des 26 hyperonymes et de leur domaine d'action.	63
13. Copies d'écran montrant la distinction entre les basiliques civiles et religieuses par l'utilisation du latin	64
14. Copie d'écran montrant la poly-hiérarchie du micro-thésaurus PACTLOS-Sujets – objet religieux	67
15. Copie d'écran illustrant la hiérarchie des concepts – Couverture	68
16. Copies d'écran montrant les différents modes de consultation des thésaurus PACTOLS	72
17. Représentation des deux schémas : onomasiologique et sémasiologique	76
18. Méta-modèle d'organisation d'une base de données terminologique	77
19. Les trois étapes de la méthode de construction d'ontologies	81
20. Copie d'écran montrant la structure de l'ontologie du cyclisme, réalisée par R. Troncy et A. Isaac, et les principes différentiels de la notion Spectateur.	82
21. Copie d'écran montrant les possibilités complémentaires de DOE	83
22. Copie d'écran montrant la présence et la situation du concept <i>Maillot</i> dans l'ontologie référentielle du cyclisme	85
23. Copie d'écran montrant la hiérarchie des relations selon le principe différentiel	86
24. Copie d'écran de l'interface de DOE permettant de saisir les spécificités des relations	87

25. Représentation schématique de la taxinomie d'Ustensile	91
26. Structure fondamentale de l'ontologie du cyclisme	92
27. Extrait du tableau général de classement des termes du thésaurus par catégorisation.	93
28. Représentation de l'ordonnancement graphique du concept Technique	94
29. Copie d'écran du logiciel DOE représentant la hiérarchie des concepts selon trois grandes rubriques : Objet, Propriété, Activité	96
30. Tableau résumant les changements qui pourraient être effectués dans la structure du micro- thésaurus PACTOLS-Sujets dans le cadre de sa transformation en ontologie.	97
31. Copie d'écran du logiciel DOE représentant la solution intermédiaire envisagée pour transformer le thésaurus en ontologie.	98
32. Copie d'écran de l'ontologie de haut niveau structurant l'ontologie de domaine issue du micro-thésaurus PACTOLS-Sujets.....	101
33. Définitions différentielles des deux premiers concepts de haut niveau structurant l'ontologie	102
34. Définition différentielle d'Objet Mobile.....	102
35. Copie d'écran mettant en évidence le principe différentiel attaché à la notion d'instrument.	104
36. Copie d'écran mettant en évidence le principe différentiel attaché à la notion d'Objet Outil, sœur de la notion Instrument	105
37. Copie d'écran montrant certaines informations complémentaires apportées à la notion Carnyx	106
38. Tableau révélant les modifications apportées à l'ontologie référentielles.....	108
39. Copie d'écran illustrant le cas particulier de l'héritage multiple avec l'exemple de Coquemar	109
40. Copie d'écran de la hiérarchie des relations rattachée à l'ontologie référentielle	110
41. Copie d'écran de l'interface de base proposé par Protégé	113
42. Copie d'écran montrant l'import dans Protégé de la modélisation de l'ontologie obtenue dans DOE pour les Classes.....	115
43. Copie d'écran montrant l'import dans Protégé de la modélisation de l'ontologie obtenue dans DOE pour les Propriétés	115
44. Copie d'écran prouvant la présence de la totalité de la modélisation effectuée dans DOE, dans Protégé	117

Table des acronymes

- **ADBS** : Association des professionnels de l'information et de la documentation
- **AdIFI** : Archéologie de la France-informations
- **AFNOR** : Association Française de NORmalisation
- **ANR** : Agence Nationale de la Recherche
- **CCI** : Catalogue Collectif Indexé
- **CDU** : Classification Décimale Universelle
- **CDSH** : Centre de Documentation en Sciences Humaines
- **CEPAM** : Centre d'Études de la Préhistoire, de l'Antiquité et du Moyen Age
- **CNRS** : Centre National de la Recherche Scientifique
- **CREAAH** : Centre de Recherche en Archéologie, Archéosciences, Histoire
- **DAPHNE** : Données en Archéologie, Préhistoire et Histoire sur le NEt.
- **DOE** : *Differential Ontology Editor*
- **DWP** : *Difference With Parents*
- **DWS** : *Difference With Siblings*
- **ETS** : Ensembles Thématiques Signés
- **FRANTIQ** : Fédération de Ressources sur l'ANTIQuité
- **GID** : Gestion de l'Information et du Document
- **HTML** : HyperText Markup Language
- **IEEE** : *Institute of Electrical and Electronics Engineers*
- **INA** : Institut National de l'Audiovisuel
- **INIST** : INstitut de l'Information Scientifique et Technique
- **INRAP** : Institut National de Recherches Archéologiques Préventives
- **IRIT** : Institut de Recherche en Informatique de Toulouse
- **ISO** : *International Organization for Standardization*
- **ITM** : *Intelligent Topic Manager*
- **LGRD** : Logiciel de Gestion et de Recherche Documentaire
- **MAE** : Maison de l'Archéologie et de l'Ethnologie

- **MAFE** : Missions Archéologiques Françaises à l'Étranger
- **MOM** : Maison de l'Orient Méditerranéen
- **NA** : Note d'Application / **SN** : *Scope note*
- **OPALES** : Outils pour des Portails Audiovisuels Éducatifs et Scientifiques
- **OWL** : *Web Ontology Language*
- **OWL DL** : *Web Ontology Language Description Logics*
- **PACTOLS** : Peuples et cultures, Anthroponymes, Chronologie relative, Toponymes, Œuvres, Lieux et Sujets
- **PIC** : Plateforme Intégrée de Consultation
- **RDF** : *Resource Description Framework*
- **RDFS** : *Resource Description Framework Schema*
- **SBC** : Système à Base de Connaissances
- **SHS** : Sciences de l'Homme et de la Société, anciennement Sciences Humaines et Sociales
- **SKOS** : *Simple Knowledge Organization System*
- **SIGB** : Système Intégré de Gestion de Bibliothèque
- **SUO** : *Standard Upper Ontology*
- **SWAD-Europe** : *Semantic Web Advance Development for Europe*
- **SWP** : *Similarity With Parent*
- **SWS** : *Similarity With Siblings*
- **TA** : Terme Associé / **RT** : *Related Term*
- **TG** : Terme Générique / **BT** : *Broader Term*
- **TGE ADONIS** : Très Grand Équipement Accès Unifié aux données et documents numériques des sciences humaines et sociales
- **TG G** : Terme Générique en Relation Générique
- **TG P** : Terme Générique Partitif
- **TMF** : *Terminological Markup Framework*
- **TML** : *Terminological Markup Language*
- **TS** : Terme Spécifique / **NT** : *Narrower Term*
- **TS G** : Terme Spécifique en Relation Générique.
- **TS P** : Terme Spécifique Partitif
- **UMR** : Unité Mixte de Recherche
- **UMS** : Unité Mixte de Service

- **URI** : *Uniform Resource Identifier*
- **W3C** : *World Wide Web Consortium*
- **XML** : *eXtensible Markup Language*

BIBLIOGRAPHIES

Travaux cités

Indications bibliographiques

Travaux cités

- [BACHIMONT 2000] « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances ». In : CHARLET, Jean *et al.* *Ingénierie des connaissances. Évolutions récentes et nouveaux défis*. Paris : Eyrolles, p. 305-323.
- [BANEYX 2007] *Construire une ontologie de la pneumologie : Aspects théoriques, modèles et expérimentations*. Thèse de doctorat en informatique médicale. Paris 6.
- [BERNERS-LEE 2001] The Semantic Web, *Scientific American*, <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>
- [BOULOGNE : 2004] *Vocabulaire de la documentation*. Paris : ADBS éditions. pp. 144-145.
- [CHANDRASEKARAN *et al.*1999] What are ontologies and why do we need them? *IEEE Intelligent Systems*. 14(1):20-26.
- [CHARLET 2002] *L'Ingénierie des connaissances : Développements, résultats et perspectives pour la gestion des connaissances médicales*. Mémoire d'habilitation à diriger des recherches. Université Pierre et Marie Curie, spécialité : informatique. XIII-125p.
- [CHARLET *et al.* 2004] Ontologies pour le Web sémantique. In *Revue i3*, numéro Hors Série « Web sémantique ».
- [FÜRST 2004] *Contribution à l'ingénierie des ontologies : une méthode et un outil d'opérationnalisation*. Thèse de doctorat en informatique. École polytechnique de l'Université de Nantes.
- [CLAUBERG 1647] *Elementa philosophiae ; sive Ontosophia...*Groningae : typis J. Nicolai, 311 p.
- [GUARINO 1997a] Some organizing principles for a unified top-level ontology. *Proceedings of the AIII Spring Symposium on Ontological Engineering*.
- [GUARINO 1997b] Understanding, building and using ontologies: a commentary to 'using explicit ontologies in kbs development.
- [GUARINO et POLI 1995] Formal ontology in conceptual analysis and knowledge representation. *Special issue of the International Journal of Human and Computer Studies*. 43(5/6):625-640.
- [GOCCLENIUS 1613] *Lexicon philosophicum*. Francofurti : typis viduae M. Beckeri, 1143 p.
- [GRUBER 1993a] *Formal ontology in conceptual analysis and knowledge representation*. Chap. Towards principles for the design of ontologies used for knowledge sharing. Kluwer Academic Publishers.

- [GRUBER 1993b] « A translation Approach to Portable Ontology Specifications ». *Knowledge Acquisition*, vol. 5, n° 2, p.199-220 (b)
- [HEGEL 1994] *Science de la logique*. Paris : Aubier.
- [ISO 2788:1986] - Documentation -- *Principes directeurs pour l'établissement et le développement de thésaurus monolingues*
- [ISO 5964: 1985] - Documentation -- *Guidelines for the establishment and development of multilingual thesauri*. 61 p.
- [KANT 2001] *Critique de la raison pure*. Paris : Flammarion, 749 p.
- [LE ROBERT 2006] *Le nouveau Petit Robert : dictionnaire alphabétique et analogique de la langue française*. Dictionnaires Le Robert : Paris, 2837 p.
- [MENON 2007] « Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan. » *Documentaliste – Sciences de l'information*, vol. 44, n°1, p. 18-28.
- [POLI 1996] « Ontology for Knowledge organization ». In : GREEN, Rebecca (dir.) *Knowledge organization and change: proceedings of the Fourth international ISKO Conference, July 15 – 18, 1996. Washington (DC)*. Francfort : Indeks Verlag, p. 3193-319.
- [PSYCHE *et al.* 2003] « Apport de l'ingénierie ontologique aux environnement de formation à distance ». *Revue Sticef*, vol. 10. http://sticef.univ-lemans.fr/num/vol2003/psyche-06s/sticef_2003_psyche_06s.pdf
- [RASTIER 1995] « Le terme : entre ontologie et linguistique ». *La banque des mots*, n°7, p. 35-65
- [RASTIER 1987]. *Sémantique interprétative*. Paris : PUF,
- [RASTIER *et al.* 1994] *Sémantique pour l'analyse*. Paris : Masson,
- [SMITH et WELTY 2001] “Ontology: Towards a New Synthesis”. In : SMITH, Barry et WELTY, Christopher (dir.) *Formal ontology in information system: proceedings of the international conference on formal ontology in information system, Ogunquit, Maine, USA, October 17 – 19*. New York: ACM Press, p. III-X.
- [SOWA 1999] *Knowledge representation: Logical, philosophical and computational foundations*. Brooks Cole Publishing Co.: Pacific Grove, CA USA.
- [STUDER *et al.* 1998] “Knowledge engineering: Principles and methods”. *IEEE Transactions on Data and Knowledge Engineering*, vol. 25, n° 1-2, p. 161-197.
- [TRONCY et ISAAC 2002]. DOE : une mise en œuvre d'une méthode de structuration différentielle pour les ontologies. In 13^e Journées Francophones d'Ingénierie des Connaissances (IC'02), Rouen, 28-30 mai 2002, pages 63-74.

- [USCHOLD et GRUNINGER 1996] Ontologies: Principles, Methods and Applications”. *Knowledge Engineering Review*, vol.11, n°2, p. 93-136
- [USCHOLD et KING 1995]. Towards a methodology for building ontologies. *In* Workshop on Basic Ontological Issues in Knowledge Sharing. International Joint Conference on Artificial Intelligence.
- [VAN HEIJST *et al.*] Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 45(2/3), 183-292.
- [WOLFF 1730] *Philosophia Prima sive Ontologia*. Francofurti et Lipsiae : prostate in officina libraria Rengeriana, 696 p.

Indications bibliographiques

Systèmes d'organisation des connaissances et ingénierie des connaissances

- CAUVET, Corine et ROSENTHAL-SABROUX, Camille (dir.). *Ingénierie des systèmes d'information*. Paris : Hermès-Science Publication, 2001, 353p.
- CONTAT, Odile. *Langages documentaires et nouvelles technologies : l'avenir des langages et leur positionnement au cœur des systèmes d'informations dans le contexte de la presse*. Mémoire de DESS en Sciences de l'information et de la documentation spécialisées. Institut national des techniques de la documentation du CNAM. Soutenu le 5 novembre 2003. disponible sur http://memsic.ccsd.cnrs.fr/mem_00000057.en.html
- GHOUAS DZIRI, Anissa. « Ontologie et thésaurus : des systèmes d'organisation des connaissances différents et complémentaires. » In : TIMIMI Ismaïl (dir.) *Indice, index, indexation : actes du colloque international organisé les 3 et 4 novembre 2005 à l'université Lille-3 par les laboratoires CERSATES et GERICO*. Paris : ADBS éditions, pp. 191-202.
- MANIEZ, Jacques. *Actualité des langages documentaires : fondements théoriques de la recherche d'information*. Paris : ADBS éditions, 2002.
- MENON, Bruno. « Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan. » *Documentaliste – Sciences de l'information*, vol. 44, n°1, 2007, p. 18-28.
- MENON, Bruno. « L'évolution des langages documentaires. » *Les dossiers de l'ingénierie éducative*, n° 49, décembre 2004.
- ROMARY, Laurent et VAN CAMPENHOUDT, Marc. « Normalisation des échanges de données en terminologie : le cas des relations dites « conceptuelles ». » *Conférence TL4-2001, Nancy, 3 et 4 mai 2001*. <http://www.termisti.refer.org/tia4.pdf>
- ZACKLAD, Manuel. « Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI). » In : ARSENAULT C. et DALKIR K. (dir.). *35^e Congrès annuel de l'Association Canadienne des Sciences de l'Information. Partage de l'information dans un monde fragmenté : Franchir les frontières*. Montréal : CAIS/ACSI, 2007. Disponible sur http://www.cais-acsi.ca/proceedings/2007/kacklad_2007.pdt

Ontologies et représentation des connaissances

- GRUBER, Thomas R. *Formal ontology in conceptual analysis and knowledge representation*. Chap. Towards principles for the design of ontologies used for knowledge sharing. Kluwer Academic Publishers. 1993.

- GRUBER, Thomas R. « A translation Approach to Portable Ontology Specifications ». *Knowledge Acquisition*, vol. 5, n° 2, 1993, p.199-220
- GUARINO, Nicola. Some organizing principles for a unified top-level ontology. *Proceedings of the AIII Spring Symposium on Ontological Engineering*, 1997.
- GUARINO, Nicola. Understanding, building and using ontologies: a commentary to ‘using explicit ontologies in kbs development, 1997.
- GUARINO, Nicola et POLI, Roberto. Formal ontology in conceptual analysis and knowledge representation. *Special issue of the International Journal of Human and Computer Studies*, 1995. 43(5/6):625-640.
- POLI, Roberto. « Ontology for Knowledge organization ». In : GREEN, Rebecca (dir.) *Knowledge organization and change: proceedings of the Fourth international ISKO Conference, July 15 – 18, 1996. Washington (DC)*. Francfort : Indeks Verlag, p. 3193-319.
- RASTIER François. « Ontologies. » *Revue des sciences et technologies de l'information*, série : Revue d'Intelligence artificielle, 2004, vol. 18, n°1, p. 15-40
- RASTIER, François. « Le terme : entre ontologie et linguistique ». *La banque des mots*, n°7, 1995, p. 35-65.
- RASTIER, François. *Sémantique interprétative*. Paris : PUF, 1987.
- RASTIER, François *et al.* *Sémantique pour l'analyse*. Paris : Masson, 1994.
- SOWA, John F. *Knowledge representation: Logical, philosophical and computational foundations*. Brooks Cole Publishing Co.: Pacific Grove, CA USA. 1999.

Ingénierie ontologique

- BACHIMONT, Bruno. « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances ». In : CHARLET, Jean *et al.* *Ingénierie des connaissances. Évolutions récentes et nouveaux défis*. Paris : Eyrolles, 2000, p. 305-323.
- BACHIMONT, Bruno. « Modélisation linguistique et modélisation logique des ontologies : l'apport de l'ontologie formelle. » In : Actes de la conférence « IC'2001 », 2001, Grenoble.
- BANEYX, Audrey. *Construire une ontologie de la pneumologie : Aspects théoriques, modèles et expérimentations*. Thèse de doctorat en informatique médicale. Paris 6. Soutenue le 06 février 2007
- BERNES-LEE, Tim *et al.*. The Semantic Web, *Scientific American*, 2001, disponible sur <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

- BOURIGAULT Daniel., AUSSENAC-GILLES Nathalie et, CHARLET Jean. « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas. » *Revue d'Intelligence Artificielle (RIA)*. Numéro spécial sur les Techniques Informatiques et Structuration de Terminologies. PIERREL J.M. et SLODZIAN M. (Ed.). Paris : Hermès. 18 (1) : 87–110. 2004.
- CHANDRASEKARAN, Bruce *et al.* What are ontologies and why do we need them? *IEEE Intelligent Systems*. 1999, 14(1):20-26.
- CHARLET, Jean. *L'Ingénierie des connaissances : Développements, résultats et perspectives pour la gestion des connaissances médicales*. Mémoire d'habilitation à diriger des recherches. Université Pierre et Marie Curie, spécialité : informatique. Soutenue en décembre 2002. XIII-125 p.
- CHARLET, Jean *et al.* Ontologies pour le Web sémantique. In *Revue i3*, 2004, numéro Hors Série « Web sémantique ».
- FÜRST, Frédéric. *Contribution à l'ingénierie des ontologies : une méthode et un outil d'opérationnalisation*. Thèse de doctorat en informatique. École polytechnique de l'Université de Nantes. Soutenue le 25 novembre 2004.
- HORRIDGE, Matthew *et al.* *A practical guide to building OWL ontologies using the Protégé-OWL plugin and CO-ODE Tools. Edition 1.0*. The University of Manchester. 2004. <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>
- ISAAC, Antoine. Conception et utilisation d'ontologies pour l'indexation de documents audiovisuels. Thèse de doctorat en informatique : concepts et langages. Université Paris IV-Sorbonne. Soutenue le 8 décembre 2005. <http://www.few.vu.nl/~aisaac/papers/Isaac-These-Finale.pdf>
- MALAISE Véronique *et al.* « Vers une combinaison de méthodologies pour la structuration de termes en corpus. » In : POLITY, Yolla *et al.*(dir). *L'organisation des connaissances : approches conceptuelles*. Paris : l'Harmattan, 2005, pp. 179-189.
- NOY, Natalya F. et MCGUINNESS, Deborah. *Développement d'une ontologie 101 : Guide pour la création de votre première ontologie*. Traduit de l'anglais par Anila Angjeli, BnF, Bureau de normalisation documentaire. <http://www.bnf.fr/pages/infopro/normes/pdf/no-DevOnto.pdf>
- PSYCHE, Valéry *et al.* « Apport de l'ingénierie ontologique aux environnements de formation à distance ». *Revue Sticef*, vol. 10. 2003 http://sticef.univ-lemans.fr/num/vol2003/psyche-06s/sticef_2003_psyche_06s.pdf
- SMITH, Barry et WELTY, Christopher. "Ontology: Towards a New Synthesis". In : SMITH, Barry et WELTY, Christopher (dir.) *Formal ontology in information system: proceedings of the international conference on formal ontology in information system, Ogunquit, Maine, USA, October 17 – 19 2001*. New York: ACM Press, p. III-X.
- STICEF. *Recueil 2004*. Lyon : Institut national de recherche pédagogique, vol. 11, 362 p.

- STUDER, Rudi *et al.* “Knowledge engineering: Principles and methods”. *IEEE Transactions on Data and Knowledge Engineering*, vol. 25, n° 1-2, 1998, p. 161-197.
- TRONCY, Raphaël. *Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies : applications à la description de documents audiovisuels*. Thèse de doctorat en informatique : systèmes et communications. Université Joseph Fourier – Grenoble I. Soutenue le 5 mars 2005. <http://homepages.cwi.nl/~troncy/Publications/Troncy-PhD04.pdf>
- TRONCY, Raphaël et ISAAC, Antoine. DOE : une mise en œuvre d'une méthode de structuration différentielle pour les ontologies. In 13^e Journées Francophones d'Ingénierie des Connaissances (IC'02), Rouen, 28-30 mai 2002, pages 63-74.
- USCHOLD, Mike et GRUNINGER, Michael. Ontologies: Principles, Methods and Applications”. *Knowledge Engineering Review*, vol.11, n°2, 1996, p. 93-136
- USCHOLD, Mike et KING, Martin. Towards a methodology for building ontologies. In Workshop on Basic Ontological Issues in Knowledge Sharing. International Joint Conference on Artificial Intelligence. 1995.
- VAN HEIJST Gertjan *et al.* Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 1997, 45(2/3), 183-292.

Ontologie en philosophie

- CLAUBERG. *Elementa philosophiae ; sive Ontosophia...* Groningae : typis J. Nicolai, 1647, 311 p.
- GOCCLENIUS. *Lexicon philosophicum*. Francofurti : typis viduae M. Beckeri, 1613, 1143 p.
- GRENET, Paul Bernard. *Ontologie, Analyse spectrale de la Réalité*. Paris : Beauchesne et ses fils, 1963, 214 p.
- HEGEL, G.W.F. *Science de la logique*. Paris : Aubier, 1994.
- KANT, Emmanuel. *Critique de la raison pure*. Paris : Flammarion, 2001, 749 p.
- MEYER, Michel. *Pour une histoire de l'ontologie*. Paris : PUF, 1999, 172 p.
- WOLFF, Christian. *Philosophia Prima sive Ontologia*. Francofurti et Lipsiae : prostate in officina libraria Rengeriana, 1730, 696 p.

Normes et standards

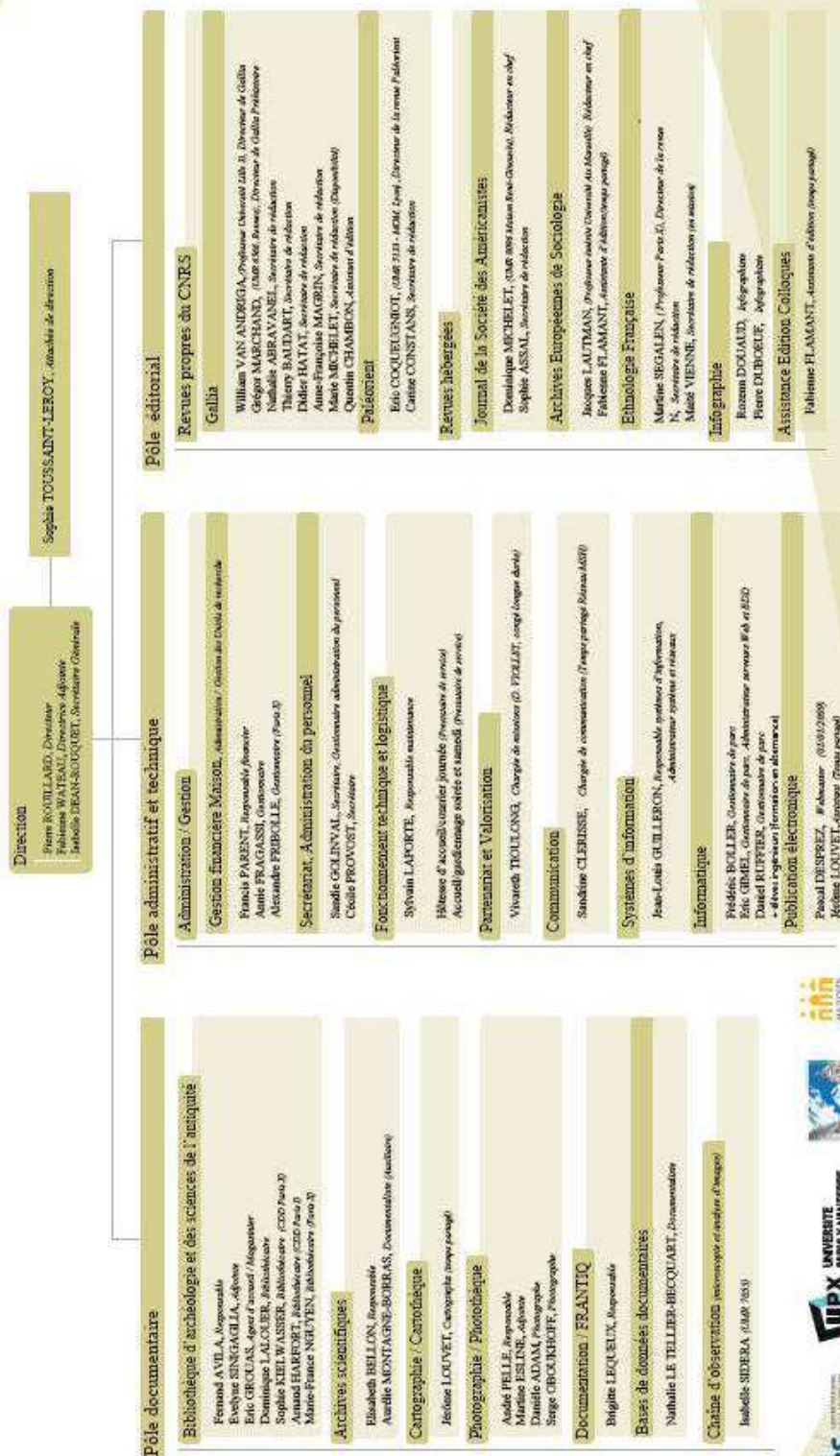
- ISO 16646. *Computer applications in terminology – Terminological markup framework*, 2003, 49 p.

- ISO 704. *Travail terminologique – Principes et méthodes*, 2000, 41 p.
- ISO 12620. *Aides informatiques en terminologie – Catégories de données*, 1999, 58 p.
- ISO 12200. *Applications informatiques en terminologie – Format de transfert de données terminologiques exploitables par la machine (MARTIF) – Transfert négocié*. 1999, 118 p.
- ISO 2788. *Documentation – Principes directeurs pour l'établissement et le développement de thésaurus monolingues*, 1986.
- ISO 5964. *Documentation – Guidelines for the establishment and development of multilingual thesauri*. 1985, 61 p.
- OWL. Recommandations du W3C du 10 février 2004. <http://www.w3.org/2004/OWL>
- SKOS. Publications du W3C. <http://www.w3.org/2004/02/skos/>
- RDF. Recommandations du W3C du 10 février 2004. <http://www.w3.org/RDF/>
- RDF-Schema. Recommandations du W3C du 10 février 2004. <http://www.w3.org/TR/rdf-schema/>
- XML. Publications du W3C. <http://www.w3.org/XML/>

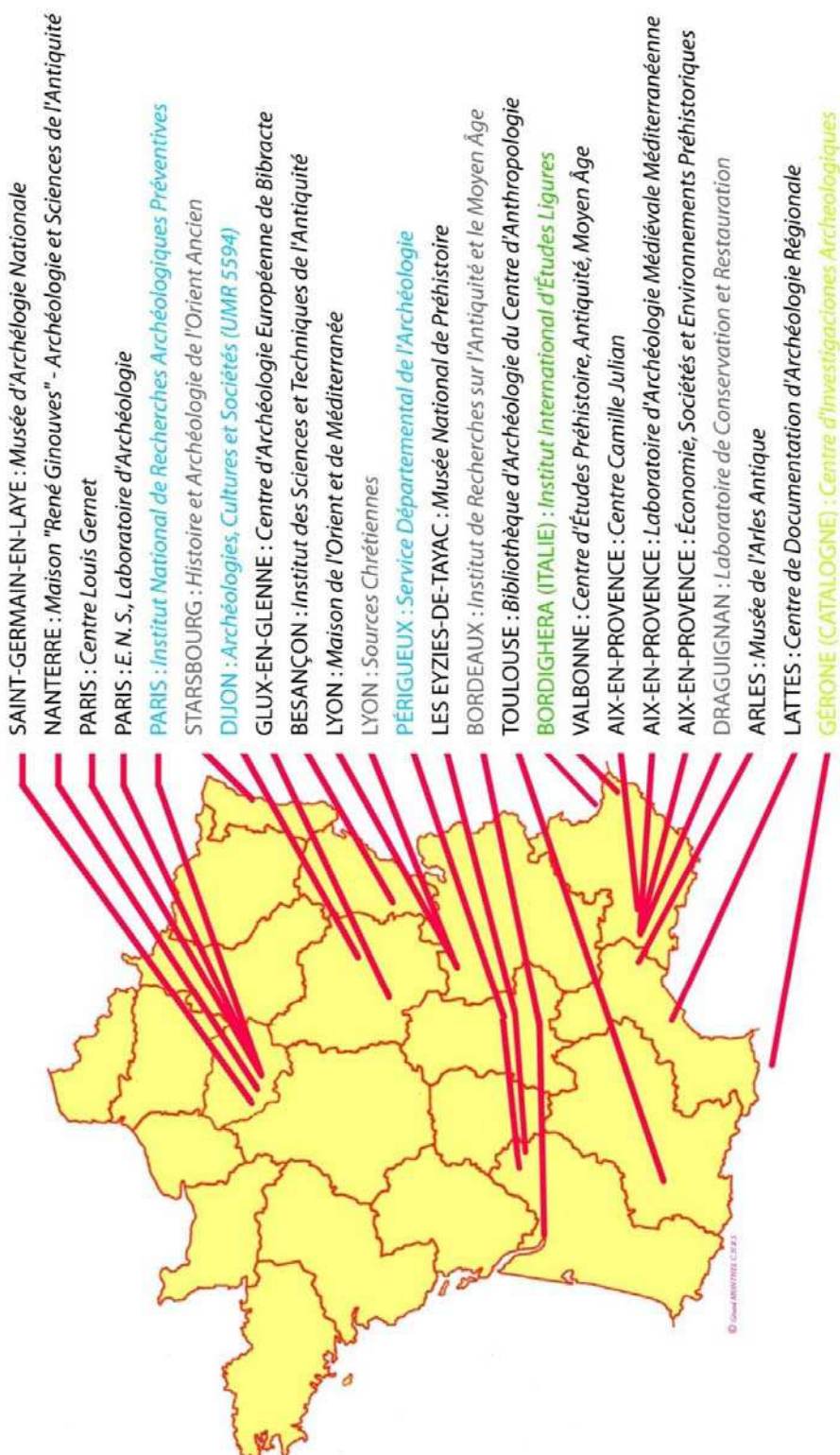
ANNEXES

- **I.** Organigramme de l'UMS 844 à la MAE
- **II.** Carte des producteurs participants au réseau FRANTIQ
- **III.** Liste des ouvrages et sites Internet de référence utilisés pour la création des PACTOLS et les définitions des Notes d'Application, élaborée par Madame Lequeux
- **IV.** Liste complémentaire des ouvrages et sites Internet de référence utilisés pour compléter les Notes d'Application et les définitions encyclopédiques de l'ontologie
- **V.** Définitions différentielles relatives à l'ontologie de haut niveau de utilisée pour l'ontologie créée à partir du micro-thésaurus PACTOLS-Sujets
- **VI.** Arbre des concepts de l'ontologie différentielle sur l'archéologie et les sciences de l'Antiquité
- **VII.** Diapositives de la présentation des projets FRANTIQ pour l'Université d'été du TGE ADONIS

Annexe I : Organigramme de l'UMS 844 à la MAE



Annexe II : Carte des producteurs participants au réseau FRANTIQ



Annexe III : Liste des ouvrages et sites Internet de référence utilisés pour la création des PACTOLS et les définitions des Notes d'Application, élaborée par Madame Lequeux

Audouin-Rouzeau, Frédérique. – Hommes et animaux en Europe de l'époque antique aux temps modernes : corpus des données archéozoologiques et historiques. – Paris : CNRS Ed., 1993. – 230 p. (Dossier de documentation archéologique ; 16)

Aurenche, Olivier (Dir.). – Dictionnaire illustré multilingue de l'architecture du Proche Orient ancien. – Lyon : MOM, 1977. – 391 p. (Collection de la Maison de l'Orient méditerranéen ; 3. Série archéologique ; 2) (IFAPO. Publication Hors-série)

Balfet, Hélène, M.F. Fauvet-Berthelot et S. Monzon. – Lexique et typologie des poteries pour la normalisation de la description des poteries. – Nouv. Ed. - Paris : Presse du CNRS, 1989. – 146 p. (CNRS Plus)

Bibliothèque diocésaine de Rennes. – Thésaurus. – 1988. – 625 p.

BUREAU DE RECHERCHES GEOLOGIQUES ET MINIERES (Orléans). – Thésaurus franco-allemand Sciences de la terre / BRGM, CNRS et Bundesanstalt für Bodenforschung. – 1974. – pagination multiple.

Camps-Fabrer, Henriette, Bourrelly, L., Nivelles N ; - Lexique des termes descriptifs de l'industrie de l'os. – Aix : LAPMO ; Marseille : URADCA, 1974. – 31 p.

Centre de recherche sur la documentation dans les sciences de l'Antiquité = CREDO. – Banque de données Cultures et religions antiques : thésaurus thématique et thésaurus des auteurs antiques. – Villeneuve-d'Ascq : Université de Lille III, sciences humaines, lettres et arts, 1987 cop. – 2 vol. (297 p. et 209 p.)

Chatenet, Monique et Verdier, Hélène (Dir.). – Thésaurus de l'architecture. – Paris : Ed. du Patrimoine, 2000. – (Documents et Méthodes ; 7)

CNRS. Laboratoire d'information et de documentation en géographie – Intergéo (Paris). – Thésaurus : index des noms de lieux / N. Voionmaa. – 1988. – 292 p.
Dictionnaire Larousse. – xx vol.

Dictionnaire Le Robert

France. – Direction du Patrimoine. Inventaire général. – Base Thésaurus : vocabulaire normalisé, dictionnaire raisonné (5.000 termes en juin 2000) Thèmes : urbanisme ; architecture religieuse, publique, domestique, industrielle, artisanale et commerciale ; génie civil, parcs et jardins, objets et mobiliers civils et religieux meubles civils domestiques ; objets, meubles et vêtements religieux ; instruments de musique ; instruments scientifiques ; machines de production et moteurs, avec équivalents anglais et italiens. Accès direct à la base de données ou accès par listes hiérarchique et alphabétique. [français, anglais, italien].
<http://www.culture.gouv.fr/documentation/thesarch/pres.htm>

- Garnier, François. – Thesaurus iconographique : système descriptif des représentations. – Paris : Le léopard d'Or : Direction du Patrimoine IGMRAF
- Ginouvé, René. – Dictionnaire méthodique de l'architecture grecque et romaine. T. III, Espaces architecturaux, bâtiments et ensembles. – EFA : EFR, 1998.
- Girardot, Jean-François. – Projet pour l'élaboration d'un thesaurus sectoriel de l'esclavage antique : exemples et propositions pour la structuration d'une facette. Mémoire de DESS « Informatique documentaire », Université de Lyon I, option ENSB « Conception et gestion des systèmes et des réseaux d'information » : 1988 (Dir ; de stage Pierre Lévêque)
- Guyot-Rougé, Claire. – traduction du lexique DYABOLA. Papier interne à la MOM, 1997
- Inizan M-L., M. Reduron, H. Roche et J Tixier. – Technologie de la pierre taillée, suivi par un vocabulaire multilingue... - Meudon : CREP, 1995. – 199 p. : ill. (Préhistoire de la pierre taillée ; 4)
- Lavell, Cherry. - British Archaeological Thesaurus — London : Council for British Archaeology, 1989-.69 p. (Practical handbooks in archaeology ; 4)
- Leroy-Gourhan, André (Dir.). – Dictionnaire de la Préhistoire. – Paris : PUF, 1984 et 2^e éd. 1994. – 1277 p.
- Nivelle-Muysers, Nicole. – Application d'un programme d'édition automatique de listes et journaux bibliographiques dans le domaine de l'histoire et de l'archéologie. – Marseille : CNRS, Centre d'analyse documentaire pour l'archéologie. – Cod./16/1969. – 312 p.
- Royal Commission on the Historical Monuments of England (London). – Thesaurus of archaeological terms. - 1989 ca. – vii-122 p.
- Université de Liège, Centre informatique de philosophie et lettres. - Programme MENTOR. - 75 p. puis Centre d'Histoire des Religions. Bibliographie critique et informatisée de la religion grecque antique : <http://www.ulg.ac.be/histrel/mentor.htm>
- Verdier, Hélène (Dir.). – Thesaurus des Objets mobiliers / par Aline Magnien. - Paris : Dir de l'architecture et du patrimoine, 2001 (Documents et Méthodes ; 8)
- Yon, Marguerite (Dir.). – Dictionnaire illustré multilingue de la céramique du Proche Orient ancien. – Lyon : MOM, 1981. – 310 p. (Collection de la Maison de l'Orient méditerranéen ; 10. Série archéologique ; 7) (IFAPO. Publication Hors-série)

Annexe IV : Liste complémentaire des ouvrages et sites Internet de référence utilisés pour compléter les Notes d'Application et les définitions encyclopédiques de l'ontologie

BOSC, Ernest. *Dictionnaire général de l'archéologie et des antiquités chez les divers peuples*. Paris : J. de Bonnot, 1985. 576 p.

COTTREL, Léonard (dir). *Dictionnaire encyclopédique d'archéologie*. Société d'édition de dictionnaire et encyclopédies : Paris, 1962. 573p.

Dictionnaire de la préhistoire. Paris : A. Michel : Paris, 1999. 1119p

DJINDJIAN, François. *Méthodes pour l'archéologie*. A. Colin : Paris, 1991. 401p.

HOWATSON, M. C. *Dictionnaire de l'Antiquité : mythologie, littérature, civilisation*. Paris : Robert Laffont. 1993, 1066 p.

KRUTA, Venceslas. *Les Celtes, histoire et dictionnaire : des origines à la romanisation et au christianisme*. Paris : R. Laffont. 2000, 1005 p.

LECLANT, Jean (Dir.). *Dictionnaire de l'Antiquité*. Paris : PUF, 2005. 2389 p.

LE ROBERT. *Le nouveau Petit Robert : dictionnaire alphabétique et analogique de la langue française*. Dictionnaires Le Robert : Paris, 2006. 2837p.

RACHET, Guy. *Dictionnaire de l'archéologie*. R. Laffont : Paris, 1983.

REAU, Louis. *Dictionnaire polyglotte des termes d'art et d'archéologie*. Paris : PUF, 1953. 247 p.

TLFi *Trésor de la langue française informatisé* : <http://atilf.atilf.fr/tlf.htm>

VIALOU, Denis (Dir.). *La Préhistoire : histoire et dictionnaire*. Paris : R. Laffont, 2004. 1637 p.

VILLARD, François. *Les vases grecs*. Paris : PUF, 1956. 109 p.

BL : Brigitte Lequeux

**Annexe V : Définitions différentielles relatives à l'ontologie de haut niveau
de utilisée pour l'ontologie créée à partir du micro-thésaurus PACTOLS-
Sujets**

[Objet]

--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler

[Propriété]

--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

[Objet]

--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler

[ObjetAbstrait]

--Indique si l'objet qui peut être localisé dans l'espace ou pas : c'est un type d'objet qui ne peut pas être repéré dans l'espace

[Objet]

--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler

[ObjetConcret]

--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps

[ObjetConcretSpatial]

--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial

[ObjetManufacture]

--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet résultant de l'activité humaine. Peut être considéré dans sa globalité ou en partie.

[ObjetFixe]

--Indique son degré de mobilité, s'il est transportable ou pas : il n'est pas transportable ou très difficilement

[Objet]

--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler

[ObjetConcret]

--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps

[ObjetConcretSpatial]

--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial

[ObjetManufacture]

--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet résultant de l'activité humaine. Peut être considéré dans sa globalité ou en partie.

[ObjetMobile]

--Indique son degré de mobilité, s'il est transportable ou pas : il est transportable à la main ou grâce à une énergie (eau, force animale...).

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretSpatial]
--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial
[ObjetBiologique]
--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet de nature biologique. Il peut être considéré dans sa globalité ou en partie.
[Personne]
--Indique une espèce biologique : c'est un être humain

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretSpatial]
--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial
[ObjetBiologique]
--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet de nature biologique. Il peut être considéré dans sa globalité ou en partie.
[Animal]
--Indique une espèce biologique : être vivant, organisé, doué de sensibilité et de mobilité sauf l'homme

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretSpatial]
--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial
[ObjetGéographique]
--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet de nature géographique. Il peut être considéré dans sa globalité ou en partie.
[ObjetGéoNaturel]
--Indique si l'objet existe indépendamment de l'activité administrative ou s'il est lié à celle-ci : l'objet est pré-existant à toute activité administrative

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretSpatial]

--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est spatial
[ObjetGéographique]
--Indique un angle privilégié (résultant d'une activité humaine) sous lequel on considère l'objet : c'est un objet de nature géographique. Il peut être considéré dans sa globalité ou en partie.
[ObjetGéoArtificiel]
--Indique si l'objet existe indépendamment de l'activité administrative ou s'il est lié à celle-ci : l'objet est défini par ou pour une activité administrative

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretTemporel]
--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est temporel
[Evénement] *définition non renseignée*

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretTemporel]
--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est temporel
[Processus] *définition non renseignée*

[Objet]
--Précise le statut de l'entité par rapport au discours : c'est ce dont on veut parler
[ObjetConcret]
--Indique si l'objet peut être localisé (dans l'espace ou dans le temps) ou pas : c'est un type d'objet qui peut être localisé dans l'espace ou dans le temps
[ObjetConcretTemporel]
--Précise si le repérage pour l'objet est spatial ou temporel : le repérage privilégié est temporel
[EtatDeChoses] *définition non renseignée*

[Propriété]
--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose
[PropriétéLieu]
--Indique le rôle éventuel par rapport à un objet et/ou une activité : Peut spécifier une relation de localisation
[LieuSpatial]
--Indique le type de relation de localisation dans lequel intervient le lieu : sert à localiser dans l'espace

[Propriété]

--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

[PropriétéLieu]

--Indique le rôle éventuel par rapport à un objet et/ou une activité : Peut spécifier une relation de localisation

[LieuTemporel]

--Indique le type de relation de localisation dans lequel intervient le lieu : sert à localiser dans le temps

[Propriété]

--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

[PropriétéMatière]

--Indique le rôle éventuel par rapport à un objet et/ou une activité : peut spécifier se dont est constitué l'objet et/ou l'activité

[Propriété]

--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

[PropriétéCaractère]

--Indique le rôle éventuel par rapport à un objet : peut spécifier le caractère spécifique d'un objet

[Propriété]

--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

[PropriétéFonction]

--Indique le rôle éventuel par rapport à un objet : peut spécifier la fonction d'un objet, ce à quoi il sert

[Propriété]

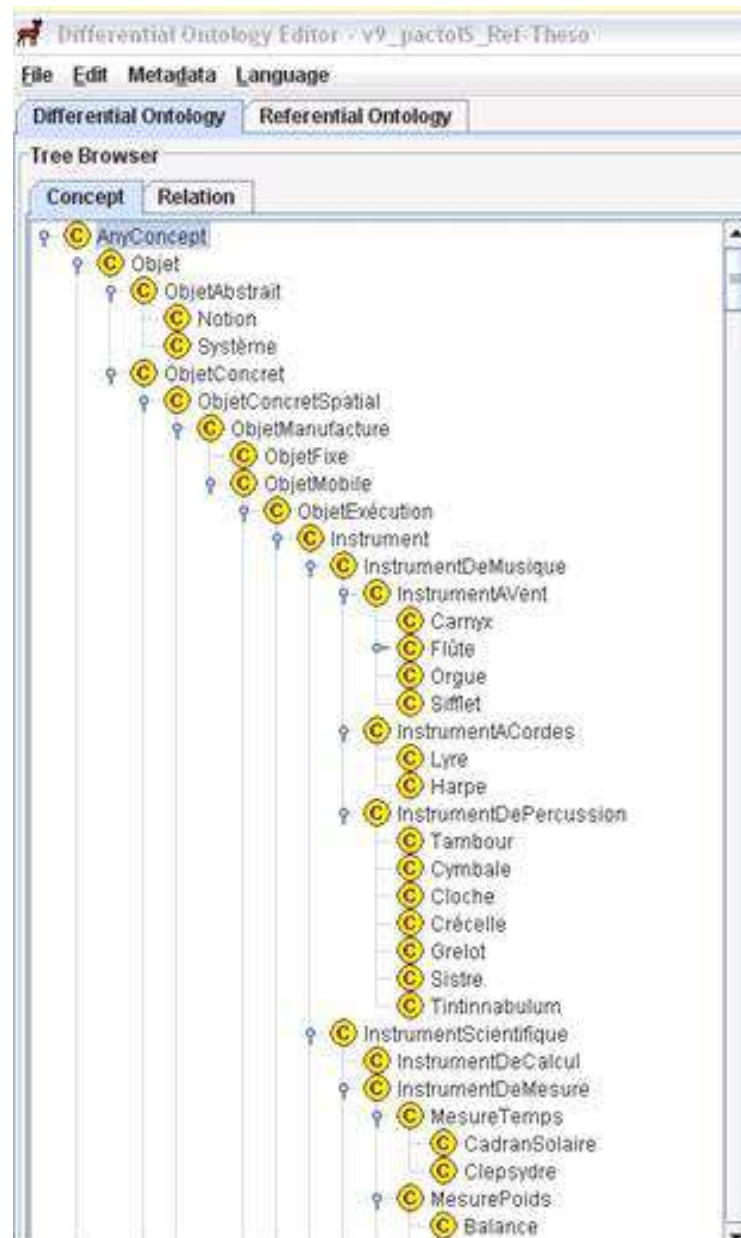
--Précise le statut de l'entité par rapport au discours : c'est ce qu'on veut exprimer à propos de quelque chose

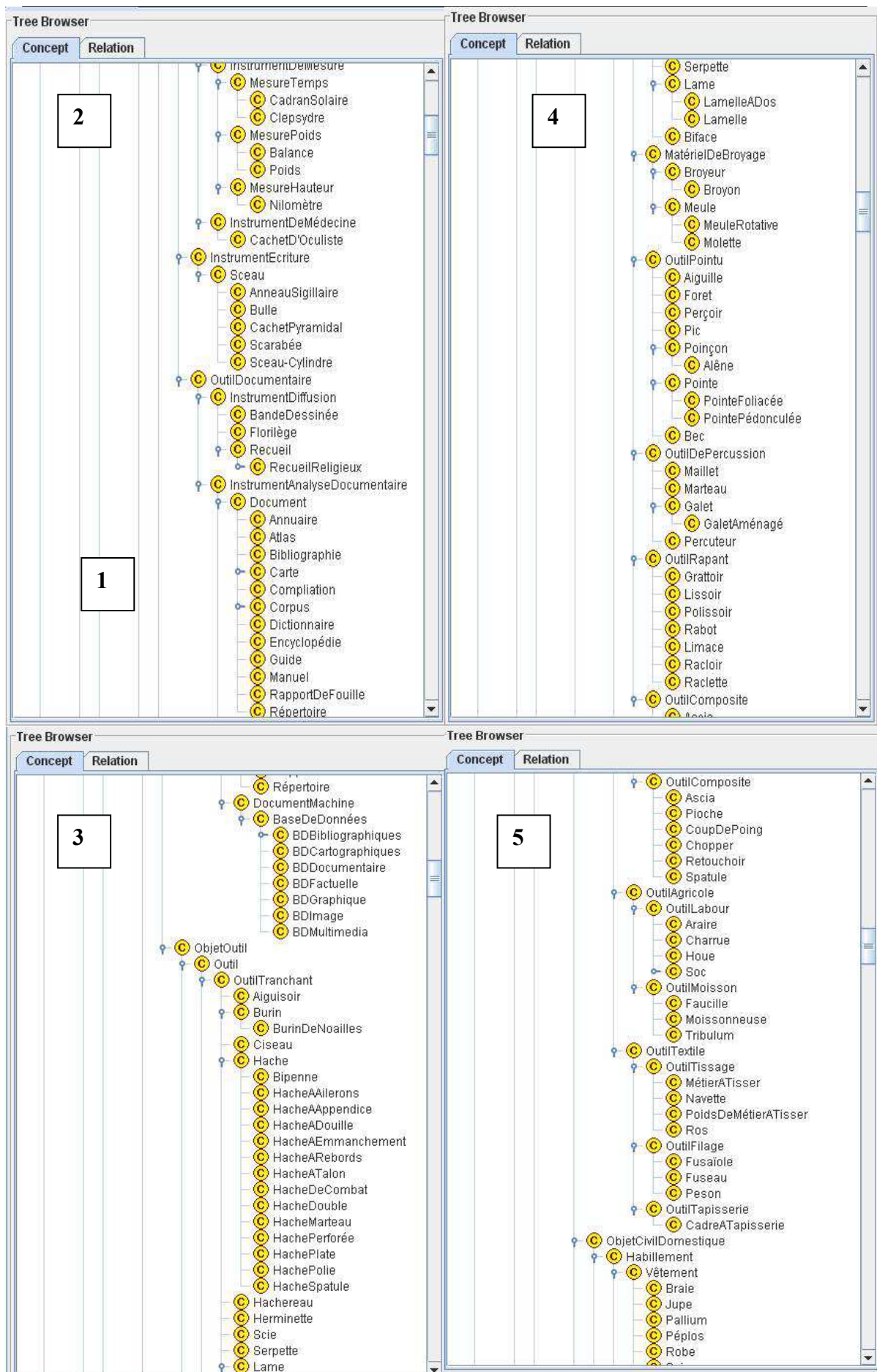
[Domaine]

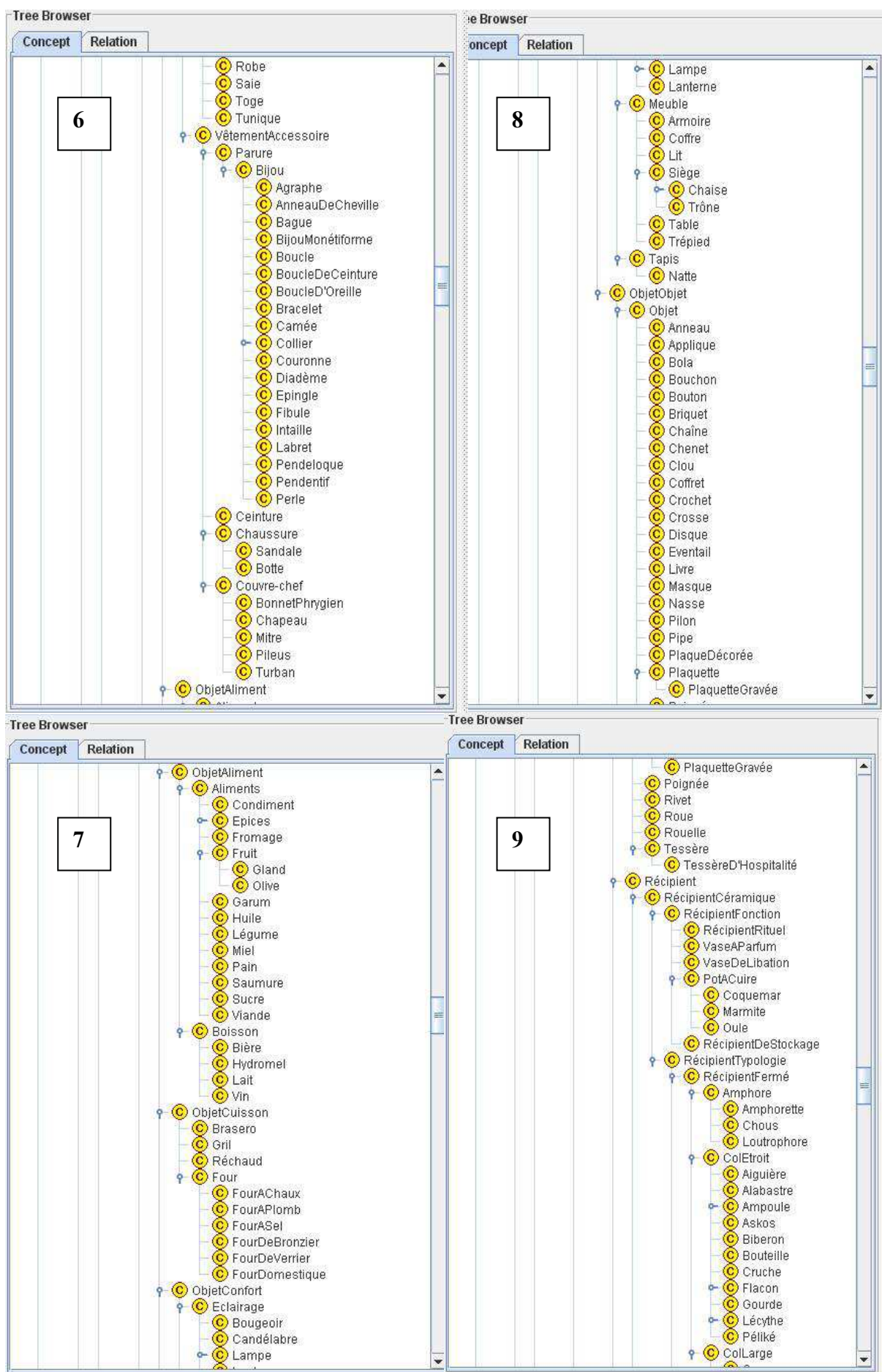
--Indique le rôle éventuel par rapport à un objet : peut spécifier le contexte de l'objet

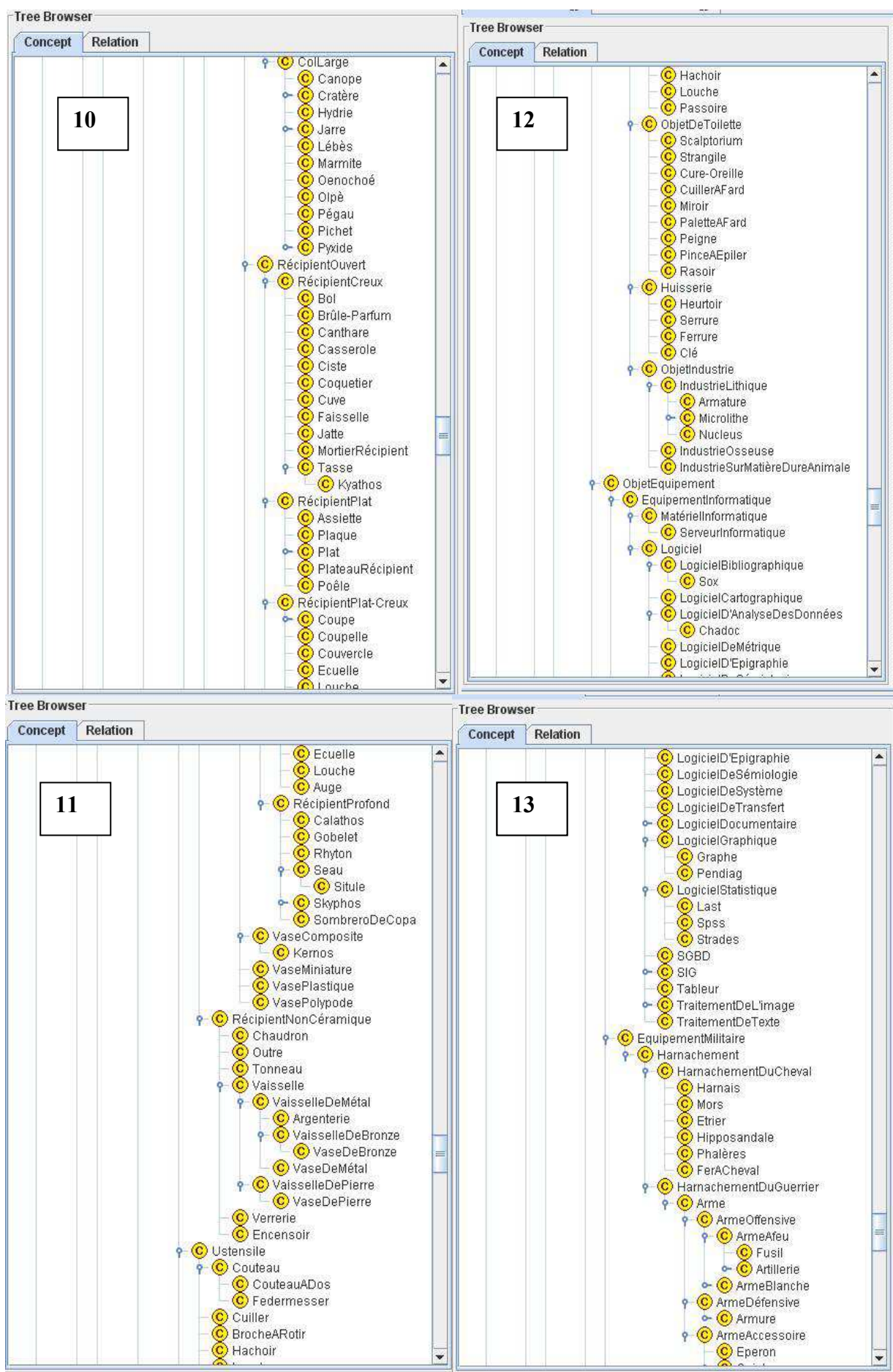
Annexe VI : Arbre des concepts de l'ontologie différentielle sur l'archéologie et les sciences de l'Antiquité

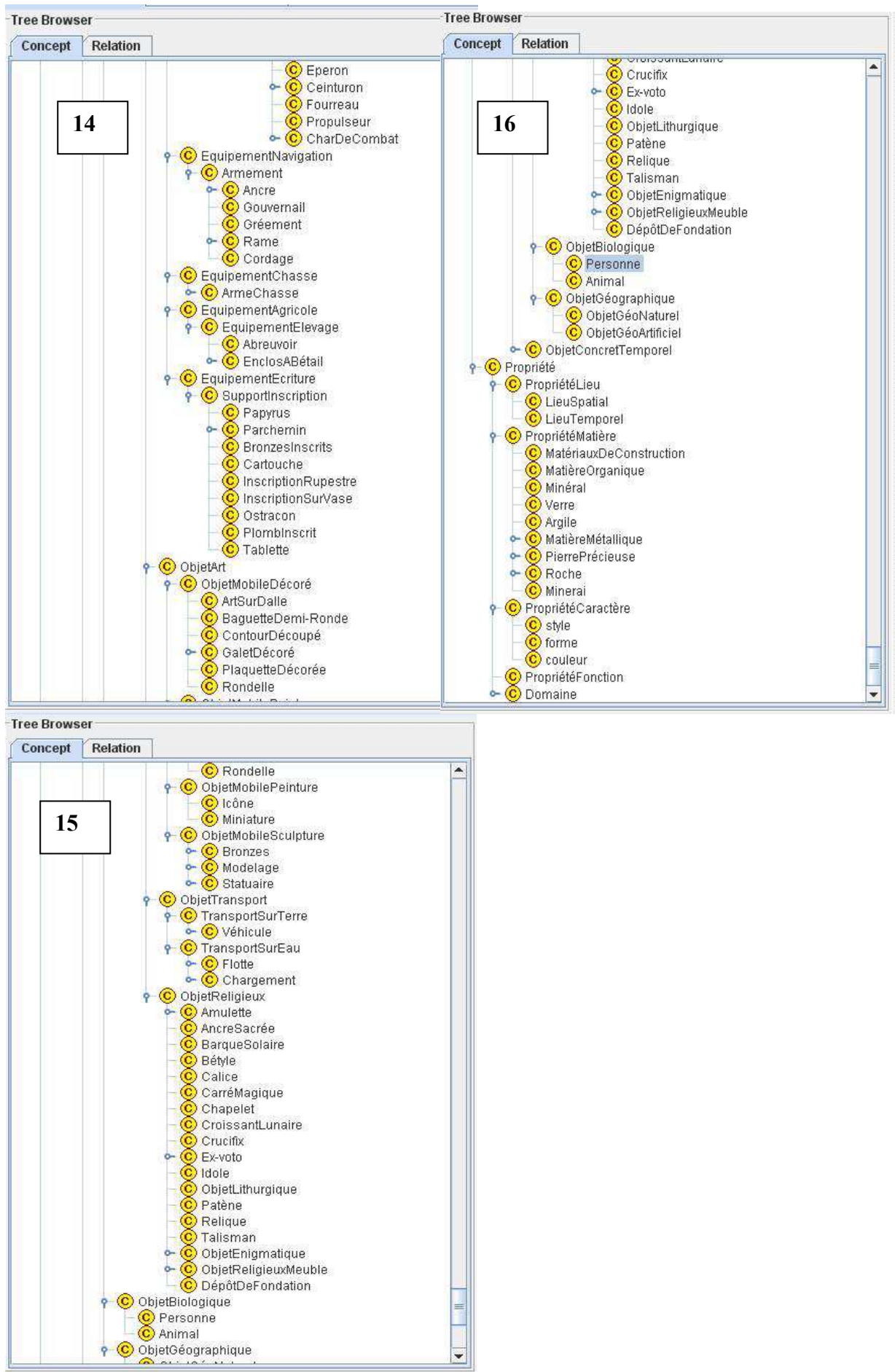
1











Annexe VII : Diapositives de la présentation des projets FRANTIQ pour l'Université d'été du TGE ADONIS

1




IdR2D-FRANTIQ

Projet innovant du TGE Adonis
Critère « Indexation sociale et contrôlée »
B. Lequeux et L. Marcheix



2



PACTOLS

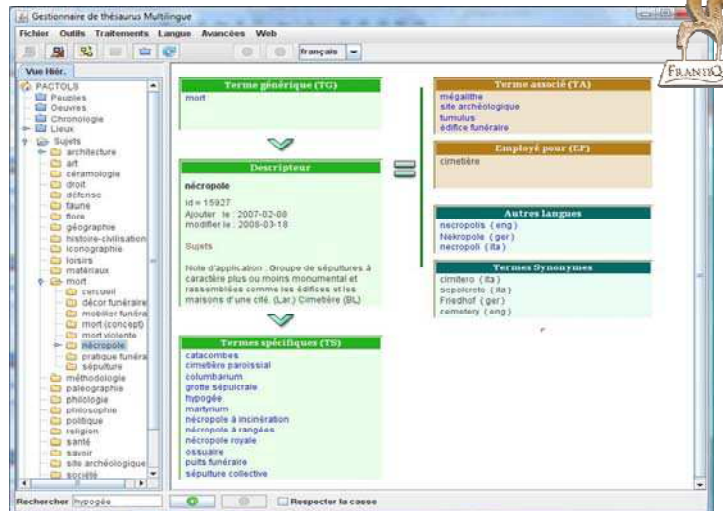
- **Thématique :**
 - Archéologie : de la préhistoire à l'ère industrielle
 - Sciences de l'Antiquité sous tous ses aspects
- **Thesaurus**
 - *Ensemble de termes normalisés fondé sur une structuration hiérarchisée et où les termes sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques*
- **Ontologie**
 - *Conceptualisation organisée en vue d'aboutir à un objet formel; les relations entre les concepts sont sémantiques mais formalisées.*

OpenTheso et OpenThesoWeb



- Critères de création
 - Autonomie par rapport aux BdD,
 - Multilinguisme, poly-hiérarchies
- Logiciel libre
 - Réalisé par Miled Rousset (MOM)
 - En cours de finalisation
- Fonctionnalités
 - OpenTheso : administration (gestion des termes du thésaurus évolutif et dynamique) et
 - OpenThesoWeb pour la consultation et la recherche

Administration



Consultation



6

Recherche

Open Theso Web De l'aide ?



Termes

Recherche

Termes : gallica

Options : Sujets Français

Choisissez une ou des lettres pour lancer la recherche

Hérarchique Alphabétique Remontée

Déroulant

- gaiese
- galène
- galène
- galier
- galet ardoisé
- galet décoré
- galet gravé
- galet peint
- galle
- gallicus
- étrusque gallo-belge
- monnaie gallo-belge
- monnaie gallo-ecloque
- gallo-grec (langue)
- étrusque gallo-romaine
- sigillée gallo-romaine ELP sigillée gauloise
- supplémentaire galvarique
- gauloise
- gauloise
- gauloise

Suivant

Termes(s) générique(s)

- zempart

↓

muros gallicus

Alus de défense gaulois. (BL)

NBR de notices: 4

Microthesaurus: Sujets

Crée le : 08/02/07 Modifié le : 08/02/07

Termes(s) Associé(s)

Enquêtes pour


Autres(s) langues(s)

- murs gallicus (eng)



7


Vers une ontologie



- **Apports**
 - Ajout de définitions pour lever les ambiguïtés
 - Relations plus développées
 - Modélisation plus rigoureuse des concepts
- **Limites**
 - Chronophage en définition et en modélisation (structuration)
 - Restriction à quelques branches du PACTOLS SUJETS

8

Vers une ontologie-2



- **Processus de transformation**
 - Modélisation de l'ontologie : DOE (*Differential Ontology Editor*)
 - Représentation et raffinement : Protégé
- **État d'avancement**
 - Déstructuration de la hiérarchie existante
 - Création de taxinomies à l'aide du logiciel DOE
 - Taxinomie des concepts
 - Taxinomie des relations

Taxinomie des concepts

Taxinomie des relations

Perspectives

- Amélioration de la recherche, indexation automatique
- Cette application peut servir à d'autres disciplines grâce à l'ontologie de haut niveau
- Intégration globale du thésaurus dans une ontologie
 - Par Archéotek
 - Par l'informaticien du réseau