Enjeux conceptuels pour l'élaboration d'un modèle ontologique : le cas du domaine minier

Amélie Daloz

▶ **To cite this version:**

Amélie Daloz. Enjeux conceptuels pour l'élaboration d'un modèle ontologique : le cas du domaine minier. Sciences de l'information et de la communication. Université de Lille, 2021. Français. NNT : . tel-03670559v1

HAL Id: tel-03670559

https://hal.science/tel-03670559v1

Submitted on 23 Dec 2021 (v1), last revised 17 May 2022 (v2)

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 DE LILLE

École doctorale S.H.S. - E.D. 473 - Laboratoire GERiiCO, UR 4073

ENJEUX CONCEPTUELS POUR L'ÉLABORATION D'UN MODÈLE ONTOLOGIQUE : LE CAS DU DOMAINE MINIER

Thèse pour obtenir le grade de

DOCTEURE DE L'UNIVERSITÉ DE LILLE

En sciences de l'information et de la communication

Présentée et soutenue publiquement par

Amélie DALOZ

le 15 Décembre 2021

Sous la direction de :

M. le professeur Stéphane Chaudiron

et le co-encadrement de :

M. le maître de conférences Bernard Jacquemin

COMPOSITION DU JURY :

Monsieur le professeur **Manuel Zacklad**, CNAM Paris, Rapporteur

Madame la professeure **Laurence Balicco**, Université Grenoble Alpes, Rapporteur

Monsieur le professeur **Christophe Roche**, Université Savoie Mont Blanc, Président

Madame la professeure **Céline Paganelli**, Université Paul-Valéry Montpellier 3, Examinateur

Monsieur le professeur **Stéphane Chaudiron**, Université de Lille, Directeur

Monsieur le maître de conférences **Bernard Jacquemin**, Université de Lille, Co-encadrant

Remerciements

Je n'aurais pu accomplir cette thèse sans l'aide de mon directeur de thèse, Stéphane Chaudiron, et mon co-encadrant de thèse, Bernard Jacquemin, qui ont tous deux su me guider avec bienveillance et avec une juste exigence à chaque étape de ce long parcours. Je sors grandie de cette épreuve. Merci encore d'avoir cru en moi et d'avoir aussi compris mes limites.

Merci à tous les membres du jury pour leur intérêt, leur relecture attentive et leur présence à ma soutenance. Merci à Manuel Zacklad, professeur au CNAM à Paris, et Laurence Balicco, professeure à l'Université Grenoble Alpes, d'avoir accepté d'être rapporteurs dans une période très chargée. Merci encore à Christophe Roche, professeur à l'Université Savoie-Mont-Blanc (pour m'avoir décerné le prix jeune chercheur Toth en 2018), et à Céline Paganelli, professeure à l'Université Paul-Valéry de Montpellier 3, pour avoir accepté d'être examinateurs.

Je remercie ensuite toute l'équipe du projet MémoMines avec qui j'ai eu la chance de collaborer. Un merci tout particulier à Virginie Blondeau pour son dynamisme et ses conseils bien utiles au début de la rédaction. Merci à Éric Kergosien pour sa disponibilité et ses compétences partagées lors de nos nombreuses réunions « ontologies » en visioconférence : « et ça, c'est un *E25* ? Non plutôt un *E22* parce qu'il y a la *P12* ! » et à Steffen Lalande et Abdelkrim Beloued pour leur aide précieuse avec la plateforme OKAPI.

Merci aux enseignants-chercheurs avec qui j'ai eu le plaisir d'échanger : Patrice de la Broise, directeur actuel du laboratoire GERiiCO, et Laure Bolka, maître de conférence, qui m'ont suivie tout au long de cette thèse. Un merci particulier à Claudio Gnoli et à Marcin Trzmielewski pour nos discussions lors de l'ISKO19 et nos échanges par mail qui ont agrandi ma connaissance des SOC. Merci aux professeurs et maîtres de conférence du laboratoire GERiiCO pour leur riche programmation de séminaires tout au long de ces quatre années qui m'ont permis de mieux comprendre les enjeux d'une discipline telle que les sciences de l'information et de la communication. Merci infiniment à Delphine Spileers, responsable administrative du laboratoire, pour sa gentillesse, sa disponibilité, ainsi que ses encouragements.

Merci à l'École Doctorale Sciences de l'Homme et de la Société, plus particulièrement à Catherine Maignant, Christine Hoët-Van Cauwenberghe et Sabrina Abed, pour leurs efforts quotidiens à destination des doctorants dans une période sanitaire difficile.

Merci à toutes les personnes (anciens mineurs, amis, famille de mineurs, passionnés) que j'ai pu rencontrer tout au long de ce parcours, qui m'ont livré une partie de leur histoire et ont enrichi ma connaissance sur le domaine de la mine qui m'était alors inconnu. Merci aux membres du Centre Historique Minier de Lewarde, au musée de l'école et de la mine de Harnes, aux amis de la mine de fosse 2, aux anciens mineurs de la fosse 9 et à ONYACUM à Oignies pour la qualité de leur médiation. Merci également à Vianney Haeuw pour ses précieuses remarques pour l'évaluation de ThesoMines.

Merci à mes amies rencontrées au GERiiCO : Kaouther, Fabiola, Zoé, Juliette, Louise, Marianne, Alice, Stéphanie, Leticia et les autres pour nos échanges et rigolades en salle des doctorants, nos balades à la Citadelle, nos JJC en présentiel et à distance.

Enfin, un énorme merci à ma famille. Merci à mes parents pour leur amour inconditionnel et leur soutien dans toutes mes prises de décisions. Je tiens à remercier ma mère de toujours croire en moi et de me comprendre aussi bien. Merci à mon père pour ses encouragements et son expérience. Merci enfin à mes frère et sœurs que j'ai hâte de retrouver.

Pour finir, mille mercis à toi, Stefano, pour ton immense soutien durant ces quatre années, pour ta patience dans les moments les plus difficiles, pour ton intelligence et point de vue sur la vie, qui m'ont permis de prendre du recul sur qui je suis et ont finalement permis de mener mon projet à terme.

Sommaire

REMERCIEMENTS **2**

SOMMAIRE **4**

RÉSUMÉ ET MOTS-CLÉS **6**

ABSTRACT AND KEYWORDS **7**

REMARQUES PRÉLIMINAIRES **8**

INTRODUCTION **9**

PARTIE 1. CADRE THÉORIQUE DE L'ORGANISATION DES CONNAISSANCES **14**

CHAPITRE 1. ORGANISATION, REPRÉSENTATION ET MODÉLISATION 15
1.1 Organisation des connaissances (OC) *15*
1.2 Représentation des connaissances (RC) *27*
1.3 Modélisation des connaissances *43*
CHAPITRE 2. APPROCHES DOCUMENTAIRES DE L'ORGANISATION DES CONNAISSANCES 49
2.1 Systèmes d'organisation des connaissances (SOC) *49*
2.2 Types et hétérogénéité des SOC *50*
2.3 SOC et Web sémantique *61*
CHAPITRE 3. APPROCHES ONTOLOGIQUES DE L'ORGANISATION DES CONNAISSANCES 63
3.1 Domaines d'étude des ontologies *64*
3.2 Types d'ontologies *67*
3.3 Travaux actuels sur les ontologies dans le champ de l'OC *70*
3.4 Rôle des ontologies *71*
3.5 Éléments d'un modèle ontologique *77*
3.6 Différentes approches de conception d'une ontologie de domaine *79*

PARTIE 2. LE PATRIMOINE MINIER COMME DOMAINE DE CONNAISSANCE **84**

CHAPITRE 4. DÉFINIR LE PATRIMOINE MINIER 86
4.1 Un patrimoine à plusieurs facettes *86*
4.2 Un patrimoine inscrit à l'UNESCO *88*
4.3 Un patrimoine langagier en danger *91*
4.4 Un patrimoine souterrain à rendre visible *94*
4.5 Un patrimoine documentaire hétérogène et dispersé *95*
CHAPITRE 5. VERS UNE MÉTHODOLOGIE DE CONCEPTION DES SOC DU DOMAINE MINIER 99
5.1 Constitution d'un corpus documentaire hétérogène *99*
5.2 Des ressources lexicales à la terminologie de la mine *120*
5.3 De la terminologie minière au thésaurus *124*
5.4 Du thésaurus au modèle ontologique *131*
5.5 Du modèle ontologique réduit à l'ontologie de domaine instanciée *142*
CHAPITRE 6. PRÉSENTATION DES TROIS SOC 144
6.1 Une terminologie de la mine : TerminoMines *144*
6.2 Un thésaurus du patrimoine minier : ThesoMines *147*
6.3 Une ontologie de la mine : OntoMines *149*

PARTIE 3. ENJEUX CONCEPTUELS DES SOC DU PATRIMOINE MINIER **158**

CHAPITRE 7. COUVERTURE INFORMATIONNELLE 159
7.1 Recouvrement et différenciation conceptuels *159*
7.2 Évaluation de la saturation *162*
7.3 Processus d'enrichissement : démarche évaluation et perfectionnement *165*
CHAPITRE 8. FINESSE DE DESCRIPTION OU GRANULARITÉ 170
8.1 Les entités nommées *172*

8.2 Les objets matériels ... *186*
8.3 Les activités ... *188*
CHAPITRE 9. EXPRESSIVITÉ DES FORMALISMES ... 192
9.1 La structure de la terminologie ... *192*
9.2 Les relations hiérarchiques du thésaurus ... *193*
9.3 Le réseau de l'ontologie ... *196*
CHAPITRE 10. INTEROPÉRABILITÉ ... 203
10.1 Interopérabilité syntaxique ... *203*
10.2 Interopérabilité sémantique ... *207*
CONCLUSION ... **211**
BIBLIOGRAPHIE ... **219**
ANNEXES ... **235**
LISTES ET TABLES ... **298**

Résumé et mots-clés

La thèse s'inscrit dans le cadre des recherches portant sur le web sémantique et l'interopérabilité des données. Son objectif général est d'analyser les enjeux conceptuels posés par la représentation des connaissances (*knowledge organization*) du patrimoine minier de l'ancien bassin du Nord–Pas-de-Calais (désormais Hauts-de-France). Trois systèmes d'organisation des connaissances (SOC) ont été réalisés, une terminologie (TerminoMines), élaborée à partir des nombreux lexiques, glossaires et dictionnaires existants, un thésaurus (ThesoMines), conçu sur la base de la terminologie et de l'analyse d'un corpus documentaire (essentiellement des entretiens d'anciens mineurs), et une ontologie de domaine (OntoMines) fondée sur le modèle CIDOC CRM. Ces trois SOC ont été réalisés dans la perspective de contribuer à la valorisation du patrimoine minier, matériel et immatériel, culturel et industriel, tel que défini par l'UNESCO et le TICCIH. Sur le plan conceptuel, la réalisation des trois systèmes a permis de mettre en évidence leur diversité et leur complémentarité à travers quatre critères : la couverture informationnelle, la granularité ou finesse de description, l'expressivité des formalismes et l'interopérabilité syntaxique et sémantique. Le caractère innovant de la thèse porte sur la démarche de conception et l'analyse simultanée des trois SOC à l'aide d'une approche qualitative et fine menant à des résultats réutilisables pour d'autres domaines. Le thésaurus ThesoMines est librement accessible et utilisable sur la plateforme Opentheso hébergée sur HumaNum. Ce travail doctoral a été financé dans le cadre du projet ANR MémoMines (ANR 16-CE38 0001) dont l'objectif est de valoriser la mémoire minière.

Mots-clés : Organisation des connaissances, patrimoine minier, web sémantique, terminologie, ontologie, thésaurus, bassin minier, Nord–Pas-de-Calais, HumaNum.

Abstract and keywords

The thesis is part of the research on the semantic web and data interoperability. Its general objective is to analyze the conceptual issues raised by the knowledge organization of the mining heritage of the former Nord–Pas-de-Calais basin (now Hauts-de-France). Three knowledge organization systems (KOS) have been developed: a terminology (TerminoMines), based on several existing lexicons, glossaries and dictionaries; a thesaurus (ThesoMines), based on the terminology and on the corpus analysis of many documents (mainly interviews with former miners); a domain ontology (OntoMines), based on the CIDOC-CRM model. The three SOC were carried out in the perspective of contributing to the valorization of the tangible and intangible, as well as the cultural and industrial, mining heritage, as it is defined by the UNESCO conventions and the Nizhny Tagil Charter for the Industrial Heritage by TICCIH. At the conceptual level, the realization of the three systems allowed to highlight their diversity and their complementarity through four criteria: the informational coverage, the granularity or fineness of description, the expressiveness of the formalisms and also the syntactic and semantic interoperability. The innovative character of the thesis concerns the design approach and the simultaneous analysis of the three SOC by using a qualitative and fine-grained approach leading to reusable results for other domains. The ThesoMines thesaurus is freely accessible and usable on the Opentheso platform hosted on HumaNum. The doctoral work was financed within the framework of the ANR project MémoMines (ANR 16-CE38 0001), whose objective is to participate in the valorization of the mining memory.

Keywords: Knowledge organization, mining heritage, semantic web, terminology, ontology, thesaurus, mining basin, Nord-Pas-de-Calais, HumaNum.

Remarques préliminaires

Cette thèse reprend des éléments qui ont fait l'objet de publications par l'auteure. Tous les liens hypertextes présents dans ce document étaient valides au 19/10/21.

À chaque fois qu'un acronyme français est disponible, il est préféré à son équivalent étranger. Nous utilisons donc SOC au lieu de *KOS* pour Systèmes d'Organisation des Connaissances. Les acronymes étrangers dans les citations ne sont en revanche pas traduits. La liste des acronymes et des sigles utilisés est disponible à la fin du document.

Les unités linguistiques, les termes, les instances, les descripteurs, les classes et propriétés ontologiques, ainsi que les mots, expressions et citations en langue étrangère sont écrits en italique. Les définitions issues du corpus de ressources lexicales sont notées entre guillemets simples (‘’).

Les légendes sont placées au-dessus des tableaux et en-dessous des figures. Les tables des tableaux et figures sont disponibles à la fin du document.

Les recherches issues de cette thèse ont été financées par l'agence nationale de la recherche (ANR) dans le cadre du projet ANR MémoMines[1].

[1] ANR-16-CE38-0001 https://memomines.hypotheses.org/

Introduction

La progressive disparition des derniers mineurs de l'ancien Bassin Houiller du Nord – Pas-de-Calais met en danger la transmission de leur mémoire et soulève des problématiques liées à la sauvegarde d'un patrimoine à résoudre dans un laps de temps toujours plus court. La thèse étudie la mémoire sous le prisme de l'organisation des connaissances et selon une approche hybride prenant en compte les dimensions linguistiques, extralinguistiques et métalinguistiques du domaine étudié. Les questions d'accès et de représentation des connaissances d'un patrimoine sont interrogées à travers la conception de trois systèmes d'organisation des connaissances (SOC – anglais *KOS*) : une terminologie, un thésaurus et une ontologie de domaine. Le travail, inscrit dans le cadre du projet ANR MémoMines participant à la valorisation de la mémoire minière, présente la méthodologie de construction des trois SOC à partir de plusieurs ressources documentaires hétérogènes sur le domaine. Outre créer de nouvelles connaissances et nouveaux terrains d'études, l'objectif du travail est de réfléchir au moyen de tirer le meilleur profit des nouvelles technologies, et notamment du web sémantique avec l'analyse des SOC selon quatre critères : couverture, finesse, expressivité des formalismes et interopérabilité. Les enjeux liés à la conception des SOC, à l'alignement avec des standards et les SOC en eux-mêmes sont des résultats de ce travail. L'analyse des documents sur le domaine montre une valeur essentielle dans la réalisation de tout travail : celle de partage basée sur des codes de communication communs au sein de différentes communautés d'acteurs. Par l'ambition de conceptualiser un domaine spécifique dans une période délimitée et révolue qui est celle de l'exploitation minière, les enjeux soulevés sont à poser dans le contexte (numérique) actuel et à transposer sur de nouveaux territoires ou même territoire en évolution. Cette thèse marque enfin l'importance de la représentation de la dimension sensible dans tout processus de transmission des savoirs.

Plus largement, la recherche s'insère dans les sciences de l'information, qui ont comme origine la rédaction en 1934 par Paul Otlet de l'ouvrage fondateur : *Traité de documentation. Le livre sur le livre* (Otlet 1934) et la fondation avec Henri La Fontaine du Mundaneum en 1895. L'ambition de rassembler, classer et partager toutes les connaissances du monde est novatrice à l'époque et est aujourd'hui plus que centrale.[2] En 1984, Le Coadic définit la science de l'information comme la « science qui étudie la communication de l'information » (Le Coadic 1984, 168). L'auteur en donne cette représentation thématique (Cf. figure 1), avec pour thèmes centraux : les théories de l'indexation (liées à la recherche de l'information, *information retrieval*, et, en particulier, celles qui concernent l'indexation automatique des documents textuels), la loi de Bradford (analyse bibliométrique relative à la performance des recherches bibliographiques dans les revues scientifiques), les systèmes de recherche en ligne « de

[2] « Connu pour être le père des sciences de l'information modernes, Paul Otlet fonde le Mundaneum en 1895 avec Henri La Fontaine (Prix Nobel de la paix en 1913). Son traité est la synthèse du travail qu'il réalise au sein de cette vaste entreprise destinée à rassembler, à classer et à partager toutes les connaissances du monde. Dans cet ouvrage novateur, il expose ses idées futuristes et visionnaires sur l'information envisagée en fonction de ses différents supports, son organisation et ses modes de circulation. Se basant sur les technologies en cours ou en projet à son époque, telles que la radio, le téléphone ou la télévision, il imagine une sorte de cerveau mécanique qui intègre du son et de l'image, permette de combiner les informations et de les partager grâce à leur mise en réseau. Ce texte qui fait encore référence se montre précurseur de l'informatique et d'Internet. » Paragraphe introducteur de l'ouvrage sur https://lesimpressionsnouvelles.com/catalogue/le-livre-sur-le-livre/

documents ou de données et les études descriptives des technologies correspondantes, banques de données bibliographiques, référentielles et textuelles, et banques de données factuelles », les techniques des bibliothèques (comme les techniques de constitution et de mise à disposition des collections en site, techniques en matière de bâtiments, pour l'accueil des lecteurs, l'organisation des fonds et des services, techniques de partage des ressources etc.), l'analyse des co-citations (la fréquence des citations reçues par un article est l'indice de l'importance scientifique de cet article) et des mots-associés (mots-clés utilisés lors de l'indexation d'un article qui sont retenus et analyse des co-occurences).

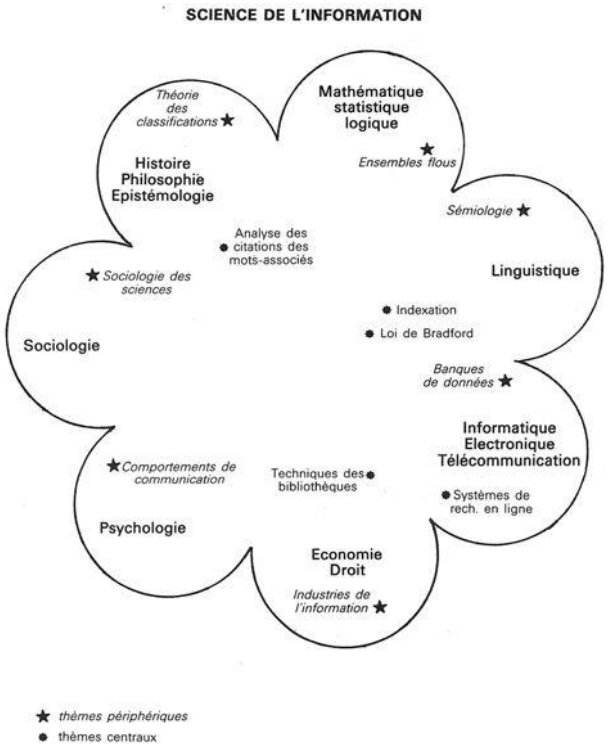


Figure 1 : Représentation des thèmes périphériques et centraux de la science de l'information par (Le Coadic 1984)

Le cadre dans lequel nos travaux se situent sont les théories de l'indexation, les systèmes de recherche en ligne, les techniques des bibliothèques et l'analyse des mots-associés que nous étudions du point de vue du champ d'étude de l'organisation des connaissances.

L'organisation des connaissances est un champ considéré par Dahlberg (Ohly 2016) comme une sous-discipline de la science de la science, avec une orientation de recherche descriptive. D'après elle, les champs d'application de l'organisation des connaissances ne sont pas seulement les sciences de l'information mais tous les domaines nécessitant des taxonomies, avec pour seule limite les objectifs de la classification (Ohly 2016). Dans nos travaux, nous verrons que le domaine du patrimoine minier est choisi comme champ d'application direct de l'organisation des connaissances. Hjørland liste les principales activités de l'organisation des connaissances telles que : la description de document, l'indexation, la classification, activités effectuées dans les lieux de savoirs[3] (Hjørland 2008, 86). Comme champ d'étude, le domaine s'intéresse à la nature et à la qualité des systèmes d'organisation des connaissances (SOC ou *KOS* en anglais) : « *KO as a field of study is concerned with the nature and*

[3] Lieux de savoirs tels qu'ils sont définis dans (Jacob 2014)

quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations, works and concepts » (Hjørland 2008, 86). La nature des SOC est constamment redéfinie par les auteurs. Nous verrons que récemment, Zacklad en propose une nouvelle typologie : il identifie les SOC symboliques, les SOC algorithmiques, les SOC visuels, les SOC incarnés et insiste surtout sur l'importance « d'articuler de manière fine et complémentaire » ceux-ci « pour rendre compte de la diversité croissante des usages dans les environnements postdigitaux actuels » (Zacklad 2018, 103).

En ce qui concerne le terrain d'application, les données de la thèse sont issues des travaux effectués dans le cadre du projet ANR MémoMines[4] visant la sauvegarde de la mémoire minière de l'ancien bassin houiller des départements du Nord et du Pas-de-Calais. En décembre 1990, la remontée du dernier morceau de charbon marque l'arrêt définitif de l'activité houillère du Bassin minier du Nord–Pas-de-Calais. Ce morceau de charbon, comme de nombreux termes techniques de la mine, a une appellation spécifique pour les mineurs de fond : c'est une « gaillette », en ch'ti[5], diminutif de gaille « noix « (Guiraud 1978). Si l'activité cesse en raison de l'épuisement du gisement, du coût de revient et de la concurrence d'autres sources d'énergie, elle reste ancrée dans les mémoires de ceux qui l'ont vécue. Aux mémoires individuelles s'ajoute une riche documentation aussi bien technique, scientifique, pédagogique que culturelle, composée de rapports d'activité des compagnies minières, de rapports de recherche, d'illustrations, mais aussi d'œuvres fictives, de romans, de films contant l'histoire des mineurs dans les moindres détails. Cette documentation, dispersée sur le territoire, détenue par plusieurs institutions différentes, indexée ou non par des langages souvent hétérogènes est parfois difficilement accessible. En 2012, après un long travail débuté en 2003 par les acteurs du patrimoine minier, le Bassin minier du Nord–Pas-de-Calais est inscrit au titre de *paysage culturel évolutif vivant* sur la liste de l'Unesco avec 353 éléments représentatifs. Ce sont des infrastructures d'exploitation, des bâtiments d'habitat, des terrils, des fosses qui ont été inventoriés, analysés en tant qu'objets de patrimoine. Aujourd'hui, le patrimoine immatériel et son élément constitutif principal, la langue comme vecteur (UNESCO 2003a) est quant à lui en danger. Dans ce cadre, le projet ANR MémoMines a pour objectif de valoriser scientifiquement la mémoire minière grâce à des approches en sciences de l'information et de la communication. Les objectifs du projet sont multiples[6]. L'un d'entre eux est de créer des outils permettant d'accéder au corpus vidéo lié à la mémoire de la mine en cours de constitution[7]. Nous proposons pour cela la conception de trois SOC : une terminologie, un thésaurus et une ontologie de domaine. Dans le contexte de l'*open science*, le corpus vidéo ainsi que les SOC seront mis à disposition des différentes communautés, professionnelles et amateures, et plus particulièrement les professionnels de l'information culturelle et scientifiques (musées, historiens, linguistes…). Nos travaux consistent

[4] https://memomines.hypotheses.org/
[5] Le chti, ch'ti, chtimi ou patois du Nord est l'appellation donnée au picard parlé en région Nord–Pas-de-Calais.
[6] https://memomines.hypotheses.org/
[7] https://medihal.archives-ouvertes.fr/MEMOMINES

donc à indexer, décrire et représenter les connaissances issues de documents hétérogènes selon une approche intégrant la réalité linguistique, extralinguistique et métalinguistique du domaine d'étude.

Plus spécifiquement, l'objectif de nos travaux est d'étudier les spécificités de chacun des trois SOC et d'interroger leur fonctionnement au regard de critères tels que la couverture informationnelle, l'interopérabilité, la finesse de description dont le traitement des entités nommées et l'expressivité des formalismes. Ces critères ont été identifiés comme potentiellement pertinents pour répondre aux problématiques que nous développons ci-dessous.

Notre ambition est de pouvoir représenter une vision globale du monde de la mine du bassin houiller du Nord–Pas-de-Calais véhiculée par une étude à la fois linguistique, métalinguistique et extralinguistique de corpus hétérogènes. Nous nous demandons si et comment cette vision peut être utilisée pour conceptualiser le domaine étudié, puis comment généraliser sans omettre la richesse du domaine. La couverture informationnelle et la saturation des modèles réalisés sont étudiées dans ce cadre (Cf. chapitre 7).

Pour être effective, une représentation doit pouvoir être cadrée et avoir des limites. Nous faisons l'hypothèse que la prise en compte des entités nommées tels que les acteurs, des entités spatiales et temporelles permet de rendre effective cette limite. Nous étudions plus précisément les références à celles-ci dans les SOC, puis dans un entretien avec un ancien mineur (Cf. chapitre 8).

Le fait de pouvoir partager les connaissances identifiées sur le web sémantique permettrait de pouvoir les comparer plus facilement avec celles contenues et décrites par et dans d'autres systèmes utilisant les standards du web sémantique et de réaliser des alignements entre connaissances équivalentes ou des rapprochements entre connaissances liées. Pour cela, le modèle conceptuel de référence dans le domaine du patrimoine, le CIDOC CRM, est utilisé. Grâce à son langage, il devra être possible d'établir des inférences entre connaissances pour en créer d'autres et alimenter ainsi la vision du domaine en même temps qu'il évolue. Le modèle sera donc étudié sous plusieurs angles : selon sa granularité (Cf. chapitre 8), c'est-à-dire sa finesse de description, sur l'expressivité de son formalisme (Cf. chapitre 9) et sur son interopérabilité syntaxique et sémantique (Cf. chapitre 10).

Nous souhaitons en effet mettre en lumière les connaissances qui ne peuvent pas être exprimées par les modèles de référence. Le problème d'alignement de connaissances spécifiques avec des SOC plus larges est étudié. Notre découpage des connaissances est réalisé au niveau lexical. De nombreuses ressources définissent le vocabulaire minier. Sur la base de celui-ci, nous visons à établir une ontologie de domaine et voulons montrer l'importance d'étudier le lexique utilisé par des communautés expertes ayant vécu le domaine (sachant que plusieurs niveaux de vie du domaine existaient : travail au fond, travail au jour), le lexique étudié systématiquement par des spécialistes de la langue, le lexique des professionnels de la documentation, le lexique des professionnels des archives, le lexique des professionnels des musées, et le lexique utilisé par des profanes, passionnés par le domaine et s'exprimant sur le web. Nous voulons

montrer pourquoi il est difficile d'établir des équivalences entre tous ces lexiques et revenir sur les types de lexiques plus à même d'aider à concevoir une ontologie.

Dans cette thèse, plusieurs problèmes sont ainsi soulevés et traités dont l'identification de concepts dans un document et leur « traduction » en termes-concepts, la structuration puis la représentation sémantique de concepts pour indexer non plus un seul fonds documentaire, mais tous documents sur un même domaine. Nous souhaitons ainsi montrer la complémentarité des systèmes.

Dans une première partie (partie 1), nous exposons comment les notions d'organisation, de représentation et de modélisation sont utilisées dans les travaux anciens et actuels pour appréhender (et visualiser) un domaine complexe dans sa globalité. Puis, nous présentons diachroniquement et synchroniquement les outils de représentations des connaissances traditionnels et soulevons les enjeux actuels posés par les ontologies. Dans une deuxième partie (partie 2), nous présentons et délimitons le terrain et le domaine d'application de la recherche qui est le patrimoine minier, puis décrivons la méthodologie de conception des trois outils de représentation des connaissances : terminologie, thésaurus et ontologie de domaine que nous présentons finalement d'une manière factuelle. Dans une troisième partie (partie 3), nous montrons grâce à une analyse comparative comment les SOC sont de véritables outils pour valoriser un patrimoine. Pour démontrer ceci, quatre critères info-documentaires sont relevés pertinents : la couverture informationnelle, la granularité de description, l'expressivité des formalismes et l'interopérabilité. Outre la création des systèmes disponibles en accès libres et leurs critères d'évaluation, nous ouvrons sur la transférabilité de notre approche à de nouveaux domaines.

PARTIE 1. Cadre théorique de l'organisation des connaissances

CHAPITRE 1. Organisation, représentation et modélisation

L'essentiel, c'est toujours de distinguer. (Werner 2006)

Ce chapitre insère le contexte de l'étude qui est le champ de l'OC (1.1). Il se focalise tout d'abord sur trois enjeux du champ au regard de travaux réalisés lors de ces dix dernières années (1.1.1). Nous présentons ensuite les différentes approches identifiées par Hjørland (Hjørland 2008) et développons celles qui intéressent nos travaux (1.1.2). Hjørland définit le champ de l'OC en séparant les processus, des systèmes d'organisation des connaissances. Nous nous intéressons particulièrement aux SOC et les définissons à travers des typologies récentes. Les SOC répondent à des enjeux de représentation des connaissances, notamment au niveau du langage à utiliser. Nous commençons par donner différentes définitions et exemples de la dualité représentation mentale (1.2.1) représentation graphique (1.2.2) des connaissances. Les préoccupations actuelles se placent au niveau d'une représentation globale des connaissances et nous voyons ce que propose le web sémantique dans ce cadre (1.2.3), puis présentons les langages de représentation en vigueur (1.2.4). Nous questionnons finalement la représentation de l'aspect sensible des connaissances (1.2.5). Enfin, nous choisissons de développer un type de représentation, qui est la modélisation (1.3) et nous expliquons ensuite en quoi la structuration des connaissances est nécessaire (1.3.1) puis développons les différents problèmes que posent les ontologies (1.3.2) et notamment la modélisation, la généricité, la linguisticité des connaissances. Nous posons enfin la question des ontologies en documentation.

1.1 Organisation des connaissances (OC)

De 1895 et jusqu'à sa mort (en 1944), Paul Otlet, bibliographe belge, met en œuvre une théorie globale de la documentation. Une des premières occurrences de l'expression « *Organization of Knowledge* » est située dans un de ses écrits (Otlet 1903, 74). Celui-ci la distingue alors de la documentation. Jugeant l'expression imprécise, il s'attache à la définir. Il applique celle-ci à l'ensemble des faits et des considérations, extérieurs à la connaissance, impliquées dans la vie, qui participent à l'organisation de la connaissance et tout ce qui est né de la création, de la conservation, de l'enseignement, de la diffusion de cette connaissance[8]. Dans ce cadre *knowledge* est défini comme la systématisation de ce qui est connu sur un sujet donné (Otlet 1903, 74). Comme exemple de sujet (*matter*) pouvant faire partie de cette *Organization of Knowledge*, l'auteur donne l'enseignement d'une discipline dans toutes ses dimensions (chaires spéciales, cours pratiques et séminaires, conférences de vulgarisation, instituts de recherche et tous les moyens d'encourager l'étude d'une science : concours, bourses d'études, bourses de voyage, etc.). Le sens est donc large.

[8] « Documentation should be distinguished with no less care from the Organization of Knowledge. This imprecise expression, which it is convenient to use in the absence of a unique term with a definite meaning, is here applied to the whole of the facts and considerations which are involved in the life and external organisation of knowledge itself, everything that has risen from creating it, preserving it, teaching it, disseminating it » (Otlet 1903, 74)

Quelques années plus tard, Henry Evelyn Bliss, bibliothécaire américain, utilise la même expression « *Organization of Knowledge* » dans ses deux ouvrages (Bliss 1929, 1933) qu'il confronte au système des sciences en 1929 et qu'il étudie en bibliothèques avec une approche thématique des livres (1933). Le domaine d'investigation alors constitué est établi principalement sur des questions relatives aux systèmes de classification et aux systèmes de relations conceptuelles tels que les thésaurus (Murguia et De Sales 2013). Dans la lignée des travaux de Bliss, un champ d'étude international se forme et Ingetraut Dahlberg crée en 1973 la revue *Knowledge Organization*[9] qui deviendra une revue bi-mensuelle officielle de la *International Society for Knowledge Organization* (ISKO), dont Dahlberg sera la première présidente. En France, l'organisation des connaissances est une discipline étudiée sous le prisme des sciences de l'information et de la communication (Couzinet 2012). En 2013, Murguia et De Sales (Murguia et De Sales 2013) s'attachent quant à eux à réfléchir aux associations théoriques possibles entre les concepts de Bliss et d'Otlet ainsi que celle de l'ISKO que nous définissons ci-dessous. Les auteurs montrent que Dahlberg a accordé une stature particulière à l'œuvre de Bliss, ce qui implique qu'elle a, avec des auteurs ultérieurs, développé l'idée et l'applicabilité de l'œuvre, faisant de Bliss le fondateur d'un discours, emprunté et réinterprété par ISKO. Les auteurs montrent également qu'Otlet n'est pas le fondateur d'un discours lié à l'organisation des connaissances tel qu'il existe dans ISKO, mais que ses travaux sont présents dans les propos de Hjørland.

La conférence internationale ISKO, créée en 1989 est entièrement dédiée aux échanges et questionnements sur la discipline de l'OC. Comme nous pouvons le voir sur le site web dédié, une trentaine de pays sont représentés[10]. De nombreux actes de colloque sont écrits suite à chaque événement organisé, représentant une partie de l'évolution du champ. Parmi ceux-ci, des travaux récents traitent du futur de la discipline ainsi que celui de l'organisation de l'information (Ohly 2016) lors de la 13ème conférence en 2014 à Cracovie, ou de la théorie, de la sémantique et de l'OC comme l'ouvrage de Babik, Ohly, et Weber (Babik, Ohly, et Weber 2017) qui rassemble les actes de plusieurs événements organisés en Allemagne. Comme nous pouvons le lire sur le site web dédié, la mission de l'ISKO est :

> « […] *to advance conceptual work in knowledge organization in all kinds of forms, and for all kinds of purposes, such as databases, libraries, dictionaries and the Internet*.[11] »

La visée de cette mission est donc assez large, d'après Favier, l'OC concerne même l'étude et la recherche en général :

> « L'organisation des connaissances a pour rôle de produire un langage commun normalisé entre des documents et contenus divers dont le regroupement est nécessaire pour l'étude et la recherche. » (Favier 2017, 6)

[9] Cf. https://www.isko.org/ko.html
[10] https://www.isko.org/ mentionne : Le Brésil, le Canada et les Etats-Unis, la Chine, la France, l'Allemagne, l'Inde, l'Iran, l'Italie, la Tunisie, l'Algérie, le Maroc, la Pologne, Singapour, l'Espagne, le Portugal, le Royaume-Uni, le Nigéria, le Cameroun, la Belgique, les Pays-Bas, le Luxembourg, l'Australie, l'Autriche, la Georgie, l'Europe du Nord, la Russie et la Slovaquie
[11] https://www.isko.org/about.html

Ceci rejoint les propos de Dahlberg dans (Ohly 2016) qui considère l'OC comme une métascience, donc une science des sciences.

La revue *Knowledge Organization*[12] possède cinq grands objectifs :

> « (1) *clarify theoretical foundations (general ordering theory, philosophical foundations of knowledge and its artifacts, theoretical bases of classification, data analysis and reduction)* ;
>
> (2) *describe practical operations associated with indexing and classification, as well as applications of classification systems and thesauri, manual and machine indexing* ;
>
> (3) *trace the history of knowledge organization* ;
>
> (4) *discuss questions of education and training in classification* ;
>
> *and* (5) *problems of terminology in general and with respect to special fields.* »[13]

Si la visée de l'OC est assez large, elle possède tout de même un sens particulier, c'est-à-dire du point de vue des bibliothèques et de la science de l'information. Hjørland divise en 2008 le champ en deux objets : les *Knowledge Organizing Processes* (*KOP*) traduits par *processus d'organisation des connaissances* (POC) et les *Knowledge Organizing Systems* (*KOS*) traduits par *systèmes d'organisation des connaissances* (SOC) :

> « *Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of "memory institutions" by librarians, archivists, information specialists, subject specialists, as well as by computer algorithms and laymen.* » (Hjørland 2008, 86).

Nous représentons les propos de Hjørland sur le schéma de la figure 2 ci-dessous et sur lequel nous identifions cinq catégories autour desquelles gravitent les questionnements du champ de l'OC dans ce sens particulier : les activités, les agents, les lieux, les objets. Sont entourés en orange les objets sur lesquels cette thèse se focalise : les SOC qui organisent les concepts, les documents et les représentations des documents. Les liens qui relient les catégories et entités sont des relations d'instance (*de type*), relations d'objet (*de*), relations sémantiques spécifiques (*s'intéresse à, consiste en, effectuées par/dans, utilisés pour*).

[12] Cf. https://www.isko.org/ko.html
[13] https://www.isko.org/ko.html

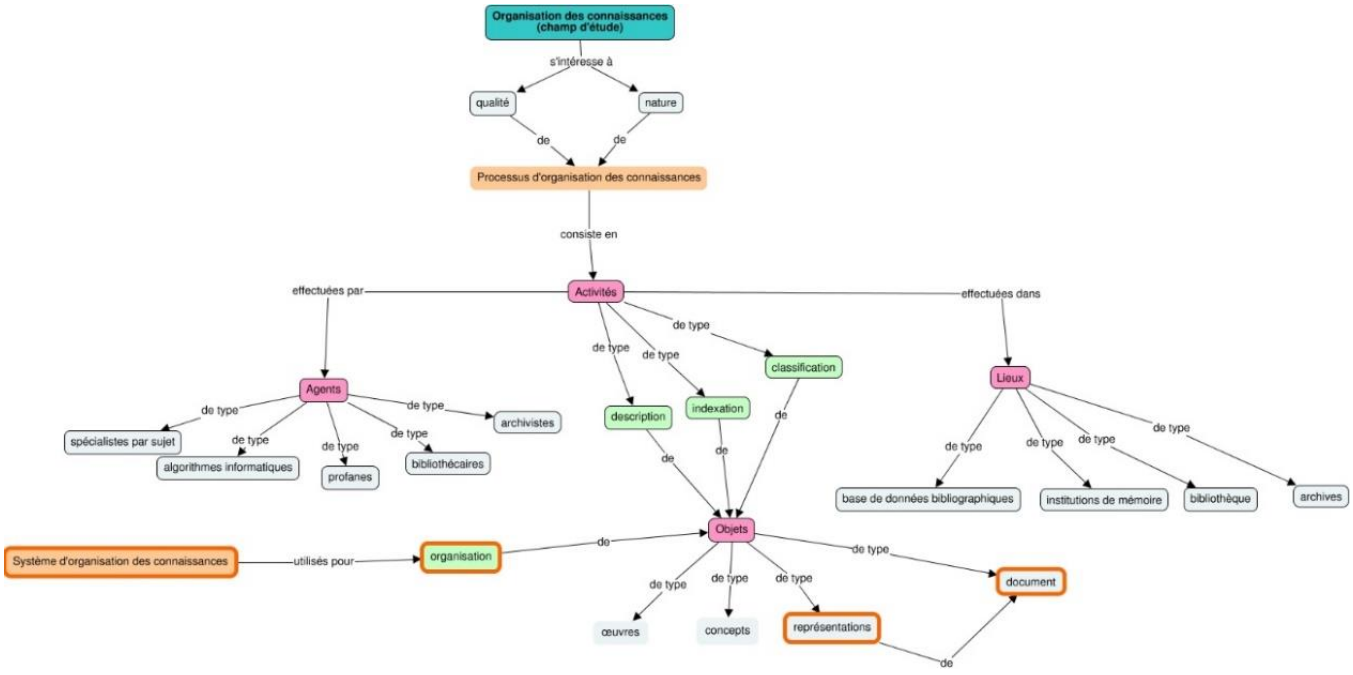


Figure 2 : Schéma représentant la définition de l'OC issue de (Hjørland 2008) du point de vue particulier de la LIS disponible à l'adresse : https://cmapscloud.ihmc.us:443/rid=1WV9S58GL-1HMJC51-VV987J

À propos du sens particulier et du sens large de l'OC (*social division of mental labor*) Hjørland nous met en garde sur le fait que : « *There exists no closed "universe of knowledge" that can be studied by KO in isolation from all the other sciences' study of reality* » (Hjørland 2008, 87). En d'autres termes, il faut toujours étudier les objets de l'OC en contexte, avec des théories appropriées. D'après l'auteur, le fait que l'OC s'insère directement dans la théorie de la connaissance en placant au centre le concept de *connaissance* peut permettre de ne pas se laisser dépasser par d'autres champs d'étude comme la gestion de la connaissance et l'informatique qui jouent déjà un rôle important dans les futurs environnements de communication et d'échange de connaissances. Le champ de l'OC fait en effet face à des enjeux collaboratifs, virtuels et éthiques et d'ouverture que nous développons dans la section suivante.

Enfin, en ce qui concerne le concept de connaissance, Agostinelli le rappelle, c'est « la façon dont on pense que les connaissances se construisent [qui] induit un choix dans les modèles d'organisation et de gestion » (Agostinelli 2012, 7). L'auteur met ainsi en opposition deux approches : une cartésienne et une kantienne. La première possède une vision rationnelle, qui va du simple au complexe et dans ce cas les connaissances ont une structure existante, unique et invariante. Le modèle sémantique entre dans ce point de vue. Dans la deuxième approche, les connaissances sont le résultat de constructions mentales. C'est le modèle sémiotique. En 2012, Courbières montre l'importance de la prise en compte de la sémiotique dans la théorie de la documentation (Courbières 2012).

Dans nos travaux, le concept de **connaissances** se définit à travers le concept d'information. La figure 3 que nous avons trouvée sur le web (@gapinvoid[14]) illustre de manière plaisante la différence entre

[14] https://www.gapingvoid.com/

information et *knowledge* : si les nœuds représentent l'information, elle est alors une disposition de plusieurs unités colorées de même forme. La connaissance (*knowledge*) est représentée par ces mêmes unités (nœuds de même couleur, même emplacement, même forme) avec l'ajout de liens perpendiculaires ou horizontaux reliant d'une certaine façon ces unités entre elles. Cette représentation, comme toute représentation, mène à une interprétation, elle n'est donc pas fiable si elle n'est pas accompagnée d'un contexte. D'après nous, le schéma de la figure 3 pourrait aussi bien représenter l'approche cartésienne que l'approche kantienne. La représentation ne dit presque rien sur la construction des connaissances et laisse libre l'interprétation.

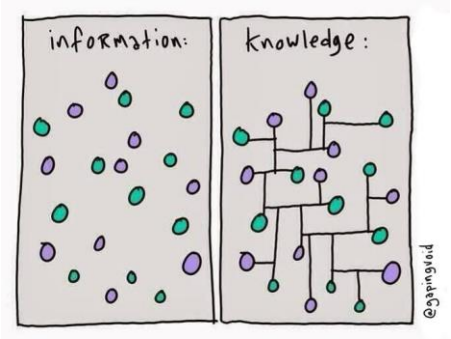


Figure 3 : Schéma représentant la différence entre information et knowledge @gapinvoid

Nous nous appuyons sur la définition d'information que donnent Arsenault, Da Sylva, Forest, et *al.* D'après ces auteurs, l'information est :

> « un ensemble de données dont la structure et le contexte permettent de véhiculer un sens, et donc susceptible d'informer la personne à qui les données sont transmises. Étant dès lors informé, l'individu est en mesure – en traitant et en assimilant cette information et aussi en établissant des liens avec d'autres informations acquises et emmagasinées auparavant – de générer des connaissances » (Arsenault et *al.* 2009, 105).

Pour comprendre cette définition il est nécessaire de revenir sur les termes *donnée*, *structure*, *contexte*, *sens*, *lien* et *connaissance* que nous replaçons rapidement dans leur contexte scientifique, puis dans celui de notre étude.

La donnée est une « [r]eprésentation d'une information sous une forme conventionnelle destinée à faciliter son traitement[15] » (Boulogne et Institut national des techniques de la documentation 2004, 88). De par son caractère formel et la notion de traitement, le terme a largement été utilisé au sein des sciences dites « dures » à tel point que la production scientifique s'est transformée en production de données menant aux données massives ou *big data* et engendrant un certain nombre de problématiques, dont celles de traitement (automatique). Dans les sciences humaines et sociales, le terme a mis un peu de temps à être utilisé, les chercheurs ayant du mal à se l'approprier. Schöch redéfinit la donnée dans ce champ et introduit la notion de *smart data* (que nous traduisons par données intelligentes) qu'il compare aux *big data* : « *Smart data is data that is structured or semi-structured; it is explicit and enriched,*

[15] JO du 17/01/1982, arrêté du 22/12/1981 relatif à l'enrichissement du vocabulaire de l'informatique.

because in addition to the raw data, it contains markup, annotations and metadata. » (Schöch 2013, 4). Dans ce cadre, des réflexions sur l'hétérogénéité des données issues du numérique et plus récemment du web et ses implications sur la création de corpus, mais aussi sur le traitement des métadonnées sont traitées dans (Kergosien et *al.* 2017) ou (Caron, Defiolle, et Lay 2019). Les données ont donc vocation à être structurées comme le mentionnent Mehta et Sahni (Mehta et Sahni 2005). Médiatiquement et en raison de la valeur marchande actuelle qu'elles ont, elles sont parfois présentées comme le « pétrole » du XXIème siècle (mais cette proposition est largement critiquée).

La structure fait référence à un ensemble cohérent et construit, à l'agencement d'unités formant un système. Dans le domaine de l'informatique, une structure est un « type d'organisation et de représentation des données en rapport avec une application ou un langage spécifiques ». Les domaines de la linguistique et de l'informatique travaillent tous deux sur la notion de structure ce qui les rendent pertinents pour l'analyse des données, de l'information puis de la connaissance. La structure répond en effet à des enjeux d'organisation et de représentation pour créer de la connaissance.

Le contexte représente l'environnement sensible, perceptible. L'enjeu d'une réception équivalente à une production initiale et intentionnelle est contenu dans la notion de contexte. De nombreux travaux insistent sur l'importance de la prise en compte du contexte lors de la réutilisation des données. Le contexte permet en effet de reconstituer le sens initial de production des données. Celui-ci peut être noté sous la forme de métadonnées, c'est-à-dire, des données sur les données.

D'après le TLFi[16], le sens est une « idée, signification représentée par un signe ou un ensemble de signes » ou une « représentation intelligible évoquée ou manifestée par un signe ou une chose considérée comme un signe ». Dans cette acception, la question du signe ainsi que celle de la représentation sont soulevées. La représentation étant centrale dans nos travaux, nous la traitons en 1.2 puis nous revenons sur celle du signe ci-dessous :

La question du signe est posée dans (Eco et Klinkenberg 1988) et est l'objet d'étude de la sémiotique, approche utilisée en sciences de l'information-documentation et en linguistique. Nous puiserons dans cette dernière discipline et plus particulièrement en linguistique générale, pour utiliser le terme comme un signe linguistique (Gaudin 2005), résultat d'une association entre un sens et une forme (L'Homme 2005), ou entre un signifiant et un signifié pour Saussure. Le signe est issu de la tradition aristotélicienne où « *le sens d'un signe est conçu comme représentation mentale (concept), et défini par ce à quoi il renvoie dans le monde (le mot « chien » « signifie » ce quadrupède à poils ras)* » Zacklad (2007, 7) cite Lacour (2004), lui-même citant Rastier (1997)[17]. Nous pensons en revanche comme Béjoint et Thoiron (2000) que le concept ne se résume pas au signifié[18]. La notion de concept est traitée en 1.2.1 et 3.5.1. Nous ne nous occuperons pas des problématiques posées par ceci, comme la remise en question de l'arbitraire du signe (qui postule que prévoir comment se créerons les relations entre désignations et

[16] TLFi : Trésor de la Langue française informatisée http://atilf.atilf.fr/
[17] Ces nombreuses re-citations montrent comment l'information traverse les différents documents et illustre le renouvellement de l'accès à cette information.
[18] L'ambiguïté de la relation entre signe et concept remonterait à Saussure. (Béjoint et Thoiron 2000)

choses désignées est impossible). En sémiotique, Pierce place le signe au milieu de trois éléments relevant de la syntaxe, de la sémantique et de la pragmatique : le Representamen, l'Objet et l'Interpretant[19]. Les travaux de thèse de Bui (2003) traitant de la structuration sémantique des contenus des documents audiovisuels selon les points de vue de la production, utilisent cette approche. Nous verrons que le modèle sémiotique s'oppose au modèle sémantique selon des approches kantienne ou cartésienne, ceci dépendant de notre point de vue sur la construction des connaissances (Agostinelli 2012). Plus précisément, nous considérerons dans cette thèse les enjeux d'utilisation d'un modèle sémantique pour les organiser et les représenter. Dans le cadre de la notion de document, celle de signe signifie « texte analysable » pour Müller (2008) qui définit le document non plus comme rapport entre un fond et une forme, mais entre « un fonds et […] une structure (langage informatique), à la fois comme forme (ensemble de données organisées selon une structure stable), comme signe (un texte analysable), comme médium enfin (la trace de relations sociales reconstruite par les dispositifs informatiques) » (Müller 2008, 38). Enfin, nous verrons que les signes peuvent être enregistrés par une classe d'un modèle ontologique, le CIDOC CRM, sous la classe *E90 Symbolic Object* qui « comprend des ensembles de signes de toute nature, qui peuvent servir à désigner quelque-chose ou à communiquer un certain contenu propositionnel » (Du Château 2010, 86).

Ensuite, les choses existent en lien avec des autres. En fonction de la nature des choses reliées entre elles et la nature du lien, on parle de relations ou de propriétés (Cf. 3.5.2), de facettes (Cf. 2.2.2.1) et même de colle sémantique (*semantic glue*) comme dans (Binding, May, et Tudhope 2008, 281). Nous utiliserons le terme lien comme Paquette, qui le définit comme une :

> « Composante d'un modèle des connaissances qui met en relation deux connaissances, la première étant l'origine du lien et la seconde sa destination. Dans un modèle graphique, un lien est représenté par une flèche allant de l'origine vers la destination. Cette flèche est surmontée d'une lettre indiquant le type de liens qu'elle représente » (Paquette 2002, 341).

L'auteur donne comme exemple le lien d'instanciation (I), de composition (C), de spécialisation (S), de précédence (P). Nous aurons l'occasion de revenir sur ces types de liens lors de la construction du thésaurus et de l'ontologie. L'utilisation de l'appellation « signe passeur » au lieu de « lien », concernant le lien hypertexte est traitée dans (Davallon et Jeanneret 2004) (Cf. 1.2.3 pour le traitement de l'hypertexte dans un contexte de web sémantique).

Le concept de connaissance se définit donc comme résultat de ce processus et dans notre cas, les données sont tout d'abord de nature lexicales (identifiées dans des documents). Nous les structurons (grâce à des structures existantes) étudiées sous le prisme de théories linguistiques et terminologiques (conceptuelles) afin qu'elles aient du sens. Leur contexte est documentaire et vise des récepteurs (humains ou machines) dont une typologie reste à produire. Nous identifions les liens qu'ont ces

[19] Le Representamen est la forme que prend le signe, l'Objet correspond au référent (objet réel ou imaginaire ou état du monde) auquel renvoie le Representamen. L'Interpretant est le sens donné au signe au moment de sa perception. (Bui 2003)

informations avec d'autres dans les documents pour créer des connaissances, qui, à leur tour, doivent être organisées de la même manière (avec une structure conceptuelle qui donne du sens, un contexte favorable à des récepteurs spécifiques (aussi bien scientifiques comme les chercheurs en linguistique, en histoire, que des professionnels de la documentation, des musées, des archives), puis une formalisation des liens entre ces informations pour pouvoir générer d'autres connaissances.

1.1.1 Enjeux actuels et généraux de l'organisation des connaissances

Nous ne reviendrons pas sur tous les enjeux soulevés par l'OC mais nous notons ci-dessous trois enjeux génériques soulevés lors des dix dernières années dans ce cadre et qui concernent nos travaux :

Le premier est collectif et collaboratif. En 2012, Beau (2012, 9) écrit que l'enjeu de l'organisation des connaissances se situe au niveau de la conception collective de la connaissance[20]. Cet enjeu se place dans des enjeux plus larges tels que la mondialisation et la globalisation, répondant eux-mêmes à des problématiques culturelles et politiques d'une part et économiques d'autre part. Pour cela, le partage et l'ouverture des connaissances favorisant la conception collective nécessitent un cadre de réflexion à propos des aspects multilingues et éthiques.

Le deuxième est culturel et éthique. Favier (2017) traite du terme *infrastructure* dans le domaine des humanités numériques et place quant à elle « [l]'intérêt porté à la diversité culturelle et à un fondement éthique (*ethical warrant*) de l'espace virtuel commun que nous construisons […] au cœur des enjeux de l'organisation des connaissances à l'âge des réseaux. (Favier 2017, 13) ».

Le troisième est l'ouverture. En 2019, pour le 12ème colloque ISKO France, le thème des données et des mégadonnées ouvertes en SHS posaient de nouveaux enjeux pour l'état et l'organisation des connaissances. Lors de ce colloque, Clavier (2019, 1) a notamment montré « comment les visualisations ont progressivement pris le pas sur les modèles de représentation des connaissances classiques de nature logocentrée » c'est-à-dire dont l'analyse est centrée sur les énoncés et les seules matières langagières, et ce que cela impliquait en ce qui concerne les normes, les standards et les modes de représentation des connaissances (Cf. 1.2).

Ces enjeux semblent d'après nous pouvoir catégoriser les principales orientations guidant les travaux dans le champ de l'organisation des connaissances. Nous présentons dans la prochaine sous-partie les différentes approches de l'OC puis développons celles qui intéressent nos travaux en prenant soin de noter les enjeux plus spécifiques à chacune des approches.

1.1.2 Approches en organisation des connaissances

En 2008, Hjørland (2008, 87) dénonce l'absence d'une théorie claire pour le champ de l'OC qui s'établit sur de la pratique et présente six grandes approches qu'il reprend de (Broughton et *al.* 2005) et qu'il développe : l'approche traditionnelle, d'analyse par facette, de recherche d'information, une approche

[20] « La connaissance peut être conçue collectivement : c'est là tout l'enjeu de l'organisation des connaissances. » (Beau 2012, 9)

orientée utilisateurs et point de vue cognitif, une bibliométrique, une d'analyse de domaine. Il ne développe pas l'approche sémiotique, l'approche critique herméneutique, l'analyse de discours, l'approche étudiant le genre, la représentation de document, la typologie et description de document, les langages *markup* et l'architecture de document. Nous ne développerons ci-dessous que les approches qui influencent nos travaux, c'est-à-dire l'analyse par domaine, par facette, de discours, typologique et descriptive de document et d'architecture documentaire.

L'analyse de domaine (*domain analytic approach*) nécessite de définir ce que l'on entend par domaine. D'après Hjørland, un domaine est une « spécialisation dans la division du travail cognitif, qui est théoriquement cohérente ou socialement institutionnalisée » (Hjørland 2017)[21]. Nous verrons en quoi pour notre part le patrimoine minier peut être considéré comme un domaine de connaissance (Cf. partie 2). Dans ce type d'approche, la subjectivité et l'objectivité des points de vue est une question particulièrement soulevée, question qui est peu présente en approche traditionnelle et qui est ici analysée d'une manière systématique. Dans l'article, Hjørland insiste sur le caractère dynamique des domaines, toujours en développement et surtout qui dépendent de la théorie. D'après nous, un domaine peut être appelé domaine seulement à partir du moment où il a été vécu puis étudié par un certain nombre de personnes. Celles-ci peuvent alors être appelées des acteurs. Les SOC se réalisent grâce à l'étude systématique du domaine, par un raisonnement inductif ou déductif. Les travaux dans ce sens montrent l'importance d'une réalisation collective et la nécessité d'impliquer les acteurs dans la réalisation, afin de ne pas représenter le point de vue d'une seule personne. Le domaine est en effet la somme des points de vue de ses acteurs et sa représentation arrive à saturation lorsque toutes les expériences ont été considérées comme pouvant être conceptualisées. Tandis que l'expérience est unique (un individu vit sa propre réalité), l'étude, elle, tend à être partagée, c'est-à-dire que l'on souhaite donner à d'autres des outils standardisés pour permettre à ces personnes de connaître le même domaine. Il y a alors un fort but de transmission.

L'analyse par facette (*facet analysis*) est traitée dans (Vickery 1963 ; Maniez 1999 ; Broughton 2006a ; La Barre 2010 ; Hudon et Hadi 2017 ; Hudon 2020). Dans ce cadre, des questions sur la représentation sémantique du contenu des documents grâce au web sémantique sont posées. Les facettes peuvent en effet être analysées telles que des catégories de relation (partie, processus, agent…) comme les propriétés des ontologies. Nous développons d'avantage cette approche dans le cadre des classifications à facettes (2.2.2.1). Même si cette approche aurait pu être pertinente pour nos travaux, elle n'a pas fait l'objet d'une étude plus poussée.

[21] Traduit par l'auteure à partir de : *« A domain, on the other hand, is a specialization in the division of cognitive labor that is theoretically coherent or socially institutionalized. »* (Hjørland 2017, 439)

L'analyse de discours (*discourse-analytic approaches*) permet de faire émerger des connaissances (Clavier 2014). L'auteure montre bien la différence entre les points d'entrées d'analyse d'un corpus en information ou en communication. En sciences de l'information :

> « l'analyse porte sur les textes, le langage et la recherche d'une signification stable, représentable et formalisable [a]lors que pour les sciences de la communication, plus intéressées par les aspects énonciatifs et pragmatiques du discours, la focale porte sur les dimensions subjectives, sociales et historiques » (Clavier 2014, 141).

Cette approche nous sert dans l'étude de documents oraux retranscrits et s'articule avec l'approche « *document architectures* » pour retirer le sens d'unités linguistiques (donc du point de vue des sciences de l'information).

L'analyse typologique et descriptive de document (*document typology and description*) s'intéresse au document en tant qu'ensemble de support d'information et constituant une unité autonome (Boulogne et Institut national des techniques de la documentation 2004, 80). Elle tente de définir des types spécifiques de document en étudiant leur caractéristique discriminante (étude de la nature, du genre, du format, de la/des thématiques, de la langue etc.) et s'attache à leur description (création de résumé, indexation, notes) pour permettre leur lecture, leur classement, leur indexation et accès, leur organisation dans une bibliothèque ou dans un musée ou leur conservation. Un exemple de système réalisé avec cette approche est la « Liste d'autorité Typologie documentaire pour l'indexation des archives locales »[22] qui comme son nom l'indique fournit une liste de termes pour indexer des archives locales et qui distingue par exemple le *livret ouvrier* du *livret militaire*. L'hétérogénéité des documents pose la question de l'interopérabilité entre les différents systèmes, lors de l'interrogation de leur contenu et de leurs métadonnées, par exemple. La description des documents numériques peut se réaliser grâce à l'extraction (automatique, semi-automatique ou manuelle) des entités nommées (personne, lieux, dates etc.) ou des concepts. La question que nous nous posons dans ce cadre est : est-ce que, et si oui, en quoi le type de document influe sur l'analyse sémantique de celui-ci? Est-ce que la formulation d'un résumé descriptif normalisé de chaque document permet l'interopérabilité de documents hétérogènes ? Cette approche pose également la question de la granularité dans l'établissement de la typologie et du choix des éléments de description.

L'analyse de l'architecture documentaire (*document architectures*) permet enfin d'interroger la structuration d'un document et admet l'identification de plusieurs niveaux sémantiques. Par exemple, pour un texte issu d'un oral transcrit ou d'un écrit, les niveaux lexique, locutions, phrases, tours de parole, paragraphe, document peuvent être étudiés sous le prisme du sens. Les travaux de Diemert (2012) montrent par exemple comment décrire sémantiquement des documents audiovisuels structurés. Le séquençage thématique des vidéos est récent, on ne s'intéresse plus qu'à la description globale d'une vidéo mais on souhaite pouvoir accéder, comme pour les textes, à des parties. Dans ce sens, le

[22] http://data.culture.fr/thesaurus/page/ark:/67717/T3

séquençage de vidéos avec thématisation est récemment[23] apparu sur la plateforme YouTube, afin que l'utilisateur puisse naviguer plus facilement au sein d'une vidéo. Nous étudierons l'impact du choix de l'unité de séquençage d'un entretien vidéo sur sa représentation sémantique (granularité) (Cf. chapitre 8). D'autre part, définir l'architecture des documents peut permettre d'effectuer des requêtes sur des parties de ce document (la possibilité de faire des requêtes sur le titre, parfois sur le résumé est déjà en vigueur sur un certain nombre de bases de données bibliographiques) mais nous pouvons imaginer requêter l'introduction, la conclusion, ou bien, si le document est divisé thématiquement, faire une requête d'un terme à partir d'une thématique. Il faudrait néanmoins tout d'abord définir des types de documents (les documents scientifiques ont normalement une introduction, un résumé, des mots-clés, une conclusion) mais les autres ne sont pas toujours aussi systématiques. De ce fait, nous considérons les approches d'architecture et typologique du document, complémentaires.

Dans cette section, nous avons montré en quoi ces différentes approches en OC peuvent être questionnées dans le cadre de notre recherche et peuvent aider dans la description et l'indexation de documents hétérogènes à plusieurs niveaux.

1.1.1 Systèmes d'organisation des connaissances (KOS)

Nous venons de le voir, d'après (Hjørland 2008) les SOC sont utilisés pour organiser les documents, les représentations de documents, les œuvres et les concepts. Nous traitons des SOC organisant les documents et représentant des documents en (chapitre 2), avec l'approche documentaire de l'OC, et l'organisation des concepts et des œuvres en (chapitre 3) avec une approche ontologique. Zacklad définit les SOC comme relevant d'une forme d'écriture codifiée :

> « […] les SOC sont des ensembles de termes codifiés constituant un lexique disposant ou non de règles d'association explicites et permettant de faciliter les opérations de gestion portant sur des contenus documentaires ou sur les caractéristiques de situations non préalablement documentées (modèle documentaire ou modèle de situation) » (Zacklad 2010, 135).

Ils aident donc à résoudre des problématiques pouvant survenir lors de la gestion documentaire ou bien à chaque fois qu'une situation nécessite d'être modélisée, c'est-à-dire simplifiée et représentée à l'aide d'un modèle. Les modèles doivent décrire au mieux une situation précise ou un contenu documentaire. La modélisation pose un certain nombre de questions traitées en 1.3. et est notamment étroitement liée au problème des ontologies.

Polity, Henneron et Palermiti définissent quant à eux les SOC comme « toutes sortes de schémas d'organisation allant des simples listes alphabétiques ou faiblement structurées (listes d'autorité, glossaires, dictionnaires, nomenclatures, etc.) à des schémas classificatoires hiérarchiques (plans de classement, classifications générales ou spécialisées, taxinomies, listes de vedettes matières, etc.) ou

[23] Nous avons commencé à remarquer ceci courant 2021.

encore à des organisations privilégiant des relations non exclusivement hiérarchiques (thésaurus, réseaux sémantiques, ontologies, etc.) « (Polity, Henneron, et Palermiti 2005, 13). Nous verrons que les SOC sont en effet hétérogènes et nécessitent des critères pertinents permettant de les distinguer (2.1).

Les éléments constitutifs des SOC sont de plusieurs natures. En 2011, Vandenbussche propose une liste des différents éléments constituant des SOC dans le domaine de l'Ingénierie des Connaissances (Vandenbussche 2011). Cette liste contient les concepts, les relations, les attributs, les contraintes et les instances. Nous définissons les concepts et les instance en (3.5.1) et les relations en (3.5.2) dans le cadre des ontologies. Les concepts des SOC possèdent des attributs qui les caractérisent comme l'URI (identifiant unique d'un concept), le code (place d'un concept dans la hiérarchie) ou le libellé (permettant de nommer le concept), et les contraintes s'appliquent sur les SOC décrits en langage formel ou logico-mathématique pour restreindre l'utilisation des concepts. Ces contraintes doivent être explicitement définies. Par exemple, le concept représenté par la classe « ville » ne permettra pas de choisir parmi les instances des « personnes ».

Zacklad définit donc des typologies de SOC (Zacklad 2007, 2010, 2018). Si en 2007, l'auteur comparait les classifications, les thésaurus, les ontologies et les folksonomies du point de vue de la recherche ouverte d'information, l'article de 2010 ajoute à la catégorie des SOC les index de moteurs de recherche que l'auteur considèrent être des SOC automatiques et oppose donc la forme d'écriture codifiée manuelle de celle automatique. L'auteur mentionne également en 2010 la catégorie des langages documentaires, les schémas de classification et les langages de représentation des connaissances issus de l'intelligence artificielle. En 2018, il présente des critères d'analyse permettant d'identifier quatre grandes familles de SOC actuelles (tableau 1, nous mettons en gras les SOC que nous rencontrons et réalisons dans l'étude) et en 2010, il présente cinq critères d'évaluation qu'il reprend en 2018 (tableau 2).

Tableau 1 : Typologie des SOC selon (Zacklad 2018)

Type de SOC	**Terme spécifique**	**Exemple de SOC**
SOC symboliques	SOC épistémiques spécialisés	Taxinomies scientifiques (Cf. 2.2.2.2)
	SOC épistémiques généralistes	**Systèmes de classification** (schémas de classification (Zacklad 2010)) CDD, CDU, classification à facettes (Cf. 2.2.2.1)
		Systèmes d'indexation, **termes contrôlés** (Cf. 2.2.1), **thésaurus de spécialité**, **terminologies** (Cf. 2.2.2.3)
	SOC cognitifs et logiques	Réseaux sémantiques, **cartes mentales**, **ontologie formelles**, ontologies sémiotiques, prototype (Cf. 1.2.2)
	SOC vernaculaires	**Systèmes de classement et d'indexation vernaculaires** professionnels et personnels, dans des carnets, des sites web, des panneaux, etc. : étiquettes papiers pour le rangement d'objets ou de ressources multimédia, **titres** et **sous-titres de section**, **mots-clefs dans les bandeaux des sites**, tags, etc. (Cf. 5.1.3.1 et 5.1.3.2)
SOC algorithmiques	SOC statistiques (SOC automatiques, (Zacklad 2010)	Index constitués automatiquement par les moteurs de recherche traitant des corpus de textes ou des informations textuelles du web, qui sont de loin les plus utilisés aujourd'hui
	SOC connexionnistes	Structures de données, vecteurs de données représentant des catégories issues des méthodes d'apprentissage de l'Intelligence Artificielle

Tableau 1 : Typologie des SOC selon (Zacklad 2018)		
Type de SOC	**Terme spécifique**	**Exemple de SOC**
SOC visuels		Typologie exhaustive à produire
SOC incarnés		Objets tangibles, espace, voire temporalité - Arrangement des ingrédients sur le plan d'une cuisine - Emplacement des outils sur un plateau chirurgical Typologie exhaustive à produire

Comme nous pouvons le voir dans le tableau 1, une typologie des SOC visuels et des SOC incarnés reste encore à produire. Ce ne sera pas là l'objectif de notre étude étant donné que nous nous concentrons sur les SOC symboliques à visée documentaire, que nous définissons en 2.1.

Ci-dessous, nous reprenons les cinq critères d'évaluation des SOC définis par (Zacklad 2010), que l'auteur reprend dans (Zacklad 2018). Ce dernier mentionne qu'il « resterait à caractériser [les types de SOC définis en 2018 (tableau 1)] selon les dimensions qu'[il avait] identifiées [en 2010] » (Zacklad 2018, 100). Quant aux critères d'évaluation des SOC, ils sont formulés à partir des questions suivantes (tableau 2) :

Tableau 2 : critères d'évaluation des SOC dans (Zacklad 2010) et (Zacklad 2018)		
1	« Qui parle, qui participe, au nom de quoi ? »	Instance énonciative responsable du SOC
2	« De quels objets ? »	Source de la description : - Source primaire, les termes des SOC font référence aux caractéristiques d'une situation extradocumentaire : modèle de situation - Source secondaire, le SOC décrit le contenu d'un document : modèle de contenu
3	« Dans quel format d'expression ? »	Type et degré de formalisation : - Formalité sémiotique (linguistique) - Logico-mathématique
4	« Quel est le niveau de signification des termes du SOC par rapport aux « objets » qu'ils décrivent ? »	Niveau de signification des termes du SOC par rapport aux éléments qu'ils décrivent (type de sémantique) : - Référentiel - Rhétorico-herméneutique
5	« En utilisant quel principe d'association entre les termes pour permettre quel type de raisonnement ? »	Principes d'association entre les termes du SOC : - Typée et référentielle (requêtes de données) - Heuristiques (navigation) - Statistiques (moteurs de recherche)

Les SOC, destinés à organiser la connaissance peuvent à leur tour être organisés selon des critères que les chercheurs identifient comme pertinents (et les définissants). La formalisation de ces critères permet leur construction et leur réutilisabilité. Les résultats doivent cependant être communicables, la représentation joue dans ce cadre un rôle primordial.

La section suivante retrace les questionnements en représentation des connaissances et montre dans ce cadre les enjeux de la modélisation.

1.2 Représentation des connaissances (RC)

La question de la représentation des connaissances n'est pas récente, elle est naturellement posée dès l'apparition des premières bibliothèques, dans le but de mettre en visibilité (au sens de perceptible par

les sens, en prenant le sens de l'information comme trace sensible[24]) le savoir pour pouvoir le réutiliser à plusieurs fins différentes telles que l'éducation, la production de nouvelles connaissances, l'innovation, la recherche qui sont caractéristiques de la nature humaine et de son évolution.

Aujourd'hui, les points clés de la représentation des connaissances seraient l'inscription des connaissances sur un support (matériel), dans un formalisme perceptible et intelligible (par l'homme et la machine), organisé et structuré (faisant sens et permettant l'accès). Dans nos travaux, nous nous intéressons plus particulièrement à la représentation sémantique des connaissances pour une compréhension homme/machine. En 2000, Bachimont rappelle que « le problème posé par la représentation des connaissances est de formaliser dans un langage formel de représentation les connaissances permettant de traiter le problème à résoudre » (Bachimont 2000, 2).

De son côté, Clavier mentionne qu'il faut redéfinir les modes de représentation des connaissances (Clavier 2019). En effet, elle rappelle que la datavisualisation et ses outils ne sont plus destinés qu'aux chercheurs spécialisés en traitement de données textuelles et traitement automatique des langues, ils sont en effet ouverts à tous. Ceci implique de redéfinir également les normes et les standards de représentation des connaissances.

Si les ontologies sont au centre des préoccupations actuelles en représentation des connaissances depuis ces quinze dernières années (1.2.2.3), nous revenons tout d'abord sur les caractéristiques des représentations mentales (1.2.1) puis des représentations graphiques[25] (1.2.2) et étudions ces dernières dans différents domaines telles que la terminologie et l'information-documentation. Nous traitons ensuite des représentations ontologiques et des systèmes de représentation des connaissances. Puis, nous introduisons l'enjeu de représentation globale des connaissances et dans ce cadre, le rôle du web sémantique (1.2.3), et les langages de représentation des connaissances en (1.2.3). Nous finissons par questionner la représentation de l'aspect sensible des connaissances (1.2.4).

1.2.1 Représentation mentale des connaissances

Les concepts, les notions, le sens (d'un signe, d'un terme), les descripteurs des thésaurus ont en commun le fait qu'ils sont tous considérés comme des représentations mentales, c'est-à-dire comme une représentation que l'on se fait, par la pensée, d'une projection sensorielle, d'un concept ou d'une situation. Ainsi, la représentation mentale des connaissances correspond à la projection d'un concept, la représentation mentale des savoirs correspond à la projection sensorielle et la représentation mentale du sens et des notions correspond à la projection d'une situation.

Dans nos travaux, nous définissons le concept comme une « représentation mentale et abstraite » (Bui 2003, 249) qui « retient les caractéristiques communes à un ensemble d'objets » (L'Homme 2004, 25) où les objets sont des instances partageant des traits communs identifiables (Kassel 1999) ». Dans cette

[24] D'après Beau qui travaille sur les systèmes d'information, « une *information* est une trace sensible (olfactive, auditive, visuelle, tactile ou gustative) de la réalité observable (le fait) "en forme" indépendante de sa manifestation concrète » (Beau 2015, 16).
[25] La représentation graphique n'est pas la seule opposée à la représentation mentale, il existe par exemple la représentation algorithmique dans laquelle se placent par exemple les systèmes experts, outils capables de reproduire les décisions cognitives d'un expert pour un domaine donné.

optique, le concept est considéré comme l'« unité de base de la représentation des connaissances » permettant de « catégoriser les objets » (Bui 2003, 249)[26]. L'application de cette définition permet « une description formelle [d'un] domaine où le concept est développé comme une classe et où les propriétés de chaque concept décrivent les différents attributs de la classe et les contraintes des facettes » (Bui 2003, 246). Un concept ne peut pas être en soi formalisé donc sa représentation (graphique ou en tout cas pas mentale) est nécessaire à des fins de manipulation.

D'après Fondin, le sens est une représentation mentale « cohérente que tout homme construit ou dégage de quelque chose observée dans son environnement » (Fondin 2005, 15)[27]. L'auteur écrit que si ce sens (donc cette représentation mentale) est partagé(e), l'homme l'appelle *information*. Ceci nous amène à dire qu'une information est initialement une représentation mentale. La question que pose Fondin et qui nous intéresse concerne le sens : « est-il immanent car déposé dans le document par son auteur, ou construit car élaboré contextuellement par les acteurs, ici par celui qui lit ? » (Fondin 2005, 15). Ceci montre toute la difficulté de représenter le sens à partir de documents. D'après Bachimont, Gandon, Poupeau et *al.*, plusieurs conceptions de la sémantique et du sens sont possible lorsque l'on souhaite « représenter des connaissances, interpréter des contenus » en explicitant « le sens d'un contenu », puis en rendant « exploitable cette explication par un traitement informatique » (Bachimont et *al.* 2011, 25). Pour nous, tout comme le concept, le sens ne peut pas être formalisé, seule sa représentation graphique peut l'être.

Le sens d'un terme est la représentation mentale qu'on en a. Cette approche est psychologique, issue des sciences cognitives « où le sens serait "dans la tête", comprendre le cerveau permettant alors de saisir ce qu'il pense » (Bachimont et *al.* 2011, 25). Rastier déclare la même chose du sens d'un signe (Rastier 1997). Zacklad mentionne en revanche que ceci est incompatible avec « les épistémologies de la philosophie pragmatique (J. Dewey) ou de la tradition herméneutique (cette dernière étant largement répandue dans les sciences humaines et sociales) » (Zacklad 2010, 140). Nous retrouvons les approches sémantiques opposées aux approches sémiologiques (le web socio-sémantique), le pragmatisme s'opposant aux conceptions cartésiennes et rationalistes sans pour autant renoncer à la logique.

D'après (AFNOR 1981) cité par Vandenbussche : « [u]n concept de thésaurus peut être défini comme « une représentation mentale stable d'un aspect de la réalité qui peut être pensé en l'absence de cette réalité » (Vandenbussche 2011, 32). D'après Hudon, un descripteur de thésaurus représente un concept unique et n'est signifiant que dans les rapports qu'il entretient avec d'autres descripteurs (Hudon, 1994, 37).

Quant à la notion, elle est définie en terminologie comme une « représentation mentale d'un objet constitué à partir d'une combinaison unique de caractères (appelés aussi caractéristiques). Les caractères

[26] Bui définit le concept en Intelligence Artificielle, en sémantique linguistique et en linguistique (Bui 2003, 249).
[27] « L'information est-elle un objet réel, objectivable, ou un objet social ? À travers cette interrogation est posée toute la question du « sens » et de son attribution. Le sens, cette représentation mentale cohérente que tout homme construit ou dégage de quelque chose observée dans son environnement, ce qu'il appelle information si ce sens est partagé, est-il immanent car déposé dans le document par son auteur, ou construit car élaboré contextuellement par les acteurs, ici par celui qui lit ? » (Fondin 2005, 15)

jouent un rôle essentiel dans le travail terminologique ; ils servent à définir et à délimiter une notion […] et déterminent la place de la notion dans un système de notions […]. Ils peuvent être regroupés en différentes catégories (types de caractères) » (Groupe de travail « Terminologie et documentation » 2014, 18). Contrairement au concept qui dit ce que la chose est en soi, la notion dit ce que l'on sait de la chose. La notion implique une étude, un jugement (Termium Plus 1971).

La volonté de définir la représentation mentale est à intégrer dans des perspectives de représentations mentales globales influant sur la recherche d'information sur le web, ou sur l'architecture de l'information des systèmes. Dans ce cadre, la différence entre utilisateurs (ou concepteurs) novices et experts est mise en valeur. Dans le domaine de la conception de sites web, Chevalier montre la difficulté d'utilisation des sites selon que l'utilisateur ou le concepteur soit novice ou expérimenté (Chevalier 2008). L'auteur mentionne notamment la mauvaise structuration du site web influant sur la difficulté des novices à construire une représentation mentale globale et cohérente du site à la différence des utilisateurs expérimentés et des concepteurs (Chevalier 2008). Ainsi, une mauvaise structuration de l'information mène à une représentation mentale incomplète et à des difficultés d'utilisation de l'objet réalisé. Dans l'étude de Facetag, qui est un prototype fonctionnel d'un outil de marquage sémantique collaboratif conçu « pour améliorer l'architecture de l'information d'un système de *tagging* social de *bookmarks* », Doria (2013) cite Quintarelli, Resmini, et Rosati sur le souci qu'à « l'ordre généralement alphabétique de présentation des tags [et qui] réduit les possibilités de navigation rapide, et rend difficile la construction d'une représentation mentale globale et cohérente des tags présents » (Doria 2013, 95). Ici, l'auteure montre la faiblesse d'une structure dont l'ordre ne fonctionne pas comme l'organisation mentale humaine des connaissances. Enfin, dans le domaine de l'œnologie et des connaissances sensibles, donc ce que nous considérons comme du savoir, l'objectif de Moutat « est de comprendre comment un lecteur non spécialiste de la dégustation parvient à construire une représentation mentale globale des propriétés olfacto-gustatives d'un vin, à partir de ses seules facultés interprétatives » (Moutat 2019, 2). Elle montre que le caractère imagé des termes décrivant les vins « sont motivés et peuvent s'avérer déterminants dans le travail interprétatif du néophyte » et qu'une transversalité du sens est possible entre le domaine olfactif et gustatif parce que « sur le plan sensible, des phénomènes relevant de domaines sensoriels différents peuvent partager des traits fondamentaux et manifester une communauté d'organisation […] entre la perception (qualités organoleptiques du vin) et le langage (propriétés sémantiques de la terminologie œnologique) » (Moutat 2019, 17). En d'autres mots, au niveau du savoir, les facultés interprétatives permettent une représentation mentale globale de propriétés sensitives.

1.2.2 Représentation graphique des connaissances

La représentation graphique des connaissances est associée à la notion de visualisation de l'information, domaine plus récent que la recherche d'information. Concernant la visualisation, un panorama des outils et de méthodes est donné dans (Vidal 2006). Elle oppose la visualisation linéaire à la visualisation

hiérarchique, à la visualisation des réseaux, à la visualisation multidimensionnelle, à la visualisation vectorielle. Nous ne traiterons pas de celles-ci.

La représentation graphique des connaissances implique d'identifier et définir :

- les connaissances à représenter (leur type),
- la position de celles-ci par rapport aux autres,
- les signes (symboles, formes, codes, conventions) qui permettront de représenter les connaissances, leur position et leur sens.

Les représentations graphiques sont utilisées afin de : faciliter la recherche d'information, visualiser ou illustrer des résultats, représenter le contenu d'un ou de plusieurs documents, représenter un domaine en partie et en détail ou dans sa globalité, orienter l'utilisateur… En somme, les représentations graphiques aident à guider les représentations mentales de potentiels utilisateurs de la connaissance.

Les domaines de la terminologie, de l'information et des documents nécessitent des représentations graphiques qui leur sont propres. Ainsi, il sera parfois nécessaire de représenter des notions, ou bien des faits (connaissances factuelles) ou des concepts (connaissances ontologiques). L'enjeu dans chacun de ces domaines, nous le répétons, est bien de représenter quelque chose qui fasse sens (le même sens) aux yeux du concepteur et de l'utilisateur.

1.2.2.1 Domaine de la terminologie

En terminologie, on représente des notions et les relations entre elles grâce aux schémas notionnels (Groupe de travail « Terminologie et documentation » 2014). Trois types de ces schémas existent : le champ notionnel, l'arbre de domaine, ou encore le système de notions (Cabré i Castellví et *al.* 1998). Les schémas notionnels se présentent sous la forme d'arborescence « parfois complexe et difficilement gérable avec les outils traditionnels ». (Groupe de travail « Terminologie et documentation » 2014, 101). Les outils de *mind mapping* permettent de créer des cartes heuristiques afin de représenter et d'organiser les liens qui relient les notions entre elles. Ces outils sont également utilisés dans la construction d'ontologies. (Groupe de travail « Terminologie et documentation » 2014, 101).

- Les liens entre les notions sont thématiques : on construit un champ notionnel pour les regrouper entre elles.
- Les relations entre les notions clés d'un domaine sont catégoriques (classes d'objets) : on construit un arbre de domaine, ou arbre analogique. La forme est pyramidale (arborescente). « Ce type de représentation n'est pas propre à la terminologie. D'autres disciplines en font usage. Ces arbres ont pour noms : organigramme, structure arborescente, arbre généalogique et thésaurus. » (Zafio 1985, 161).

- Les notions sont reliées hiérarchiquement selon une structure rigoureuse et logique : on construit un système notionnel.

En 1995, Rastier montre pourquoi la terminologie comme ensemble de termes ne peut pas représenter un champ de connaissance. La raison qu'il avance est que les textes de spécialité, qui ne sont pas faits de mots, « représentent bien plus de sens que toute représentation établie sur des entités » (Rastier 1996). L'auteur montre ainsi la limite des réseaux arborescents (tel l'arbre de domaine) qui « se heurtent au même type d'objection que la syllogistique classique, dont ils ont hérité : leur parcours n'apprend rien que l'on ne sache déjà » (Rastier 1996, s.p.). Ainsi, la terminologie n'est pas suffisante pour représenter un champ de connaissance. De plus, l'auteur rappelle que le sens porté par les mots est « toujours le produit d'une interprétation ». Le danger est d'isoler les significations de leur contexte historique de création et de leur interprétation. Une solution qu'il propose serait de « décrire les parcours interprétatifs et les contraintes linguistiques sur le parcours »[28], donc prendre en compte la variation diachronique dans les représentations. L'apport d'autres disciplines (notamment la linguistique et plus particulièrement l'étymologie) serait bénéfiques à la représentation des connaissances. D'autre part, Condamines montre comment et pourquoi la collaboration entre la discipline de la terminologie et celle de l'intelligence artificielle est possible. En effet, « la terminologie et l'intelligence artificielle ont en commun de travailler sur la représentation des connaissances, la première via l'étude des concepts désignés par les termes, la seconde via l'étude de la formalisation de la connaissance » (Condamines 1994, 36). Enfin, comme l'a montré Maryvonne Holzem, l'intelligence artificielle a fait de la documentation un outil terminologique, « indissociable de l'activité terminologique » permettant d'élaborer un outil commun : le thésaurus terminologique (Holzem 1999, 165). La « terminographie » correspond au travail d'indexation. Dans ce sens, la terminologie permet de représenter les connaissances en les indexant. Au niveau terminologique, trois éléments sont indissociables et reliés entre eux par des liens étroits : le concept, la définition et le domaine. Comme le mentionnent Béjoint et Thoiron « le domaine indique l'appartenance du concept (et du terme qui le désigne) à un système conceptuel, la définition servant à différencier les concepts à l'intérieur de ce système » (Béjoint et Thoiron 2000, 183). Ceci signifie que toute représentation des connaissances par des termes doit inclure un concept, une définition et un domaine.

1.2.2.2 Domaine de l'information-documentation

Dans ce domaine, ce ne sont pas les connaissances qui sont représentées mais les informations (pour créer de la connaissance). Dalbin mentionne la représentation graphique de l'information, utilisée dans les années soixante sur papier pour les schémas fléchés des thésaurus, puis sur ordinateur dans les activités de veille et montre que celles-ci ont pris un nouvel essor sur le Web pour la recherche et la

[28] À propos du sens porté par les mots : « il est toujours le produit d'une interprétation, cette interprétation fût-elle fondée sur des inférences par défaut, qui permettent d'hériter des traits sémantiques d'un type lexical préconstruit. Si le sens résulte d'une interprétation, pour en rendre compte, il convient de décrire les parcours interprétatifs et les contraintes linguistiques sur les parcours. En diachronie, le sens d'un mot est constitué par l'histoire de ses interprétations. Prendre pour point de départ une signification conduirait à oblitérer cette histoire » (Rastier 1996).

présentation des résultats, avec par exemple le moteur Kartoo dès 2001 (Dalbin 2007, 49-50). Celui-ci était un métamoteur de recherche à interface cartographique et interactive. Il n'existe plus depuis 2010.

À propos des langages documentaires, et plus particulièrement des thésaurus Couzinet mentionne différentes formes que peuvent revêtir leur représentation graphique (Couzinet 2011, 180) :

- un arbre hiérarchique sur lequel seule la relation hiérarchique est représentée,
- un schéma fléché ou « *arrowgraph* » tel un quadrillage avec des coordonnées facilitant le repérage des descripteurs. Les termes spécifiques sont les plus éloignés du nom du domaine, placé au centre du quadrillage. Le lien de hiérarchie est représenté par une flèche, « d'autres liens hors du cadre représentés par des traits désignent des relations d'association » (Couzinet 2011, 180).
- une forme circulaire avec une répartition du vocabulaire dans des zones concentriques.
- par tableau graphique (« *graphic display* ») ou terminogrammes, avec des cadres réunissant chaque concept par champ sémantique. Différentes flèches lient les termes génériques, associés, employés pour (EP), et termes spécifiques au descripteur.

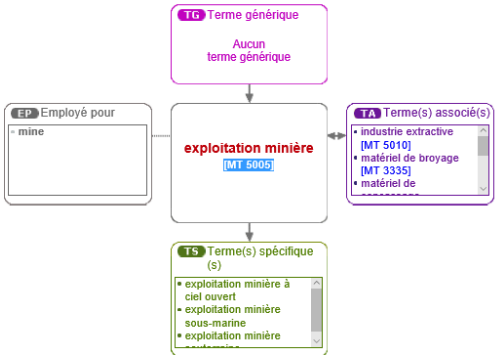


Figure 4 : Exemple de terminogrammes dans Motbis pour le descripteur *exploitation minière*

Un outil de visualisation Skosplay[29] est un exemple d'application gratuite permettant de rendre compte et de visualiser des thésaurus, des taxonomies ou des vocabulaires contrôlés exprimés à l'aide de l'ontologie *Simple Knowledge Organization System* (SKOS, Cf. 1.3.2). L'utilisateur peut ainsi choisir entre différentes représentations dont une visualisation arborescente (en arbre), une visualisation carrée (*icicle*) ou visualisation circulaire (*sunburst*, issu de l'approche surfacique). L'utilisateur pourra aussi choisir le type de rendu sous la forme : soit d'un tableau de correspondance des langues, soit sous la forme d'index (alphabétique, alphabétique avec attributs, arbre hiérarchique statique, arbre hiérarchique dépliable, permuté, *KeyWord In Context* (KWIC) à partir des mots contenus dans les libellés de la langue choisie). Ces types de visualisations diffèrent par rapport à l'agencement des informations et à leur ordre.

Aussenac-Gilles, Baziz, et Hernandez dont les travaux se situent en recherche d'information (RI), ont par exemple montré l'efficacité d'une approche combinant une représentation à base de concepts (noyau

[29] https://skos-play.sparna.fr/play/

sémantique) et l'utilisation de la pondération pour expliciter la représentativité des concepts qu'ils estiment « importante pour classer et comparer des documents répondant à une requête. » (Aussenac-Gilles, Baziz, et Hernandez 2006, 15). Ils s'appuient notamment sur le réseau de *synsets* de *Wordnet*. Les synsets sont des éléments qui décrivent une signification commune et partagée dans la même langue, donc des ensemble de synonymes.

Nous présentons ci-dessous cinq types de représentations graphiques utilisées en OC : les cartes heuristiques, les cartes conceptuelles, les réseaux sémantiques, les graphes conceptuels, les représentations ontologiques.

Les cartes heuristiques (mentales, *mind map*) ont été formalisées par le psychologue Tony Buzan, qui a écrit de nombreux ouvrages sur l'apprentissage, la mémoire et le cerveau. Des formes graphiques de ce type apparaissent néanmoins bien plus tôt, déjà au III^{ème} siècle dans l'introduction aux catégories d'Aristote, le philosophe Porphyre représente hiérarchiquement le savoir à partir de celles-ci. La représentation réalisée sera appelée « arbre de Porphyre ». On peut considérer celui-ci comme l'ancêtre des ontologies.

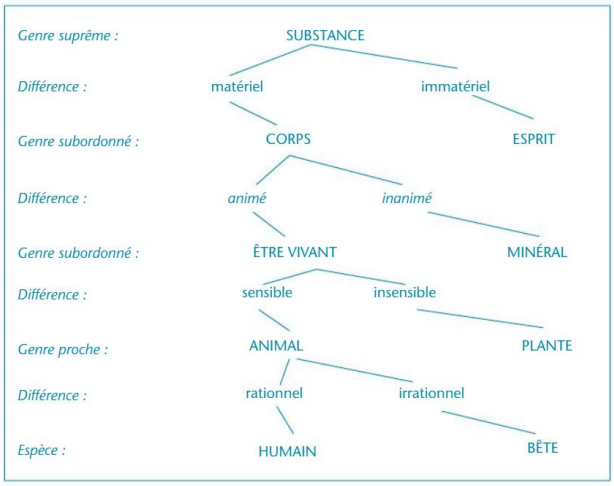


Figure 5: Représentation de l'arbre de Porphyre issue de (Delengaigne et Garcia 2018)

Buzan et Buzan ont écrit l'ouvrage *The Mind Map Book* dans lequel ils montrent les bénéfices de l'apprentissage avec les *mind maps*. À ce propos, ils notent qu'à la différence du texte linéaire qui montre juste les faits, les *mind maps* montrent les relations entre ces faits, donnant aux étudiants une compréhension plus profonde du sujet (Buzan et Buzan 2006, 233).

Les cartes mentales sont très utilisées dans les approches pédagogiques, d'apprentissage, et coopératives de création de contenus. De nombreux outils de *mind-mapping* existent actuellement gratuitement sur le web (Mindomo, FreeMind, Mindmeister) et proposent des fonctionnalités des plus simples aux plus complexes et variant en fonction des formats d'exportation des cartes, de leur disponibilité sur différentes plateformes, des abonnements, des fonctionnalités collaboratives. Ils peuvent aussi être utilisés pour réaliser des schémas notionnels dans le cadre de l'élaboration des terminologies, ils

permettent en effet « de créer des cartes heuristiques aussi appelées cartes cognitives afin de représenter et d'organiser les liens qui relient les notions entre elles. Ces outils sont également utilisés dans la construction d'ontologies » (Groupe de travail « Terminologie et documentation » 2014, 101). Enfin, au niveau de la représentation du savoir, l'Encyclopédie Universalis représente par exemple le lien entre ses différents articles à l'aide de cartes mentales (figure 6). La représentation permet à l'utilisateur de naviguer plus facilement d'articles en articles et de se représenter plus aisément la couverture informationnelle de l'encyclopédie. Dans ce cas, elle peut aussi aider à délimiter un domaine d'étude (Cf. chapitre 4)[30].

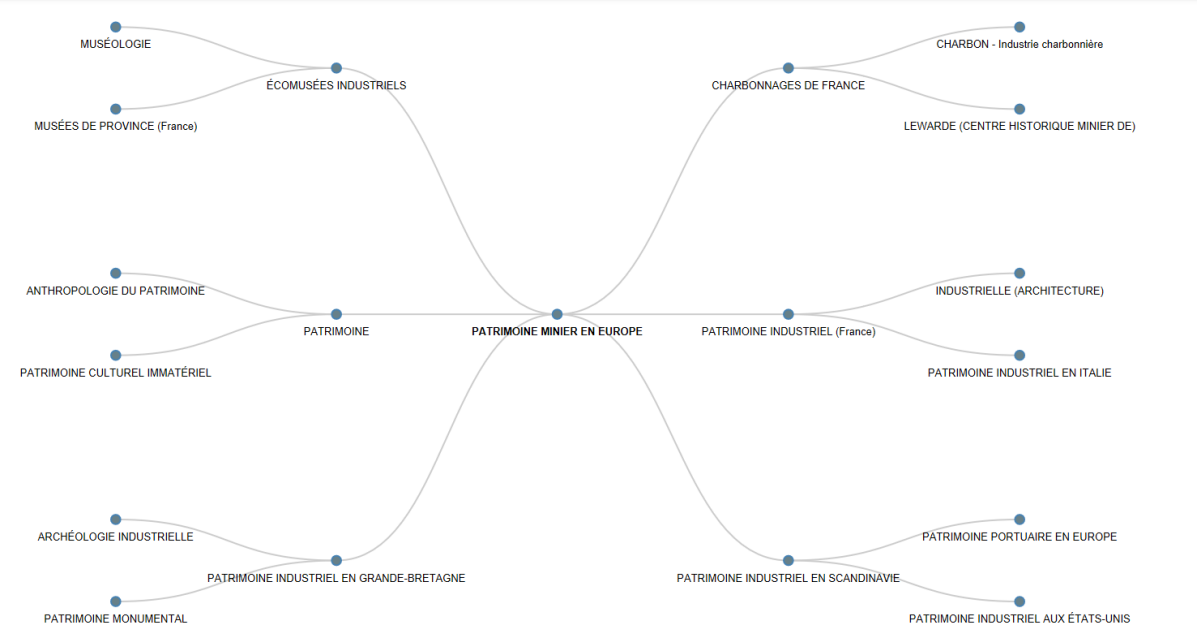


Figure 6 : Carte mentale Patrimoine minier en Europe issue de l'Encyclopédie Universalis © *Encyclopædia Universalis*

Un autre exemple est la représentation des *Linked Open Data* (figure 7) en carte heuristique. Cette représentation montre par exemple le lien entre le catalogue de bibliothèque universitaire SUDOC et la classification de DEWEY ou IdRef (base qui fournit des identifiants et référentiels pour l'enseignement supérieur et la recherche). Chaque couleur représente un type de domaine (DEWEY fait partie du domaine des publications).

[30] De ces articles, ceux de charbonnages de France, de charbon-industrie charbonnière, de Lewarde (centre historique minier de), puis de patrimoine, de patrimoine culturel immatériel, de patrimoine industriel (France), d'industrielle (architecture) nous intéressent plus particulièrement.

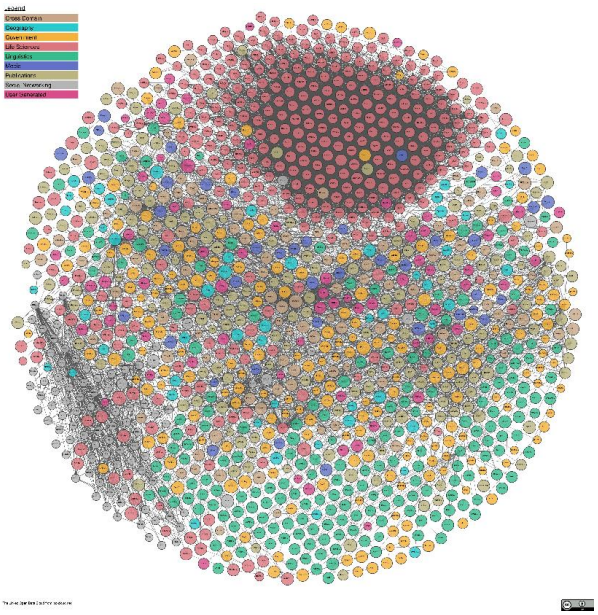


Figure 7 : Représentation heuristique des *Linked Open Data* à l'adresse https://lod-cloud.net/clouds/lod-cloud.svg

Contrairement aux cartes conceptuelles, les cartes heuristiques présentent les données de manière arborescente.

La carte conceptuelle est quant à elle un « système graphique de représentation des concepts et des relations entre les concepts » (Paquette 2002, 337). Les cartes conceptuelles sont des modèles dit de « de représentation des connaissances » comme les ontologies sémiotiques définies par Zacklad (2010). Elles sont toutes deux « construites par des ingénieurs de la connaissance, sur la base de l'analyse de documents primaires mais également à l'aide d'entretiens avec des experts ou par l'analyse *ad hoc* d'un domaine » (Zacklad 2010, 142). CMapTool est un exemple d'outil gratuit de création de cartes conceptuelles avec un cloud, ce qui permet de réaliser un certain nombre de cartes que nous utilisons dans nos travaux (figure 8).

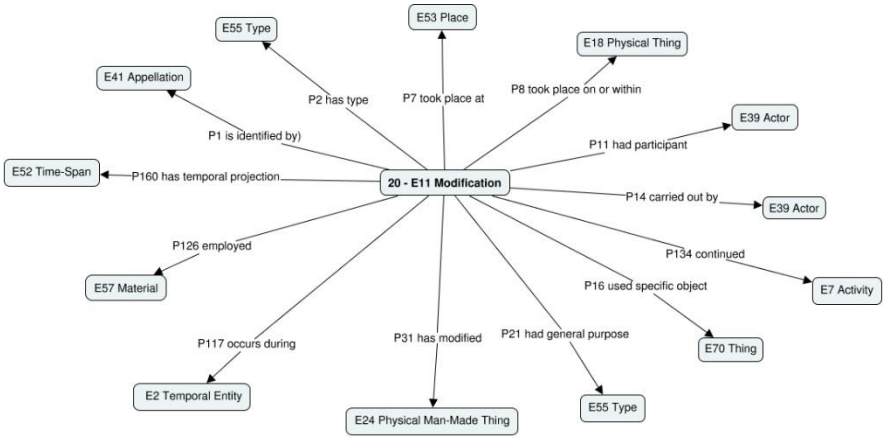


Figure 8 : Carte conceptuelle réalisée pour représenter la réduction du CIDOC CRM pour la classe *E11 Modification* (Cf. Représentation des modèles obtenus)

Les relations entre les concepts sont de nature sémantique et sont représentées par des lignes fléchées sur ou au-dessus desquelles les identifiants des relations sont placés (ci-dessus *P21 had general purpose*).

Le réseau sémantique, en linguistique ou en intelligence artificielle est un « système de représentation des connaissances prenant la forme d'un graphe composé de nœuds représentant des objets, des concepts ou des événements reliés par des arcs qui décrivent les relations entre les nœuds » (GDT). Ce type de réseau est également appelé réseau associatif. Le formalisme de ceux-ci n'est d'après Rastier « qu'une représentation graphique de la logique des prédicats du premier ordre » (Rastier 2004, 25). Dans sa thèse, Hernandez cite Quillian (1968) qui définit le réseau sémantique comme « une représentation graphique d'une conceptualisation d'une (ou plusieurs) connaissance humaine » (Hernandez 2005, 39). La représentation d'un réseau sémantique est un graphe étiqueté et orienté utilisant des arcs reliant des nœuds de départ à des nœuds d'arrivée, les arcs pouvant être multiples ou simples et leur nature permettant parfois des inférences. Ce type de graphe manquant de précision sémantique (Woods 1975 ; Brachman 1977), de nouveaux formalismes ont été définis comme les frames, les logiques de description et les graphes conceptuels (Hernandez 2005, 39). Nous présentons les graphes conceptuels ci-dessous.

La figure 9, issue des travaux de Ghio (1997) est un exemple de réseau sémantique et se lit : l'avocat (fruit) est une sorte de fruit, un fruit est une sorte d'aliment, un aliment est un objet de l'action de « manger »; de plus, une fourchette est une sorte d'ustensile; un ustensile est un instrument de l'action de « manger ». L'auteur, travaillant sur un dispositif de décodage acoustico-phonétique, montre comment ce type de réseau pourrait permettre de lever l'ambiguïté phonétique des verbes /mãʒ/ et /lãʒ/.

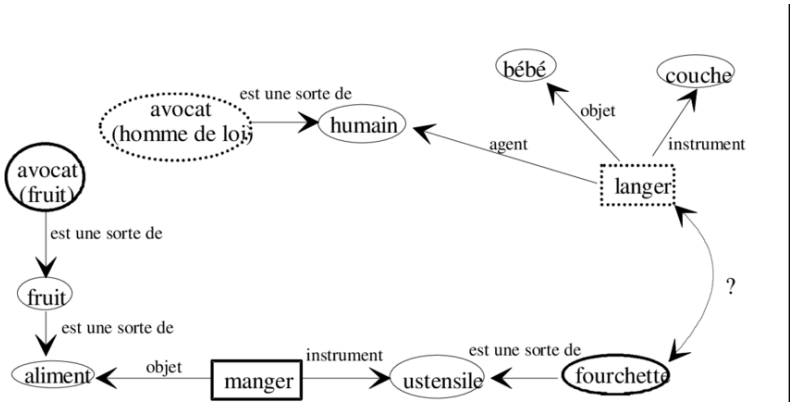


Figure 9 : Exemple d'un réseau sémantique (Ghio 1997, 47)

Sur cette figure 9, il est difficile de comprendre la symbolique des formes (ovale, rectangle : nous pouvons seulement interpréter par reconnaissance grammaticale que les verbes sont situés dans des rectangles, les substantifs dans les ovales) ainsi que celle de l'épaisseur des bordures, non clairement explicitées dans le document original de Ghio (Ghio 1997, 47). Si nous considérons ces ovales et ces rectangles comme des « nœuds », il nous semble qu'ils n'ont pas la même nature, de même que ce qui s'apparente à des liens orientés (flèches accompagnées d'une inscription) qui sont à la fois des rôles sémantiques (objet, agent, instrument) et des relations de spécialisation/généralisation (est une sorte de). Nous remarquons ici, comme le mentionne Hernandez (2005, 39), le manque de précision sémantique de ce réseau, menant à des confusions entre les relations, les classes et les individus. Nous n'utiliserons donc pas ce type de représentation dans nos travaux.

Les graphes conceptuels sont une évolution par rapport aux réseaux sémantiques. Malaisé revient sur l'origine de ceux-ci :

> « Selon leur fondateur, John F. Sowa, les graphes conceptuels (GC) sont un système de logique basé sur les graphes existentiels de Charles Sanders Peirce et sur les réseaux sémantiques de l'intelligence artificielle. Ils expriment le sens sous une forme précise d'un point de vue logique, lisible par un humain et manipulable par un outil informatique. […] Avec leur représentation graphique, ils servent de langage de modélisation et de spécification lisible tout en étant formel » (Malaisé 2005, 35).

Ainsi, les trois composants des graphes conceptuels sont : la logique, la lisibilité humaine et la manipulation informatique.

Comme le mentionnent Baget, Canet, Euzenat et *al.*, la différence entre les réseaux sémantiques et les graphes conceptuels se situe au niveau de l'avancée de ces derniers en structuration des connaissances :

> « Cette structuration des connaissances a été apportée à la fois par les logiques de descriptions (A-Box et T-Box (Baader et *al.* 2003)) et les graphes conceptuels (graphe et support) par rapport à leur ancêtre commun, les réseaux sémantiques. » (Baget et *al.* 2003, 20)

Cela s'exprime au niveau des relations qui les structurent (Rastier 1996). Tandis que les relations des réseaux sémantiques de type arborescents sont « verticales d'héritage, fondée sur l'équivalence partielle des nœuds » (Rastier 1996, s.p.), les relations des graphes conceptuels sont des relations « fondées sur la différence des nœuds, comme les relations casuelles ». Rastier précise que Sowa doit « une partie de son succès à sa capacité d'intégrer ces deux sortes de relations » (Rastier 1996, s.p.).

Pour illustrer cette notion de graphes conceptuels, nous nous appuyons sur les travaux de Yengui qui s'insèrent dans le cadre de la RI (Yengui 2016). Ses travaux traitent d'un type particulier de document audiovisuel que sont les visioconférences. Dans ce cadre, il représente les concepts et les relations sous forme d'un graphe conceptuel (figure 10) : les concepts sont représentés entre crochets et sont reliés entre eux par des relations représentées par deux flèches --> entourant le concept de relation qui, lui, est entre parenthèse. Ainsi, le concept [*surgical act*] est relié au concept [Strasbourg] par la relation (*used*). La thèse de Bui explique bien ceux-ci (Bui 2003, 75).

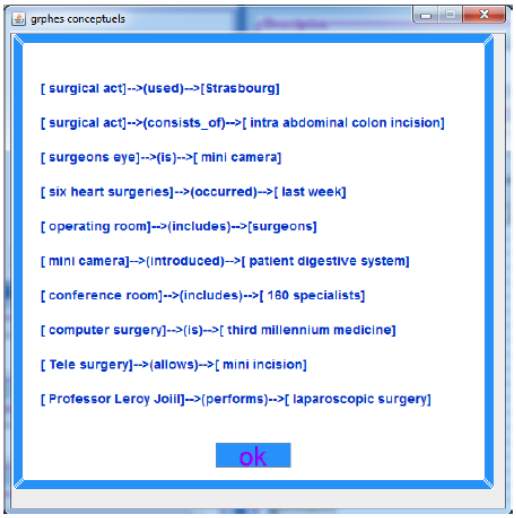


Figure 10 : Index sous forme de graphe conceptuel relatif à la visioconférence « Vidéochirurgie à Strasbourg » issu de (Yengui 2016, 159)

1.2.2.3 Représentations ontologiques

Une représentation ontologique est une modélisation, c'est-à-dire une représentation structurée de la connaissance utilisant des concepts (nœuds) liés (par des liens de type relation conceptuelle) en réseau (par exemple, les objets physiques et le temps). Une ontologie est en soi une représentation (qui peut être mentale et/ou graphique). C'est le résultat d'une vision partagée par une communauté d'acteur qui souhaite s'accorder sur le vocabulaire à employer et l'expliciter dans un domaine donné (ceci pour l'ontologie de domaine). Un langage spécifique est utilisé pour représenter les ontologies, *Web Ontology Language* (OWL), que nous développons en 1.2.4.

Les nœuds ne sont pas représentés par des unités lexicales mais par des formalisations de celles-ci et en tout cas les modèles « ne sont pas conçus pour accueillir des concepts désignés par des unités lexicales » (L'Homme 2002, 7-8).

1.2.2.4 Systèmes de représentation des connaissances

L'expression « *knowledge representation system* » est traduite par « système de représentation des connaissances ». Le Grand Dictionnaire Terminologique (GDT)[31] définit ce type de système dans le domaine de l'intelligence artificielle comme un « système de symboles et de principes utilisé pour représenter des connaissances utilisant des méthodes de représentation des connaissances telles que la logique des prédicats, les réseaux sémantiques, les frames, les espaces de problèmes... ». Nous plaçons dans cette catégorie toutes les représentations graphiques de connaissances que nous venons de voir, c'est-à-dire, les cartes heuristiques, conceptuelles, sémantiques, les graphes conceptuels et les représentations ontologiques. Dans nos travaux, nous utiliserons les cartes conceptuelles pour représenter notre réduction du modèle ontologique du CIDOC CRM, puis les graphes pour la visualisation de l'instanciation de ce sous-modèle à partir d'un entretien avec un ancien mineur.

[31] http://gdt.oqlf.gouv.qc.ca/Resultat.aspx

Les représentations graphiques tentent de rendre visibles les représentations mentales. Dans le domaine de la connaissance, la difficulté réside dans la représentation du sens et dans la standardisation de ces représentations.

1.2.3 Représentation globale des connaissances et web sémantique

D'après Charlet, Laublet, et Reynaud, la condition pour avoir une vision globale des connaissances est de réaliser « le même processus de globalisation sur la représentation des connaissances que celui que le Web fit initialement sur l'hypertexte » (Charlet, Laublet, et Reynaud 2003, iii)[32]. Pour comprendre cette idée, il faut définir d'une part les hypertextes (ou hyperliens) et d'autre part les évolutions du web depuis les années 90. Les hypertextes ont en effet été une grande innovation de ces années-là. Le vocabulaire de la documentation les définit comme un :

> « Système de renvois permettant de passer directement d'une partie d'un document électronique à un autre ou d'un document à d'autres documents choisis comme pertinents par l'auteur » (Boulogne et Institut national des techniques de la documentation 2004, 116).

Le web sémantique ou web de données ou encore web 3.0 vient comme son nom le suggère après le web 1.0 et le web 2.0. Le web 1.0 ou web documentaire est un web « où les ressources sont publiées sans autre traitement que leur mise en forme et sans autre interactivité que l'activation des liens hypertextuels » (Bachimont et *al.* 2011, 24). Ce web ne prend en compte que le format de codage des contenus et pas la sémantique de ces derniers (Bachimont et *al.* 2011, 25). Le web 2.0 ou web participatif a rendu le web accessible aux utilisateurs en leur permettant de publier du contenu avec des outils de plus en plus *userfriendly*. C'est notamment dans ce cadre qu'apparaît l'indexation sociale ou *bookmarking* social. Les utilisateurs contribuent en effet à indexer le web avec des mots-clés ou *tags* libres sur des plateformes collaboratives, permettant de partager du contenu. Des outils facilitant la veille comme les favoris, signets, *bookmarks*, marque-page (qui correspondent à des pages web choisies par l'utilisateur puis mises en mémoire sur les différents navigateurs Chrome, Firefox, Internet explorer) peuvent ainsi être indexés (Doria 2013). Enfin, le web 3.0, web sémantique ou web de données a été proposé par Tim Berners-Lee. Troncy le définit comme « une infrastructure juxtaposant au web actuel des documents structurés par des langages (pour exprimer la connaissance, pour décrire les relations entre les connaissances, pour décrire les conditions d'utilisation, les garanties et les modes de paiement) et des dispositifs permettant de trouver les ressources » (Troncy 2005, 393). Si l'on prend les connaissances au sens de structure, on peut dire que le web sémantique tend à jouer le rôle des métaconnaissances, mais à destination des machines : « le but du web sémantique est de développer un web dont le contenu s'adresse, au moins pour partie, aux machines, afin qu'elles puissent aider les utilisateurs humains » (Troncy 2005, 392-393) d'après (Charlet, Laublet, et Reynaud 2003). Par rapport

[32] « […] nous sommes bien conscients que le Web sémantique demande une approche globale et sera ce que nous obtiendrons si nous réalisons le même processus de globalisation sur la représentation des connaissances que celui que le Web fit initialement sur l'hypertexte. » (Charlet, Laublet, et Reynaud 2003, iii)

au web 1.0 et 2.0 où les hyperliens ne portaient aucune information concernant leur nature, les hyperliens du web 3.0 sont devenus sémantiques. Le web 3.0 s'appuie sur le modèle de graphe *Resource Description Framework* (RDF) qui intègre de nombreuses applications utilisant le langage XML pour la syntaxe et les URI pour la nomenclature (instance de classification). RDF exprime l'information grâce à un élément fondamental : un triplet de la forme (sujet, prédicat, objet). Le format est expliqué en 1.2.4.

Nous comprenons donc mieux l'engouement face à ce web qui, en reliant sémantiquement l'information, produit un réseau de connaissance représentable. La volonté de migrer toute ressource structurante sur le web prend ainsi son sens comme le rappelle Hernandez (2005, 69) à propos de la migration des thésaurus vers les ontologies du Web sémantique.

1.2.4 Langage de représentation des connaissances

Les langages de représentation des connaissances servent à « écrire » la visualisation d'une conceptualisation. Dans les années 2000, le langage de représentation sémantico-cognitif (LRSC), en informatique linguistique, voit le jour (Djioua 1999). Ce langage « permet de représenter des analyses sémantiques que l'on trouve en sémantique cognitive comme les structures conceptuelles de Jackendoff (Jackendoff 1983) et la grammaire cognitive de Langacker (Langacker 1987) » (Djioua 1999, 9). Émergeant dans une période de standardisation, il ne semble pas très utilisé. Six ans auparavant, en 1994, est fondé le *World Wide Web* consortium (W3C), organisme de standardisation. Le consortium recommande en effet un certain nombre de langages pour diverses utilisations. Ainsi, le langage XML est préconisé dans le balisage des documents (représentation de connaissance par objets).

Nous l'avons vu, la représentation des ontologies (Cf. 1.2.2.3) est une étape du processus de représentation des connaissances qui peut se baser sur un langage de représentation des connaissances appelé aussi « langage d'opérationnalisation » pour permettre la mise en action de « service inférentiels de type classification des concepts ou généralisation, etc. » (Charlet, Laublet, et Reynaud 2003, 49). Ceux-ci peuvent être les graphes conceptuels, les logiques de description et ils permettent :

> « […] l'un comme l'autre, d'effectuer un certain nombre d'opérations sur des ontologies : inférences propres aux structures de graphes comme la jointure ou la projection pour les graphes conceptuels, classifications dans des structures arborescentes pour les logiques de description. Les logiques de description sont le type de langage retenu par le Web sémantique pour représenter les ontologies et faire des inférences dessus. Les différents axiomes qui accompagnent chaque possibilité ne sont pas anodins en termes de connaissances représentables et de calculabilité » (Charlet, Laublet, et Reynaud 2003, 49).

Dans le cadre du web sémantique, le langage de représentation des ontologies est le *Web Ontology Language* (OWL). Celui-ci est « un langage de représentation des connaissances construit sur le modèle

de données de RDF pour définir des ontologies structurées »[33]. OWL permet par exemple de préciser l'identifiant d'une classe, l'union de deux ou plusieurs descriptions de classe ou l'intersection de deux ou plusieurs descriptions de classe.

Comme XML, RDF est un langage de balise qui permet la publication des données complètes directement comme des documents web qui peuvent être visualisés et réutilisés par n'importe quel utilisateur du web[34] (Gnoli 2020, 96). Pietriga et Lee expliquent la forme de la représentation de RDF :

> « Le sujet et l'objet sont mis en relation par le prédicat exprimé au moyen d'une propriété qui caractérise cette relation. Un modèle RDF est constitué par un ensemble de triplets formant un graphe dont les nœuds sont les sujets et les objets des triplets et dont les arcs sont les prédicats qui les mettent en relation. La relation exprimée par le prédicat pointe toujours du sujet vers l'objet ; le modèle de données de RDF repose donc sur une structure de graphe étiqueté et orienté […] « (Pietriga et Lee 2009, 175).

Les types de relations qui peuvent être représentées par RDF et XML sont les relations de subsomption et d'instanciation. La relation de subsomption[35] est une relation binaire entre un super-concept et un sous-concept, qui permet d'organiser les classes en hiérarchies. L'instanciation est une « [o]pération consistant à remplir au moyen de notions spécifiques les places d'un schéma de lexis. L'instanciation est une opération de choix notionnel « (Groussier et Rivière 1996, 106). *Resource Description Framework Schema* (RDFS) est enfin conçu pour pallier les limites du RDF. RDFS permet par exemple de préciser qu'une ressource RDF est une classe, inscrite dans une hiérarchie de classe et définit par exemple les notions de domaine et de rang où le domaine définit la classe des sujets liée à une propriété et où rang définit la classe ou le type de données des valeurs de la propriété. Par exemple, dans l'ontologie *Friend of a Friend* (FOAF)[36], la propriété *knows* relie un sujet qui est une personne à un objet qui est une autre personne :

foaf:knows rdfs:domain foaf:Person

foaf:knows rdfs:range foaf:Person

Dans le champ du web sémantique, Pietriga et Lee étudient plus précisément les langages et outils pour la visualisation et la manipulation de données (Pietriga et Lee 2009). Les auteurs montrent notamment l'intérêt des représentations visuelles de graphes RDF mais précisent que « leur portée est limitée au niveau de la taille des modèles auxquels elles peuvent faire face » (Pietriga et Lee 2009, 180). Les auteurs mentionnent également l'hétérogénéité des outils due au fait qu'RDF ne définit pas de modèle

[33] http://www.w3.org/2004/OWL/
[34] « […] *markup languages and formats for the Semantic Web, such as XML and RDF, allow publication of full data directly as web documents that can be visualized and reused by any web user.* » (Gnoli 2020, 96)
[35] « La subsomption est une méthode d'inférence en logique terminologique qui permet de classifier un concept dans une taxonomie de concepts » (Groussier et Rivière 1996, 106)
[36] « *FOAF is a computer language defining a dictionary of people-related terms that can be used in structured data* » http://www.foaf-project.org/

de présentation des données et la difficulté engendrée par leur manipulation pour les présenter à leurs utilisateurs.

Enfin, le CIDOC CRM, modèle que nous utilisons dans nos travaux (Cf. 3.4.1 et 5.4 et 6.3) est également étudié comme langage de représentation des connaissances sous la forme d'un wiki sémantique (Ribaud et Boeuf 2010). L'expérience est réalisée dans le cadre d'un cours et a pour but d'attirer des élèves (notamment du sexe féminin) vers les études d'informatique grâce à l'étude de la facilité d'appropriation du CIDOC CRM. Les résultats montrent que si le cours n'a pas eu l'effet escompté, l'aspect « gestion des connaissances » est ce qui intéresse le plus les étudiants et ce succès pourrait contribuer à « la diffusion des ontologies et du Web sémantique en général, et du CIDOC CRM en particulier » (Ribaud et Boeuf 2010, 1).

1.2.5 Représentation des connaissances sensibles

Nous considérons le savoir comme une compétence acquise par expérience vécue (et ressentie, donc avec les sens) dans un domaine pratique (domaine au sens large). Ceci nous amène à nous questionner : Comment représenter l'aspect sensible des informations ? Que fait-on du toucher, de l'olfaction, du goût dans l'ordinateur ? En psychologie, une distinction entre *sens affectifs* qui provoquent généralement une sensation de plaisir ou de douleur (odorat et goût) et *sens intellectuels* (vue, toucher et ouïe) qui sont le plus souvent à l'origine de nos représentations du monde extérieur est effectuée. Comment peut-on lier les objets du réel et ceux du virtuel pour permettre à l'utilisateur d'avoir accès à un maximum de sensations ? Il est tout à fait possible de mettre du parfum sur du papier, de recréer une ambiance passée dans un lieu actuel, de mettre du son, des images animées sur l'ordinateur, de reproduire plusieurs dimensions, mais il n'est toujours pas possible de goûter à la représentation visuelle d'un kiwi sur l'ordinateur, ni sur le papier. Comment transmettre l'odeur des choses ? Les objets de musée ne peuvent normalement pas être touchés, comment appréhender le poids d'un matériel pour se rendre compte de la pénibilité d'un travail au fond de la mine ? Est-ce que le simple fait de connaître le poids permet de se rendre compte ? Ce sont autant de questions qui rappellent que les représentations des connaissances ne devraient pas être que visuelles et sonores.

1.3 Modélisation des connaissances

Notre société et économie du savoir nécessitent de gérer les connaissances, de les acquérir et de les utiliser à bon escient pour pouvoir les transmettre. Pour cela, il faut les représenter d'une manière structurée et de façon à ce qu'elles soient perceptibles par des sens humains afin de les étudier : c'est la modélisation des connaissances. D'après le TLFi, la modélisation est une « opération par laquelle on établit le modèle d'un système complexe, afin d'étudier plus commodément et de mesurer les effets sur ce système des variations de tel ou tel de ses éléments composants ». Dans notre cas, le système complexe est l'ensemble des connaissances d'un domaine. Un système est un « ensemble structuré d'éléments abstraits, ensemble de concepts présentés sous une forme ordonnée » (TLFi). Le modèle à

réaliser est donc un système concret, qui peut s'établir grâce à des théories physiques, mathématiques, logiques, logico-mathématiques ou conceptuelles et qui permettent de représenter des structures fondamentales d'une réalité et d'en expliquer ou d'en reproduire dynamiquement (c'est-à-dire en prenant en compte les évolutions des éléments intrinsèques) son fonctionnement. Dans ce sens, Dalbin pose la question de l'intégration de la modélisation dans les systèmes documentaires en sciences de l'information et de la documentation et définit le modèle comme :

> « une représentation abstraite de la réalité au sens où il simplifie cette réalité en vue d'une utilisation précise, pour n'en conserver que les caractéristiques intéressantes par rapport au contexte ou au domaine dans lequel on se trouve. Le modèle est donc une vue subjective mais pertinente de la réalité. Il permet de simuler le fonctionnement de l'élément étudié. […] La modélisation et les modèles sont des aides à l'élaboration et à la structuration des idées, un support au raisonnement » (Dalbin 2003, 226).

En somme, la modélisation des connaissances serait une sélection pertinente de celles-ci et une simulation de leurs évolutions.

Notre domaine d'intérêt est le patrimoine et dans ce cadre, Du Château remarque que « les travaux dans le domaine de la modélisation des connaissances pour le patrimoine culturel sont relativement peu nombreux et cela malgré l'intérêt et les efforts que les institutions officielles nationales et internationales accordent à sa gestion et sa valorisation » (Du Château 2010, 73). L'auteur présente le modèle du CIDOC CRM comme une réponse pour la modélisation des connaissances des œuvres d'art. Dix ans plus tard, le modèle est toujours une référence dans le domaine du patrimoine culturel et de plus en plus de travaux l'utilisent encore selon les usages décrits par Du Château comme « modèle pivot » entre des données hétérogènes par exemple dans des travaux comme (Almeida, Freire, et Monteiro 2021) dans lesquels l'infrastructure portugaise ROSSIO est l'équivalent de DARIAH, l'infrastructure européenne de recherche pour les arts et les humanités et le thésaurus ROSSIO[37] est établi à partir des grandes classes du CIDOC CRM telles que *Agents*, *Conceptual objects*, *Physical objects*, *Places* et *Temporal entities*. Le CIDOC est toujours utilisé comme outil d'interrogation sémantique dans des projets comme l'Atlas Historique du Limousin (Morel 2018) ou encore comme un outil d'extensibilité de modèle existant comme pour le projet SCULPTEUR (Addis et *al.* 2003) qui a pour objectif de développer un système pour la navigation et la recherche dans les collections de musées à l'aide de métadonnées textuelles. Concernant ce projet, Gala et *al.* précisent : « *it extends CIDOC CRM to include concepts such as objects digital representations and their respective feature vectors* » (Gala et *al.* 2021). Comme le mentionne Picard à propos des modélisations :

> « il est nécessaire d'encoder les modélisations afin d'en faire des ontologies formelles autrement appelées informatiques ou encore ontologies computationnelles par anglicisme, afin de les rendre exploitables par les machines. Les langages alors utilisés

[37] http://vocabs.rossio.fcsh.unl.pt/rossioTesauro/en/

rendent exploitables les axiomes, c'est-à-dire les règles appliquées aux concepts ou aux relations, pour réaliser des inférences » (Picard 2014, 26) (Cf. 3.4.4).

Le passage des modélisations aux ontologies se réalise donc par encodage (également appelé implémentation) qui consiste à transformer les données en informations grâce à des codes spécifiques comme des langages de programmation. L'encodage étant une interprétation (Stutzmann 2013, 88), il mène forcément à un biais. Pour un biais minimal, (Charlet, Laublet, et Reynaud (2003, 53) citant Gruber (1993) « [l]'ontologie doit être conceptualisée indépendamment de tout langage d'implémentation » pour « permettre le partage des connaissances de l'ontologie entre différentes applications utilisant des langages de représentation différents ».

Nous verrons que le modèle ontologique que nous utilisons (CIDOC CRM) est encodé dans plusieurs langages spécifiques ce qui permet la publication d'information sous forme de données liées.

1.3.1 Importance de la structuration

Nous l'avons vu plus haut, la structure fait référence à un ensemble cohérent et construit, à l'agencement d'unités formant un système. La structure répond à des enjeux d'organisation et de représentation pour créer de la connaissance. Les connaissances étant des structures, il paraît évident que la structuration soit utilisée pour les représenter, mais ceci n'explique pas ce qu'est la structuration. Il s'agit d'un processus de segmentation de la connaissance en unités plus petites (granularité), en information puis en donnée ou inversement d'organisation, d'ordonnancement des données entre elles pour former des informations, puis de la connaissance. Le processus de structuration s'applique à tous les objets de l'OC (et tout objet de la connaissance en général) et se réalise selon différents modes régis par des normes et standards.

Nous présentons ci-dessous une liste non exhaustive des cooccurrents « objet » du terme de structuration dans notre corpus de sources scientifiques. Nous les avons regroupés par type d'objet puis sous-type d'objet. Ainsi, il est possible de structurer :

- des données, des métadonnées, des informations, des connaissances, des savoirs,
    - du lexique, des données terminologiques (terminologies, termes), des unités lexicales et sémantiques, l'environnement sémantique d'un descripteur, des réseaux sémantiques, des schémas conceptuels,
    - des activités (ici, culturelles),
    - des documents, des contenus, des discours, des corpus (multilingues), des pages, le web
    - des recherches, des idées,
    - des ontologies, des thésaurus, des classifications (à facettes),
        - des domaines,

- des notions, des concepts,
- des hiérarchies,
- des relations,

- des tableurs…etc.

Ces éléments pouvant être structurés peuvent être eux aussi des structures pour d'autres objets ou entre eux. Notre analyse a aussi permis de relever des objets de structuration, des modes de structuration et plusieurs types de structuration. Par exemple, les langages et formats informatiques sont des moyens qui permettent de structurer un type d'objet de l'OC qui est l'information. Dans ce cadre, le langage XML a pour objectif de catégoriser l'information pour pouvoir la traiter (Salvador 2016). La structuration est alors réalisée par identification de catégories pertinentes. D'autres exemples peuvent être cités : la connaissance peut être structurée grâce à des outils de TAL ou des logiques de description ou encore des graphes conceptuels. Le lexique peut se structurer grâce à des questions posées à des experts dans une discipline. Des activités culturelles peuvent être structurées par une convention de développement culturel, un champ par un index terminographique, un tableur par des onglets et des colonnes. Nous avons finalement, lors de cette requête, pu établir un tableau (tableau 3) nous permettant de relever les différents types de structuration (selon l'objet structuré, selon le mode de structuration employé, et selon le format).

Tableau 3 : Exemples de structurations organisées par type, objet, mode et type de format issues d'un corpus de textes scientifiques

Type de structuration	**Objet de la structuration**	**Mode**	**Type de format**
conceptuelle	concept	système conceptuel	
hiérarchisée		hiérarchie	arborescence
hiérarchique	des descripteurs	hiérarchie	
dynamique			
segmentale		segment	
épistémologique			
visuelle	de l'information	langage de balise	au format HTML
sémantique	des espaces documentaires	sens	
sémantique	des connaissances	sens	
documentaire	document		
ascendante			
différentielle			
	corps d'un thésaurus	en blocs	
logique		concepts, notations formelles, règles de la logique	
compréhensible		par l'outil informatique	
formelle			
évidente			
plus fine			
dialectale			
intratextuelle			

Nous verrons dans les parties suivantes quel type de structuration est utilisée dans quel outil classificatoire. Nous souhaitons structurer sémantiquement les connaissances, c'est-à-dire que notre représentation sera faite de sens. La structuration peut être formelle, hiérarchique, logique, ascendante

et sémantique, ces adjectifs ne s'excluant pas entre eux. Comme il existe des langages de représentation, il existe des logiques et des langages de structuration. La représentation des connaissances structurées s'appelle la modélisation des connaissances.

1.3.2 Problème des ontologies

Pour Bachimont, la représentation des connaissances est indépendamment liée au problème des ontologies, problème qui doit être résolu (Bachimont 2000). Dans ce cadre, il mentionne les trois points que nous développons ci-dessous dont la modélisation ontologique, la généricité des ontologies et la linguisticité des connaissances.

D'après lui, le formalisme des langages de représentation nécessite de traiter le problème de modélisation ontologique. Il s'agit de définir les « primitives de représentation et leur signification qui seront utilisées pour la modélisation formelle des connaissances » (Bachimont 2000, 2). Nous verrons que dans le contexte des ontologies, les primitives de représentation peuvent être des classes (ou des ensembles), des attributs (ou des propriétés), et des relations (ou des liens qui relient des éléments de classe). À cela s'ajoute des contraintes sur l'application de ces primitives qui doivent être logiquement cohérentes.

La généricité des ontologies pose souci dans le cadre de leur réutilisation. Bachimont propose l'hypothèse qu'« une ontologie peut être générique pour un ensemble de tâches données si elle possède un niveau de finesse de description permettant aux points de vue de chaque tâche de se représenter [mais qu'il] n'existe pas de niveau de description permettant de décrire toutes les tâches du domaine. Il n'y a donc pas de généricité universelle ou absolue, mais de généricité relative à des tâches. » (Bachimont 2000, 14). Hernandez pense au contraire que les ontologies sont réutilisables et propose pour cela d'évaluer celles-ci selon des analyses qualitatives (taux de pertinence) ou quantitatives (réutilisabilité d'une ontologie par rapport à son efficacité dans la réalisation d'une tâche donnée) » (Hernandez 2005, 83).

La « linguisticité des connaissances » (Bachimont 2000, 2) ou l'aspect fondamentalement linguistique de la pensée signifie qu'il faudrait étudier la représentation des connaissances d'un domaine uniquement grâce aux signes linguistiques. Or, nous pensons que les connaissances dite « extralinguistiques » ne peuvent être représentées par ces signes. Nous l'avons vu, ces derniers se composent d'un signifiant (image acoustique d'un signe, empreinte psychique de celui-ci) et d'un signifié, (considéré parfois comme un concept, ou non (Roche 2007)). Les questions linguistiques quant à l'arbitraire du signe sont remises en cause par Bohas (2016, résumé) qui démontre « l'élément minimal significatif est en réalité le trait phonétique, composante du phonème (lui-même constituant du morphème), lequel, en tant que propriété articulatoire, établit un lien naturel (donc non arbitraire) avec le signifié ». L'auteur démontre empiriquement que le trait phonétique est pertinent pour représenter la structuration lexicale et que les formes n'ont rien d'arbitraire relativement aux significations. Nous ne développerons pas cet aspect mais

pensons qu'il est important à prendre en compte dans des études futures pour établir le lien entre représentation mentale et production mentale, sonore ou écrite.

D'autre part, la linguisticité des connaissances soulève deux autres problèmes qui sont : l'ambiguïté du langage naturel qui motive la désambiguïsation (ce que fait la terminologie) puis, le fait qu'un signifié n'est pas un concept comme le signale Roche (2007). Afin de pouvoir représenter ces problèmes, Roche fait appel aux ontologies pour représenter les systèmes conceptuels des termes, c'est l'ontoterminologie (Roche 2008 ; Roche et *al.* 2009 ; Roche et Papadopoulou 2020).

Dans le cadre des approches documentaires, un nouveau problème est évoqué par rapport aux ontologies formalisées (représentées à l'aide de RDFS et OWL) : « elles ne capturent pas l'intégralité des référentiels utilisés en documentation (thésaurus, classifications, taxonomies, dénommés ici *KOS - Knowledge Organization Systems*) » (Dalbin et *al.* 2011, 48). Pour pallier ce problème, l'ontologie SKOS a été proposée puis recommandée dans le cadre du *W3C* pour faciliter la publication de vocabulaires structurés pour leur utilisation dans le web sémantique.

CHAPITRE 2. Approches documentaires de l'organisation des connaissances

Ce chapitre présente les différentes approches documentaires de l'OC en orientant les propos sur les SOC, sur leur hétérogénéité et le passage de ceux-ci au web sémantique. Nous revenons sur la question de l'interopérabilité des SOC et traitons en particulier les SOC culturels dans le cadre du web sémantique.

2.1 Systèmes d'organisation des connaissances (SOC)

Les SOC étudiés en organisation des connaissances sont en constante évolution et sont conçus par une communauté de pratiques hétérogènes, les rendant par conséquent hétérogènes. Afin d'étudier la variété des SOC, Gnoli[38] mentionne l'existence en ligne des répertoires donnant accès à un certain nombre de SOC tel que BARTOC *Basel Register of Taxonomies, Ontologies and Classifications* qui contient des milliers de SOC, comprenant des systèmes conçus pour les bibliothèques, les musées, les écoles et la documentation institutionnelle (Gnoli 2020). Récemment Zacklad (2018) redéfinit sa typologie de SOC qu'il avait effectuée dans (Zacklad 2010) en différenciant quatre types de SOC : symboliques, algorithmiques, visuels et incarnés (ou incorporés) et insiste sur le fait :

> « qu'il n'y aura pas de contribution pertinente des SIC à l'analyse et à la conception des dispositifs de médiation des activités individuelles et collectives sans capacité à les articuler de manière fine et complémentaires pour rendre compte de la diversité croissante des usages dans les environnements postdigitaux actuels » (Zacklad 2018, 103).

De ce point de vue, nous souhaitons travailler sur l'articulation et la complémentarité de deux types de SOC que Zacklad définit comme SOC symboliques : des SOC épistémiques généralistes tels qu'une terminologie et un thésaurus ainsi qu'un SOC cognitif et logique telle qu'une ontologie formelle afin de rendre compte de cette diversité.

Les SOC sont traités dans (Hodge 2000) avec un point de vue particulier dédié aux bibliothèques et l'ouvrage est particulièrement utile aux institutions qui se lancent dans des initiatives de bibliothèques numériques. Maniez quant à lui, interroge le terme SOC par rapport à celui de langages documentaires et montre en quoi ce premier à tendance à remplacer ce deuxième sans en être pour autant une catégorie générique (Maniez 2007). Hjørland inclut les SOC dans le champ de l'OC et agrandit le périmètre de leur étude, ils ne concernent pas que la bibliothéconomie (Hjørland 2008). Dans le domaine informatique, Vandenbussche étudie l'hétérogénéité et l'interopérabilité sémantique au sein d'un système d'information pour la définition d'un cadre formel de représentation des SOC (Vandenbussche 2011).

[38] « *an extensive online directory of thousands of KOSs, wich includes systems conceveid for libraries, for museums, for schools, for institutional documentation etc*. » mentionné dans (Gnoli 2020, 73)

L'hétérogénéité et la multiplication des SOC dans diverses applications pose la question de leur différenciation et la recherche constante de critères pertinents pour les reconnaître, les concevoir, les évaluer. Tandis que Zacklad propose des critères de définition et d'évaluation des SOC (Zacklad 2018) (Cf.1.1.1), Souza, Tudhope, et Almeida réfléchissent quant à eux à des critères pour établir une taxonomie des SOC (Souza, Tudhope, et Almeida 2012). Ils en retirent que : « *The visual rhetoric is that of a linear progression, with ontologies as the ultimate form* » (Souza, Tudhope, et Almeida 2012, 185). Les auteurs proposent leur propre taxonomie qu'ils présentent sous la forme d'un arbre hiérarchique (figure 11), rompant ainsi avec les représentations précédentes comme des continuums ou des spectres linéaires. Les auteurs distinguent les dimensions intrinsèques des dimensions extrinsèques des SOC : intrinsèque considère le SOC en tant qu'entité idéale, isolée de tout utilisateur et de l'environnement tandis que les dimensions extrinsèques sont liées à l'environnement dans lequel le KOS est utilisé.

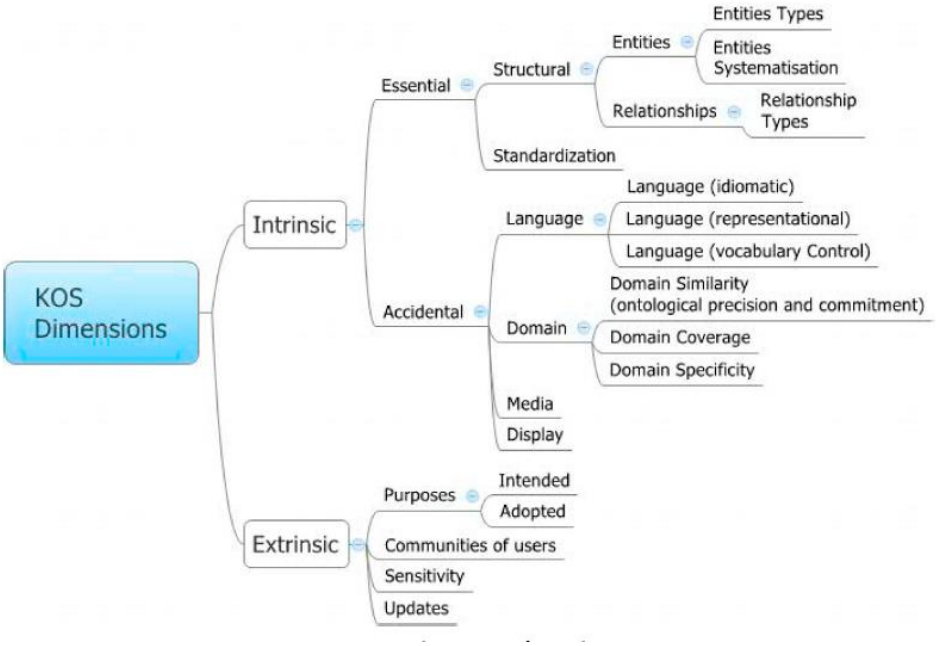


Figure 11 : « *Proposed taxonomy of KOS dimensions* » dans (Souza, Tudhope, et Almeida 2012, 189)

2.2 Types et hétérogénéité des SOC

La norme ISO Livre blanc ISO 25964-1 - Thésaurus pour la recherche documentaire, édition de janvier 2013 présente la nouvelle norme de thésaurus. Dans celui-ci, « schéma de classification », « vocabulaire contrôlé », « thésaurus multilingue », « schéma de vedettes matières / langage de vedettes-matières / liste de vedettes matières », et « thésaurus », termes traduits par les auteurs à partir de l'anglais sont considérés comme faisant partie des systèmes d'organisation de concepts (tableau 4) que nous pouvons considérer comme faisant partie des systèmes d'organisation des connaissances.

Tableau 4 : Termes français et anglais issus du livre blanc ISO 25964-1 pour le domaine Système d'organisation de concepts			
SYSTEMES D'ORGANISATION DE CONCEPTS			
fr2.39	2.06	schéma de classification	classification scheme

Tableau 4 : Termes français et anglais issus du livre blanc ISO 25964-1 pour le domaine Système d'organisation de concepts			
SYSTEMES D'ORGANISATION DE CONCEPTS			
fr2.63	2.12	vocabulaire contrôlé	controlled vocabulary
fr2.62	2.35	thésaurus multilingue	mutilingual thesaurus
fr2.64	2.56	vocabulaire structuré	structured vocabulary
fr2.40	2.57	schéma de vedettes matières, langage de vedettes matières, Liste de vedettes matières	subject heading scheme, subject heading language, subject heading list, SHL
fr2.61	2.62	thésaurus	thesaurus

Nous introduisons tout d'abord la notion générique de vocabulaire contrôlé que nous opposons au vocabulaire libre, puis développons la catégorie spécifique des SOC appelée « langage documentaire » dans laquelle nous plaçons les classifications, les taxinomies, les thésaurus et terminologies, les répertoires (ou schéma) de vedettes matières, puis nous traitons les mots-clés et folksonomies à part comme autre langage documentaire qui utilisent souvent des vocabulaires non contrôlés.

2.2.1 Vocabulaire contrôlé vs vocabulaire libre

Un vocabulaire contrôlé est une liste de termes approuvée par une communauté ou une organisation. Par exemple, les jours de la semaine ou les mois de l'année sont des termes que tout le monde utilise sans ambiguïté, ils forment donc un vocabulaire contrôlé. Dans le cadre de l'utilisation du langage pour indexer des documents et pour pallier la variation et l'ambiguïté du langage naturel, un travail de raisonnement sur ce vocabulaire et de normalisation est effectué. Ceci dans le but de servir l'indexation sujet et de permettre de désigner le plus clairement possible les sujets traités dans un contenu. On y identifie alors les groupes de termes équivalents. Le langage contrôlé, quant à lui prévoit et normalise la combinaison des signes afin de réaliser du sens. Lorsque que l'on construit un langage et un vocabulaire contrôlé, des règles d'écriture (par exemple masculin singulier pour les noms) doivent être adoptées pour l'homogénéité et la collection de termes d'un vocabulaire donné doit être structurée avec des groupements et distinctions allant du général au particulier[39]. En 2002, Maniez souligne que « la plupart des recherches sont orientées vers la modélisation des langages hiérarchiques (classifications et thésaurus) et vers le traitement automatique du langage » (Maniez 2002, 302). Il souligne aussi, à propos de la recherche d'information sur internet, qu'un « vocabulaire contrôlé de référence, assorti si possible d'un thésaurus, serait la meilleure solution d'avenir » (Maniez 2002, 311). Nous développons ci-dessous un type de vocabulaire contrôlé qui est le langage documentaire.

2.2.2 Langages documentaires

D'après Maniez, « les langages documentaires ne sont pas autre chose que des tentatives de purification et de normalisation des langues humaines en vue de faciliter la recherche des informations » (Maniez 1999, 251). Aussenac-Gilles, Baziz, et Hernandez définissent quant à eux les langages documentaires comme des ensembles organisés de termes normalisés d'un domaine, utilisés pour indexer des

[39] http://www.thesaurus.gouv.qc.ca/tag/terme.do?id=MDL1571

documents (Aussenac-Gilles, Baziz, et Hernandez 2006). Ils sont utilisés pour caractériser des données ou le contenu de documents et en permettre le stockage et la recherche.

Pour le vocabulaire de la documentation (Boulogne et Institut national des techniques de la documentation 2004, 168) le langage documentaire n'est pas une catégorie de plusieurs langages mais est un « mot ou expression choisi généralement dans le titre ou le texte d'un document pour en caractériser le contenu et en permettre la recherche. Il constitue un point d'accès. Il est à distinguer d'un descripteur, qui est un terme normalisé dans un thésaurus ».

D'après, Wersig (à partir du GDT[40]), les langages d'indexation (LI) sont spécifiques aux langages documentaires, au même titre que les langages de descripteurs, les langages de recherche documentaire, les thésaurus, les classifications et le code sémantique. Comme les langages documentaires, les langages d'indexation permettent d'exprimer le contenu sémantique d'un document. Ils s'opposent aux langages naturels (LN) et ont pour particularité d'être construits sur des langages formels. La structure des langages d'indexation a été établie sur des « modèles linguistiques ou mathématiques » (Rivier 1990, 4). Depuis l'apparition des ordinateurs et du numérique, la question de la syntaxe de ces langages ne se pose plus comme avant. Plusieurs types d'indexation existent, un exemple est l'indexation systématique de J.O. Kaiser, analysée notamment par Svenonius (Svenonius 1978) dans le cadre de la définition d'une facette (voir la partie 2.2.2.1) et définie par Rivier comme une indexation qui « consiste à analyser tous les sujets selon deux grandes catégories : les concrets (ou objets) et les processus (ou actions). Syntaxiquement, les concrets précèdent toujours les processus » (Rivier 1990, 17). Dans cet ouvrage, Rivier revient notamment sur la théorie de la construction des langages d'indexation et Allen et Aitchison produisent quant à eux, pour l'UNESCO[41], une bibliographie des langages d'indexation dans le domaine des sciences sociales (Allen et Aitchison 1984).

En 2007, Maniez se demande enfin si les langages documentaires sont devenus obsolètes sans donner de réponse puisque qu'il étudie plutôt leur dénomination au regard d'autres tels que SOC et langage contrôlé (Maniez 2007). Menon étudie quant à lui la formation et l'évolution des langages documentaires (Menon 2007). Comme Maniez (2007), il interroge le concept de langage documentaire par rapport à celui de système d'organisation des connaissances. Parmi les problématiques posées par Maniez dans ce cadre, celle de la compatibilité et de l'interopérabilité nous intéressent plus particulièrement. Ces questions, en matière de repérage d'information pertinente, sont notamment traitées dans (Feyler 2007) avec l'exemple d'OTAREN (Outil Thématique d'Aide à la Recherche pour l'Éducation Nationale), outil qui a évolué et qui a étudié l'interopérabilité entre plusieurs langages documentaires dont Motbis, la liste de vedettes-matières de Rameau, les thésaurus de l'ONU (UNBIS), de l'UNESCO, le Thésaurus Européen de l'Éducation (TEE) et le Thésaurus européen multilingue de

[40] GDT, article « Langage documentaire » de Wersig, Gernot 1976 : *Narrower term: descriptor language. Narrower term: indexing language. Narrower term: retrieval language. Narrower term: thesaurus. Narrower term: classification. Narrower term: semantic code*.
[41] La première édition du Thésaurus de l'UNESCO parue en 1977 est traduite en français en 1983.

l'Environnement (GEMET). La question de l'interopérabilité, au niveau sémantique, sera traitée dans le chapitre 10.

2.2.2.1 Classification documentaire

Les classifications documentaires les plus répandues dans les bibliothèques sont la classification décimale de Dewey et la classification décimale universelle (en Europe et en Afrique), toutes deux construites sur une approche épistémologique et divisant hiérarchiquement le savoir humain (Zacklad 2007). La classification de la bibliothèque du congrès (LCC) est aussi largement utilisée en Amérique du Nord dans les bibliothèques universitaires de la recherche comme le mentionnent Hudon et Hadi (2010). Ces trois classifications sont mises à jour régulièrement (Couzinet 2011). La classification décimale universelle a fait l'objet de plusieurs nouvelles éditions abrégées ce qui montre bien le caractère vivant des systèmes de classification et plus généralement de tout SOC.

La classification décimale universelle (UDC Consortium (The Hague) 2001) est issue de la 5ème édition de la classification décimale de Dewey (Hudon et Hadi 2010). Paul Otlet, bibliographe belge, tente de regrouper dans un même lieu organisé, le *Mundaneum*, l'ensemble de la production documentaire mondiale.

Les deux classifications sont souvent mises en comparaison : Zacklad les distingue par exemple sur leur usage : « Alors que la CDD requiert en principe une classification unique de chaque ouvrage, la CDU permet une combinaison d'indices qui, bien qu'accroissant la précision, n'en facilite pas toujours l'usage » (Zacklad 2007, 5). Toujours d'après Zacklad, la CDU possède un degré de formalisation méta sémiotique explicite au même titre que les thésaurus, les ontologies sémiotiques, les annuaires collaboratifs sur internet et les cartes conceptuelles contrairement à la Classification décimale de Dewey et les folksonomies qui n'ont pas de syntaxe pour la gestion combinée des termes et les ontologies formelles qui possèdent un méta langage formel.

Comme le signale Couzinet :

> « la question de l'organisation des savoirs a été abordée en France par des précurseurs des sciences de l'information et de la communication. Si l'on se réfère souvent à la classification élaborée par Melvil Dewey, aux États-Unis, ou à la Classification décimale universelle (CDU) établie par Paul Otlet et Henri Lafontaine, en Belgique, au début du vingtième siècle, ces dernières ont été précédées par bien d'autres et notamment par des classements de bibliothèques privées effectués par des érudits, des secrétaires ou bibliothécaires particuliers ou encore des religieux. Système de représentation des informations, l'organisation des savoirs est un des sujets centraux travaillés par les sciences de l'information et de la communication » (Couzinet 2006, 2).

D'autre part, le premier langage documentaire à facettes est créé en 1933 par Shiyali Ramamrita Ranganathan, mathématicien et organisateur du réseau des bibliothèques en Inde. Alors qu'il crée sa *Colon Classification*, dans la première moitié du XXe siècle, ce dernier élabore la notion de facette

« pour optimiser le classement des ouvrages en bibliothèque » (Amar 2000, 224). Cette notion de facette peut d'après Amar, « s'approcher en termes de « point de vue » sur un objet » (Amar 2000, 224). Ranganathan s'appuie sur les travaux d'Aristote et relève cinq catégories capables de concevoir un même objet sous différents points de vue. Ces catégories sont : Personnalité, Matière, Énergie, Espace, Temps (Ranganathan 1976).

Les langages documentaires à facettes permettent d'envisager un document sous plusieurs angles et l'utilisation « de facettes vise à introduire, dans les langages documentaires, une autre approche que celle du seul contenu : c'est à partir de propriétés extralinguistiques qu'un objet est classé » (Amar 2000, 224). Sur ce point, l'auteure précise que l'indexation par facette permet de classer les « interprétations d'un document » plutôt que le « document lui-même » et que celle-ci ne spécifie pas « le point de vue de l'interprétation (la stratégie de lecture adoptée) » (Amar 2000, 225).

Quelques années après la construction du premier langage documentaire, le *Classification Research Group* (CRG) est créé en 1952. Celui-ci pointe le souci qu'aucun schéma de classification général existant n'est satisfaisant pour organiser les documents de manière à pouvoir retrouver efficacement l'information cherchée (Classification Research Group 1969). Celui-ci se penche donc sur la définition de la classification à facettes. À partir des travaux réalisés par le CRG, Brian Campbell Vickery propose une autre méthode qui consiste à construire et utiliser des schémas spéciaux à partir de facettes domaniales (Vickery 1963). L'auteur précise que le résultat de l'analyse par facettes, si elle est « analogue aux règles traditionnelles de la division logique sur lesquelles la classification a toujours été fondée », diffère d'abord par « la rigueur avec laquelle les règles sont appliquées », puis par la liberté de combinaison des facettes pour « que chaque type de relation entre les termes et entre les sujets puisse s'exprimer » et par l'introduction de « nouvelles relations logiques […] reflétant ainsi plus parfaitement la complexité du domaine de la connaissance » (Vickery 1963, 8-9).

D'autres systèmes d'indexation sont ensuite conçus tels que le système d'indexation syntaxique PRECIS d'Austin en 1971, l'indexation relationnelle de Farradane (1950), les cadres sémantiques de Fillmore (1968) que nous ne développerons pas ici.

En 1999, Maniez mentionne que « la notion de facette apparait souvent comme « l'apport théorique le plus important de ce siècle en science de l'information » (Maniez, 1999, p. 249) mais celui-ci démontre en quoi l'approche de Ranganathan (1976) sur la classification à facettes ne convient pas. En effet, il « décrit ce qu'il considère comme une faiblesse de l'approche par facettes, soit la description ambigüe de la nature et du rôle de la facette qu'il en fait » (Hudon 2018, 71) et notamment par l'ambiguïté qui se retrouve dans les facettes Personnalité, Matière et Énergie qui sont aussi bien paradigmatiques que syntagmatiques. En 2002, l'Office québécois de la langue française la définit comme une « classification qui repose sur une analyse documentaire présentant et organisant, sur un même plan, l'ensemble des sous-caractéristiques qui peuvent diviser une facette, un sujet, ce qui a pour résultat de former des groupements homogènes qui s'excluent mutuellement ».

En 2010, La Barre propose d'appliquer la notion de facette au web sémantique (La Barre 2010)[42]. Desfriches Doria, dans un article pour la revue *Étude de communication* a cherché à « éclaircir dans quelle mesure [cette classification pouvait] servir d'outil d'organisation des connaissances, dans l'entreprise avec comme piste d'évolution, une proposition d'utiliser le tag de documents (Desfriches Doria 2012). Hudon et Hadi revisitent la classification à facettes (Hudon et Hadi 2017) et la même année Broughton montre en quoi la classification à facettes peut être étudiée comme la théorie générale pour le champ de l'OC (Broughton 2017b).

Depuis 2004, dans la lignée des travaux du CRG et plus précisément de Douglas J. Foskett and Derek Austinet est développée une classification à facette avec point de vue par phénomènes intitulée *Integrative Levels Classification* (ILC)[43]. La deuxième version de cette classification est actuellement disponible au format SKOS. Elle inclut 10845 classes et facettes couvrant l'ensemble du spectre des connaissances au sens large, ainsi que des spécificités plus profondes dans certains domaines (par exemple, les cétacés, les pays, ...) qui ont déjà été élaborés en détail.[44]

Le thésaurus à facettes, quant à lui est « un outil de classification simple et universel » et défini en 1983 par l'Association française de normalisation, comme un « thésaurus dans lequel les relations entre des termes sont établies après regroupement de ceux-ci par facettes ». D'après Maniez, s'il devenait la norme, [il] faciliterait la compatibilité entre les thésaurus » (Maniez 1999, 259).

Lors de nos travaux, nous avons pu échanger avec Claudio Gnoli qui nous expliquait que les facettes étaient des catégories de relations (partie, procès, agents), de la même manière que les propriétés des ontologies. D'après Vanda Broughton, coéditrice de la classification BC2[45],

> « *it is beginning to be taken up by researchers in the fields of automatic indexing and the semantic web as a conceptual tool to assist in the understanding of the most complex relationships between objects* » (Broughton 2006b, 68).

L'auteure développe également cette idée dans (Broughton 2009, 2011, 2017a). Comme nous l'avons indiqué plus tôt, nous ne traiterons pas de cette question dans nos travaux mais avons jugé pertinent de faire référence à ces travaux pour de futures investigations.

Enfin, encore très récemment, la notion de facette reste ambigüe et nécessiterait une définition commune au sein du champ de l'OC comme le montre Hudon (2020).

2.2.2.2 Taxinomie

Une taxinomie (anglais taxonomy), ou taxonomie est un vocabulaire contrôlé hiérarchique qui concerne un domaine ou une science, qui permet de classer des éléments en différentes catégories selon les

[42] « La classification à facettes suscite de plus en plus d'intérêt sur le web, et plus particulièrement dans le cadre d'un web que l'on veut sémantique. » (Hudon et Hadi 2010, 35).
[43] https://www.isko.org/cyclo/ilc
[44] Traduit par l'auteure du site : https://www.isko.org/cyclo/ilc
[45] BC2 ou New Bliss Classification est une classification bibliographique générale d'informations trouvées dans des documents et des dossiers de tout types réalisée à partir de la classification d'Henri Bliss et qui est considérée comme un système complémentement différent de A Bibliographic Classification (BC1).(Ranganathan 1984)

caractères qu'ils ont en commun, du plus général au plus particulier. L'établissement d'une taxinomie est réalisé d'une manière logique. Elle est depuis longtemps utilisée pour organiser les organismes vivants et notamment les plantes avec la présentation d'une première taxonomie végétale en 1813 dans (Candolle, Deterville, et Martel 1813), mais peut aussi très bien organiser les objets comme le mentionne Rey (1979, 35)[46]. La figure 12 réalisée par nos soins illustre un exemple de taxinomie catégorisant les animaux présents dans l'univers minier.

Tandis que la taxinomie est une classification (avec des classes pour représenter des catégories), le thésaurus porte sur la dualité langage/termes. Les thésaurus ont une structure plus souple que les taxinomies ; ils acceptent les termes « orphelins » c'est-à-dire des termes qui n'ont ni terme générique, ni terme spécifique, seule une relation associative, alors que la taxinomie ne l'autorise pas.

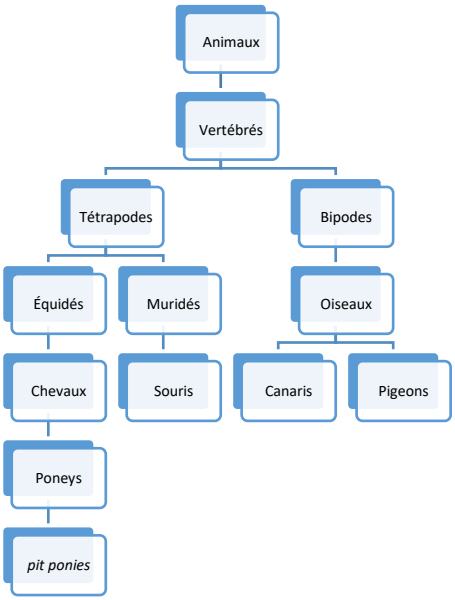


Figure 12 : Taxonomie des animaux présents dans l'univers minier, réalisée par l'auteure

2.2.2.3 Thésaurus et terminologie

Les terminologies et thésaurus ont fait l'objet de nombreuses études à la fin du XXe siècle poursuivies au XXIe. Ici terminologie ne désigne pas la discipline mais l'ensemble des termes et des expressions d'un domaine de connaissance comme résultat des choix théoriques appliqués sur la langue ou encore l'« ensemble des termes relatifs à un système notionnel élaboré par des constructions théoriques, par des classements ou des structurations de matériaux observés, de pratiques sociales ou d'ensembles culturels » (TLFi).

Leur différence réside en partie dans leur objectif de réalisation et leurs utilisateurs qui orientent le choix dans la méthodologie de construction. Le but des « thésaurus modernes » d'après Rivier est de

[46] Taxinomie : « Classification d'éléments; suite d'éléments formant des listes qui concernent un domaine, une science. Taxinomie bactériologique, botanique, chimique, zoologique. La taxonomie (...) classe les Animaux et les Plantes en différentes catégories (...). Elle range tous les êtres vivants suivant les caractères qu'ils ont en commun, des plus généraux (règne) aux plus particuliers (espèce) (Biol. t. 2 1970). Ces classes logiques, dont la constitution est étudiée par la taxinomie (anglais taxonomy), se manifestent bien dans le domaine des classifications hiérarchisées et systématiques des sciences naturelles (nomenclatures), mais aussi dans le domaine des hiérarchies d'objets spontanément observés et dénommés à l'intérieur d'une culture (folk taxonomies : taxinomies populaires) (A. REY, La Terminol., 1979, p. 35). »

« représenter d'abord des concepts, mais aussi de préciser leurs relations « (Rivier 1990, 267). Le premier thésaurus moderne (et opérationnel) a été publié en 1959 par la société Dupont de Nemours et est un thésaurus des termes de la chimie *Thesaurus of Engineering Terms*. En 1994, Condamines rapproche la terminologie de la représentation des connaissances (Condamines 1994)[47] expliquée par l'émergence d'un besoin en entreprise et propose le nouveau concept de « base de connaissance terminologique » en insistant sur le fait qu'« il est indispensable de créer de nouveaux modèles de stockage des données qui rendent compte à la fois du fonctionnement linguistique des termes et de leur lien fort avec les concepts du domaine » (Condamines 1994, 44). En 2001, l'emploi du terme thésaurus qualifié tantôt d'idéologique tantôt de documentaire par Péchoin peut permettre de créer ce lien fort (Péchoin 2001). Il définit en effet le thésaurus idéologique comme un outil servant à décrire la langue, et plus particulièrement le lexique. L'exemple le plus connu de ce type de thésaurus est le thésaurus de Roget (Roget 1982), qui se différencie de l'approche dictionnairique par sa perspective onomasiologique, en regroupant les termes en fonction des idées. Comme le signale Rivier, « à l'inverse des dictionnaires de langue habituels où l'on part d'un terme pour découvrir sa définition (donc son sens), Roget part du sens (les concepts) pour en donner tous les termes qui le représentent, sans définitions « (Rivier 1990, 19). Le thésaurus documentaire, quant à lui, est décrit comme un outil de contrôle du discours « celui de la description de certains textes (techniques ou scientifiques, le plus souvent)(Péchoin 2001, 112). Thésaurus idéologique ou documentaire, dictionnaire ou terminologie, la différence entre tous ces systèmes est bien l'objectif, qui est mis en avant pour les premiers par des adjectifs. D'un côté sont classées des idées, d'un autre, des documents. L'ambiguïté peut se révéler si l'on décide que les idées peuvent classer des documents. D'après Bernard, le thésaurus répond à la question suivante : « comment l'utilisateur final va-t-il formuler sa requête ? » (Bernard 1994, 40). Si la terminologie pouvait répondre à cette question avant l'apparition des folksonomies et l'utilisation des requêtes en langage naturel lorsque les systèmes d'information n'étaient basés que sur des vocabulaires contrôlés, elle a de plus en plus besoin de la lexicographie. L'apparition du numérique et, à la fin des années 90, du web sémantique a permis de réévaluer ces approches dans le cadre de plusieurs disciplines complémentaires : la linguistique, la terminologie, la documentation et l'informatique. Les travaux en organisation des connaissances portent un nouveau regard sur la linguistique et le lexique dont les unités sont utilisées comme mots-clés de recherche. Dans le même temps, en 2006, Hudon place le thésaurus au carrefour de la terminologie et des SIC (Hudon 2006). Dans ce cadre, le champ de l'OC est une méta-science du domaine des SIC et de la terminologie.

Le thésaurus en documentologie, est d'après le TLFi un « langage documentaire fondé sur une structuration hiérarchisée d'un ou plusieurs domaines de la connaissance et dans lequel les notions sont représentées par des termes d'une ou plusieurs langues naturelles et les relations entre notions par des signes conventionnels ». Ainsi, il existe des thésaurus de linguistique, de médecine ou de l'armement. Le thésaurus peut être considéré comme une taxinomie avec plus d'information à propos de chaque

[47] Dans cet article, le terme « organisation des connaissances » n'est pas mentionné. Dans ce champ, nous distinguons le classement d'une part (les systèmes thésaurus, classifications) et la représentation d'autre part (terminologie, lexique, schéma, dessin, image…).

concept et incluant des termes alternatifs et des termes préférés. Un thésaurus peut donc contenir des liens entre concepts associés. Par exemple, dans le thésaurus de l'UNESCO, le descripteur *Patrimoine culturel* est associé au descripteur *Paysage culturel*, c'est-à-dire que ces deux concepts sont reliés sémantiquement et conceptuellement, comme nous pouvons le voir en chapitre 4 dans la définition du patrimoine minier. Les thésaurus sont utilisés par plusieurs acteurs : les bibliothèques et centres de documentation pour indexer leurs fonds documentaires, mais aussi par la plupart des sites de commerce, avec des taxinomies, pour aider les utilisateurs dans leur recherche.

La construction de thésaurus en bibliothéconomie et en technologies de l'information est étudiée en sciences de l'information. L'objectif est de créer des vocabulaires contrôlés à des fins d'indexation ou de balisage et d'utilisation dans les systèmes de recherche d'informations. Un thésaurus est constitué de termes et de relations sémantiques. Les principales relations sont l'équivalence, la hiérarchie et l'association. Lors de l'indexation, les termes tirés d'un thésaurus sont généralement considérés comme indépendants les uns des autres, mais il est également possible de combiner des termes afin de créer des concepts plus complexes par coordination.

En ce qui concerne les relations, la relation hiérarchique générale notée TG pour terme générique et TS pour terme spécifique est représentée suivant trois axes (ISO 25964-1) avec les abréviations TSG pour le genre, TSP pour la partie, TSI pour l'instance mais ces notations ne sont pas obligatoires :

1. Espèce / Genre (relation hiérarchique générique)

Exemple : « poney TSG cheval »

2. Tout / Partie (relation hiérarchique partitive)

Exemple : « bois de mine TSP soutènement »

3. Instance (relation hiérarchique d'instance)

Exemple : « 9-9bis TSI fosse »

L'individu est une instance de concept, autrement dit, c'est l'élément décrit par le concept. Par exemple, l'individu *Amélie* est une occurrence du concept Personne.

Si l'on spécifie la relation partitive (2) avec comme « tout » un ensemble, un objet, un processus, une région, nous pouvons avoir :

Partie - Ensemble,

Membre - Ensemble

Pièce - Ensemble

Matériau - Objet

Phase - Processus

Lieu - Région...

Si l'on spécifie la relation associative, nous pouvons avoir des relations de cause à effet (par exemple grisou – explosion), de producteur et produit (par exemple abatteur – charbon), d'activité et acteur (par exemple médiation culturelle – musée), d'objet et d'emplacement (soutènement – galerie) d'objet et d'activité (pic – abattage), d'outil et sa fonction (barre à mine – forage) etc. Cette finesse sémantique n'est cependant pas notée dans le thésaurus final, on ne note que la relation d'association.

Dans notre travail, la construction d'une terminologie et d'un thésaurus provient d'un même objectif général : la préservation et la valorisation d'un patrimoine en danger. Nous souhaitons mettre en valeur la complémentarité méthodologique d'une approche lexico-sémantique et conceptuelle (L'Homme 2004) lors de l'analyse de ressources documentaires hétérogènes du domaine minier. Ainsi, pour nous permettre de les évaluer sur un même plan, nous les analysons comme systèmes d'organisation des connaissances car ils exercent tous deux ce rôle, nécessaire à l'analyse scientifique prévue dans le cadre du projet.

2.2.2.4 Répertoire de vedettes matières (RVM)

Les listes de vedettes matières (ou répertoire) sont un type de liste d'autorités, qui, elles sont apparues au début du XXIème siècle. Les listes d'autorités sont des listes de termes normalisés, alphabétiques, dont les termes sont appelés vedettes matières. Ces termes ont un caractère obligatoire, et ne peuvent comporter qu'un seul mot ou une seule expression et sont généralement complétés par d'autres éléments (Larouzière 2013). Un exemple bien connu de répertoire de vedettes matières est le langage RAMEAU utilisé par les catalogues de bibliothèques universitaires. Lors de notre recherche, nous avons pu remarquer que la recherche par sujet dans le catalogue du SUDOC[48] n'était presque jamais utilisée par les doctorants et étudiants. La nature encyclopédique du langage d'indexation matière RAMEAU lui permet d'avoir une couverture globale et une granularité variable, elle fait donc souvent l'objet d'évaluation ou de critique par les spécialistes. Boutin, Liu, Goria et *al.* ont par exemple évalué les sciences de l'information et de la communication sous le prisme de ce langage d'indexation lors du colloque international franco-tunisien de la SFSIC/ISD/IPSI en 2008 (Boutin et *al.* 2008). Ils se demandaient si RAMEAU offrait le reflet de la réalité ou une interprétation du monde et en retirent qu'il est loin de favoriser une lecture des différentes sensibilités existant dans le domaine des SIC en France et mettent en avant la nécessité de construire un thésaurus au plus proche des différentes sensibilités de cette discipline. Plus récemment, la transition de l'indexation matière de RAMEAU à l'indexation automatique est questionnée dans (Cavalié 2019). Dans nos travaux, nous utilisons ce référentiel pour identifier les vedettes-matières propres au domaine minier et retrouver la documentation associée dans différents catalogues de bibliothèques l'utilisant.

[48] http://www.sudoc.abes.fr/

2.2.2.5 Autres langages documentaires

Le contenu sémantique d'un document peut s'exprimer à l'aide de mots-clés, c'est-à-dire à l'aide de termes. Une indexation par mot-clé peut impliquer des ambiguïtés lors de l'indexation mais aussi lors de la requête par l'utilisateur, un même terme pouvant être employé dans plusieurs domaines différents. Par exemple, le terme *voile* est répertorié dans 29 domaines[49] d'après le GDT. Il est généralement choisi dans le titre ou le texte d'un document et constitue un point d'accès à celui-ci. « Il est à distinguer d'un descripteur, qui est un terme normalisé dans un thésaurus » (Boulogne et Institut national des techniques de la documentation 2004, 168).

Le terme folksonomie est un néologisme, créé par l'architecte de l'information Thomas Vander Wal, qui fusionne l'anglais *folk*, les gens, et l'anglais *taxonomy*, classification hiérarchisée. Au lieu d'être réalisée par des professionnels et des experts de la documentation, la classification est collaborativement construite par les utilisateurs du web. Ceux-ci sont donc placés au centre du processus d'organisation des connaissances. Le langage utilisé pour indexer est libre, ce sont des mots-clés, des *tags* (étiquettes) ou des méta *tags*. Avec le web 2.0 et ses technologies, une nouvelle forme d'indexation apparaît, l'indexation sociale (Francis et Quesnel 2007). Au niveau de la représentation de ces connaissances, les auteurs mentionnent qu'un des « outils les plus novateurs dans la présentation des mots clés est celui du « nuage de mots-clés » (de l'anglais *tag cloud*). Il consiste à afficher les *n* mots clés les plus populaires, triés par ordre alphabétique. Il donne à chacun un style qui dépend de la fréquence de son utilisation pour indexer les documents » (Francis et Quesnel 2007, 63). Des outils comme Voyant Tools permettent de réaliser des nuages de mots-clés afin d'analyser des textes digitaux (figure 13).

Figure 13 : Nuage de mots-clés réalisé grâce à l'outil Voyant-Tool à partir de l'article de (Francis et Quesnel 2007)

Zacklad regroupe les folksonomies avec les annuaires de ressources internet collaboratifs (liste classée de sites Web réalisée par les utilisateurs même) parce qu'ils « s'appuient sur des applications qui mettent en synergie une communauté d'indexeurs coopérant à travers le web dans une logique proche de celle du web socio sémantique » (Zacklad 2007, 9). L'auteur mentionne que le succès des folksonomies est dû au « faible effort cognitif requis par leur utilisation » ainsi qu'à la « fonction de régulation offerte par

[49] Pour le terme « voile », le GDT liste les domaines suivants : aéronautique, brasserie, céramique, cinéma, eau, éclairage, électroacoustique, énergie, géologie, habillement, imprimerie, industrie automobile, industrie de la construction, loisir, marine, médecine, métallurgie, œnologie, peinture, photographie, physique, religion, sport, technique, télécommunication, textile, transport, travail des métaux, verre.

la mise en visibilité des mots-clefs déposés par l'ensemble des utilisateurs qui permet d'avoir un effet de feed-back rapide sur leur popularité et leur degré de couverture (Ertzscheid et Gallezot 2006) » (Zacklad 2007, 10).

2.3 SOC et Web sémantique

En 2007, Maniez insiste sur le fait qu'il paraisse « difficile actuellement de trouver une définition commune des KOS qui convienne à la fois aux spécialistes du web sémantique et aux documentalistes » (Maniez 2007, 16). Pour lui, l'ambition du web sémantique semble être de participer à la définition de concepts unificateurs ou de systèmes d'harmonisation pour comprendre les différentes approches utilisées pour concevoir les outils traditionnels et ceux de l'intelligence artificielle.

Un exemple d'application du web sémantique est l'immense base de données DBPEDIA développée en français et en anglais dont le but est de maintenir des données structurées extraites de différents chapitres de Wikipédia. DBPEDIA sert aux détenteurs de données et aux développeurs. Plus particulièrement, une des utilisations envisagées est « la recherche de descripteurs partagés à l'échelle du Web pour les reprendre dans d'autres référentiels (classifications bibliothéconomiques, thésaurus, etc.) »[50].

2.3.1 Interopérabilité des SOC

L'interopérabilité peut être technique ou sémantique. En informatique (traitement des données) et en télécommunication, elle désigne la « capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner accès à leurs ressources de façon réciproque » (GDT). Au niveau sémantique, l'interopérabilité est « associée à un mode de description de l'information contenue dans une base de données (cette description forme les métadonnées) » (GDT).

Pour le domaine qui nous intéresse, les bibliothèques, archives et musées possèdent leurs propres langages, utilisés pour indexer des documents hétérogènes, leurs propres plateformes et logiciels organisant leurs données et sont ainsi un peu isolées de la production et l'échange des données sur Internet et sur le Web. Elles ne profitent donc ni de la visibilité en ligne, ni de l'apport de l'échange des données (par exemple bibliographiques ou de contenu) entre elles. Comme l'explique Cordereix concernant les bibliothèques :

> « Pourquoi ce besoin d'interopérabilité ? Pour agréger des données, pour constituer des catalogues collectifs, des portails de ressources qui vont être exposées sur Internet, et ceci avec des données qui peuvent être hétérogènes, ce que ne permet pas – ou peu –la rigidité des formats MARC par exemple. Mais le besoin d'interopérabilité vient aussi et surtout de la nécessité de rompre l'isolement des bibliothèques face aux autres acteurs de l'Internet. » (Cordereix 2016, 105)

[50] http://fr.dbpedia.org/

D’après Hubain, Hooland, et Verborgh, l’interopérabilité est un enjeu « vital » et ils rappellent dans ce cadre le rôle majeur qu’ont à jouer « les taxonomies, thésaurus, plans de classements et autres ressources terminologiques, créateurs de liens par excellence », donc les SOC (Hubain, Van Hooland, et Verborgh 2016, 78). Favier et Hadi traitent de la question de l’interopérabilité de ceux-ci et se demandent si celle-ci ne dessine pas une nouvelle conception de l’universalité du savoir (Favier et Hadi 2013). Nous ajoutons que pour ce faire, les ontologies et l’utilisation de modèles ontologiques standards comme liens formels et sémantiques entre toutes ces ressources, jouent un rôle crucial aujourd’hui. Dans ce cadre, des problématiques sur l’interopérabilité concernant des données hétérogènes liées (*linked data*) sont posées pour la sauvegarde d’un patrimoine par Kergosien, Jacquemin, Severo et Chaudiron avec l’utilisation du modèle du CIDOC CRM (Kergosien et *al.* 2015b). Nous verrons que notre méthodologie s’inspire fortement de l’approche utilisée dans ce projet. Nous discuterons et évaluerons pour notre part de l’interopérabilité sémantique entre les SOC que nous avons créés au niveau des langages de représentation en chapitre 10.

2.3.2 SOC culturels (cultural KOS) et web sémantique

Dans le domaine culturel, certains grands projets comme UNESCO *Memory of the World Program* (1992) ou Europeana (2008) ont encouragé la numérisation massive des documents historiques, des photographies, des enregistrements sonores et leur catalogage dans des dépôts numériques (Díaz-Corona et *al.* 2019). Les chercheurs ont donc pu disposer d’importantes bases de données culturelles et en explorer les diverses possibilités comme celle de la représentation dans des domaines comme la modélisation des connaissances avec le web sémantique et la réalité virtuelle comme dans des patrimoines industriels (Laubé 2016).

Les dénominations *cultural KOS* ou *cultural knowledge organization systems*, ou encore SOC culturel(s), systèmes d’organisation des connaissances culturelles ne sont pas très répandues. Pourtant elles permettraient de qualifier des types de SOC spécifiques à l’organisation des connaissances culturelles et notamment ceux orientés patrimoine culturel qui font l’objet d’un certain nombre de travaux actuels. Par exemple, Díaz-Corona, Lacasta, Latre et *al.* décrivent le processus dont ils se servent pour découvrir et décrire les SOC utilisés dans le domaine du patrimoine culturel (Díaz-Corona et *al.* 2019). Dans ce cadre, les auteurs soulèvent un certain nombre de problèmes dont le celui d’exhaustivité, le manque de définition des concepts, le problème des concepts orphelins, le manque d’exhaustivité dans les relations hiérarchiques, des problèmes de précision et d’exactitude sémantique. D’après leur étude, il en ressort que les critères de cohérence sont moins pertinents que les critères d’exhaustivité et que ceux-ci sont similaires aux critères de précision.

CHAPITRE 3. Approches ontologiques de l'organisation des connaissances

Ce chapitre présente les approches ontologiques de l'organisation des connaissances en replaçant tout d'abord celles-ci dans leurs différents domaines d'étude (3.1). Le chapitre traite ensuite des différents types d'ontologies qui peuvent être catégorisées selon plusieurs critères (3.2) puis revient sur les travaux les étudiant ou les utilisant (3.3) et sur leur rôle dans ce cadre (3.4). Enfin, nous présentons les éléments caractéristiques d'un modèle ontologique (3.5) ainsi que les différentes approches de conception d'un type particulier d'ontologie : l'ontologie de domaine (3.6).

Les ontologies sont très ancienne. Au fil du temps, elle ont souhaité répondre à : qu'est-ce qui existe ? quelles catégories conceptuelles peuvent décrire ce qui existe ? est-ce que ces catégories conceptuelles sont des connaissances ? quels liens peuvent être effectués entre elles pour pouvoir les retenir ? comment se représenter celles-ci ? comment les représenter sur un support ? comment les enregistrer sur un support pérenne ? comment transmettre les connaissances grâce à un code qui pourra être déchiffré par tous ? comment transmettre les connaissances grâce à un code qui pourra être déchiffré par tous dans mille ans ou plus ? De par leur caractère méta, elles sont traitées dans des domaines qui étudient la connaissance générale soit en questionnant son existence en philosophie, soit en essayant de trouver un code au plus proche de la réalité étudiée et correspondant aux besoins des utilisateurs de l'époque en question (linguistique, gestion et organisation de la connaissance, informatique) et si possible en exploitant ces connaissances pour en produire de nouvelles (en intelligence artificielle). Chacune de ces approches peut apporter des points de vue intéressants sur l'étude et la représentation d'un domaine que nous développons dans ce chapitre.

Très récemment, Machado revient sur le souci d'ambiguïté sémantique du terme « ontologie » qui est considéré à la fois comme un type de SOC et comme un processus de catégorisation dans les différents champs disciplinaires utilisant les ontologies (Machado 2021). D'une part, l'auteur note que l'ontologie est une évolution en termes de complexité des systèmes conceptuels traditionnels et d'autre part, comme un système qui organise les connaissances ontologiques plutôt qu'épistémologiques. L'auteur note également que l'analyse ontologique se concentre sur l'élément à modéliser et non sur les intentions qui motivent la construction du système. Dans cet article, l'évolution des deux sens est chronologiquement retracé dans la littérature (Machado 2021, 45), du passage de la philosophie (où l'ontologie désigne le processus d'étude ou d'analyse) et où *ontology* désigne la première philosophie d'Aristote qui consistait à systématiser et catégoriser toutes les entités (*entities*) qui existent dans le monde vers l'aire des systèmes d'information grâce à celle de l'intelligence artificielle. Alors que dans le champ des SOC le terme « *ontology* » désigne un produit, un artefact. La première définition d'*ontology* dans le cadre des systèmes d'informations en 1991 est la suivante : « *The ontology of a system consists of its vocabulary and a set of constraints on the way terms can be combined to model a domain* » (Neches et *al.* 1991, 40). Machado analyse les ontologies de type artefact selon trois critères, que celui-ci nomme « facette » :

- L'objectif général de l'ontologie qui oppose l'ontologie de référence à l'ontologie d'application
- Le langage formel utilisé qui oppose la logique de description à la sémantique à partir du RDF (*Resource Description Framework*)
- L'approche de modulation appliquée qui oppose les modèles ontologiques aux modèles conceptuels

Nous notons ici une définition bien connue et souvent reprise de l'ontologie : « *An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For AI systems, what "exists" is that which can be represented* » (Gruber 1995, 908).

3.1 Domaines d'étude des ontologies

À l'origine, le concept d'ontologie provient de la philosophie. Les penseurs grecs comme Parménide et Platon essaient à travers ce concept de répondre à la question générale « Qu'est-ce qui existe? ». Au début du 17ème siècle, la notion d'ontologie est utilisée en philosophie aristotélicienne avec pour objet d'étude les propriétés générales de l'être comme l'existence, la possibilité, la durée, le destin. Cette notion trouve son origine dans les idées de la métaphysique d'Aristote. Dans son ontologie, Aristote a identifié dix catégories décrivant le monde : Substance, Qualité, Quantité, Relation, Où, Quand, Position, Avoir, Action, Passion. Aujourd'hui, la recherche ontologique en philosophie a un fort penchant théorique dans le but de développer des modèles formels des catégories fondatrices et de la logique derrière tout (Sowa 2000). Notre travail ne s'insère pas dans ce champ philosophique en cela qu'il ne questionne pas la réalité des objets.

Dans le domaine de l'informatique, la notion d'ontologie a commencé à être pertinente dans les années quatre-vingt-dix. Le terme ontologie est alors utilisé pour désigner une structure de données formelle qui peut être traitée par des algorithmes. L'article de Uschold et Gruninger (1996) est pour Chaumier (2007) un texte fondateur sur les ontologies. La définition donnée dans cet article est la suivante :

> « Il s'agit du terme utilisé se référant à la compréhension partagée [*a shared understanding*] 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 » (Chaumier 2007, 81).

À partir de cette définition, plusieurs descriptions d'ontologies ont été faites donnant lieu à la définition suivante, plus communément utilisée de nos jours : Une ontologie est une spécification formelle, explicite d'une conceptualisation partagée (Studer, Benjamins, et Fensel 1998, 25)[51]. Un ouvrage de référence, (Studer 2009), définit ce qu'est l'ontologie sous tous ses angles. Dans celui-ci, Guarino, Oberle, et Staab reviennent sur la conceptualisation, la spécification formelle et explicite et sur

[51] Traduit par l'auteure à partir de : « *An ontology is a formal, explicit specification of a shared conceptualization* « (Studer, Benjamins, et Fensel 1998, 25)

l'importance d'une conceptualisation dite « partagée » (Guarino, Oberle, et Staab 2009). Ainsi, l'aspect formel signifie que l'ontologie a une syntaxe et une sémantique bien définies. Une spécification explicite signifie qu'elle peut être représentée et traitée de manière algorithmique. L'aspect de partage implique qu'elle fait l'objet d'un accord au sein d'une communauté et facilite la communication entre les agents qui la composent. Enfin, l'aspect conceptuel présente un modèle du monde réel, un point de vue simplifié sur le monde que nous souhaitons représenter dans un but précis[52]. Les outils et méthodes de ce domaine suivent et permettent de maintenir des normes et de partager des travaux sur le web en utilisant des langages formels. Gnoli (2020) citant Sowa (2000) définit enfin le but des ontologies comme : « *They are aimed at representing and exploiting knowledge in formalized ways, as part of applications in computer science, including artificial intelligence* » (Gnoli 2020, 84)[53]. Le terme d'ontologie a en effet été repris dans les sciences cognitives et en intelligence artificielle. Dans ce cadre, la principale question n'est plus « qu'est-ce que la nature de l'être ? » d'après Borst (1997, 11) mais « sur quoi un système d'IA doit raisonner pour être capable d'effectuer une tâche utile ? »[54].

Les ontologies en tant que « *schematic descriptions of the contents of domain knowledge* » (Van Heijst, Schreiber, et Wielinga 1997, 183) sont aussi utilisées en Ingénierie des connaissances où elles peuvent être exploitées pour organiser le processus d'acquisition des connaissances (*Knowledge Acquisition*) comme le montrent (van Heijst, Schreiber, et Wielinga (1997). Les auteurs s'appuient sur la définition de Gruber (1995, 908) : « *An ontology is an explicit specification of a conceptualization* » et sur le fait qu'une ontologie fournit une spécification d'une conceptualisation de notions génériques comme le temps, l'espace ou d'un domaine d'application comme la gestion des connaissances ou les sciences du vivant. D'après Prévot, Huang, Calzolari et *al.*, la conceptualisation est la connaissance informelle et pertinente que l'on peut extraire et généraliser de l'expérience, l'observation ou l'introspection. La spécification est l'encodage de cette connaissance dans un langage de représentation (Prévot et *al.* 2010).

En linguistique, l'ontologie est proche des dictionnaires sémantiques où les mots sont organisés en fonction de leur sens. Toutefois, comme le précise Hirst (2009) les lexiques (*lexicons*) ne sont pas réellement des ontologies. Les premiers sont des objets linguistiques tandis que les deuxièmes n'y sont pas. Le lexique peut néanmoins aider à la construction des ontologies. Prévot, Huang, Calzolari et *al.* différencient les ontologies formelles et les ontologies linguistiques. Dans leur ouvrage, ils considèrent que *computational lexicons* (lexiques informatiques), *lexical resources* (ressources lexicales) or *relational lexicons* (lexiques relationnels) sont synonymes même si d'un point de vue terminologique, ils n'y sont pas. Ils nomment *ontolex resource* :

[52] « *A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.* « (Guarino, Oberle, et Staab 2009)
[53] *Formalized ways* signifie langage formel.
[54] « *For AI the main question is not what the nature of being is, but what an AI system has to reason about to be able to perform a useful task* » (Borst 1997, 11)

> « *Finally, by ontolex resource we mean any attempt to combine ontological principles with lexical information in order to synergize the information content and to create a more powerful and versatile resource.* » (Prévot et *al.* 2010, 5)

OntoLex est donc l'abréviation qui désigne le résultat de la combinaison des ontologies et des lexiques informatiques, aujourd'hui considérée comme l'objectif clé pour la nouvelle génération de systèmes de connaissance. Les ontologies comme des spécifications dépendant du langage de conceptualisation, sont en effet la base de communication, le pont à partir duquel la compréhension commune est établie.

Un exemple de ressource OntoLex est Wordnet. Celui-ci est considéré comme le vocabulaire/ontologie linguistique le plus largement utilisé sur le web sémantique. Il organise les significations des mots en synonymes cognitifs, qui font référence à des concepts. Il existe d'autres versions de Wordnet pour d'autres langues, mais la version anglaise est la plus complète et contient 117 000 concepts. Alors que la lexicographie étudie les mots et leur signification et la façon de les écrire, la terminologie étudie les termes techniques et leur utilisation, les concepts et les systèmes conceptuels et leurs domaines, principalement à des fins de désambiguïsation dans des contextes professionnels. L'analyse terminologique définit les concepts grâce à une méthodologie normalisée (ISO, International Standards Organization) (Suonuuti 2001). Les synonymes, la polysémie et l'homonymie sont des sources d'ambiguïté dans la détermination des référents d'un mot, tandis que les relations d'équivalence, génériques, partitives et associatives sont les principales utilisées dans un système de concepts et sont représentées sous forme de graphe pour analyser et définir les concepts.

La combinaison termes/concepts est surtout étudiée et éclairée dans les ontoterminologies (Roche 2005, 2008 ; Roche et Papadopoulou 2020). Nous l'avons vu, ce type d'ontologie participe à la définition d'un nouveau paradigme pour la terminologie (Roche et *al.* 2009). Roche part du principe que les experts s'accordent sur la conceptualisation d'un domaine quand il est défini dans un langage formel (Roche et *al.* 2009). Les ontoterminologies séparent les unités linguistiques des concepts et les relient par des relations d'hyponymie et de synonymie. La conceptualisation est donc claire : un concept n'est pas un terme. Au niveau de la construction de ce type d'ontologie, Christophe Roche montre la limite des « ontologies textuelles », c'est-à-dire construites à partir de textes, qui ne sont pas réellement des ontologies en cela que par nature, elles comportent normalement un caractère extralinguistique, qui n'est pas porté par les unités textuelles (il nomme l'exemple des déictiques).

Nos travaux se situent quant à eux à un niveau documentaire, ce qui nous permet de faire référence aux « ontologies documentaires ». Elles permettent, grâce à un métalangage, de dépasser l'hétérogénéité des documents (notamment au niveau de la langue et au niveau des formats) et l'hétérogénéité de leur SOC afin de les rendre interopérables, ceci dans le but de permettre aux communautés d'accéder au contenu des documents et éventuellement d'échanger des données.

3.2 Types d'ontologies

Comme objet de la connaissance, les ontologies peuvent être à leur tour catégorisées. Ceci démontre bien le caractère infini de la connaissance c'est-à-dire que même les objets ayant pour but de conceptualiser peuvent à leur tour être conceptualisés et ainsi de suite. Plusieurs critères ont été identifiés pour définir celles-ci comme :

- l'objet de la conceptualisation,
- leur granularité,
- leur niveau de formalisation dans (Drame 2014),
- le besoin des utilisations (Fadili et Chakiri 2014),
- les éléments d'un modèle ontologique,
- leur rôle,
- les approches de leur conception.

Nous les définissons ci-dessous selon leur granularité conceptuelle.

3.2.1 Ontologies de haut niveau

Les ontologies de haut niveau, appelées *top-level ontology* ou *upper level ontology* en anglais, ou encore ontologie abstraites, de représentation des connaissances ou ontologies génériques représentent comme leur nom l'indique, des connaissances générales sur le monde grâce à des concepts génériques appartenant « à l'univers de modélisation commun aux ontologies des niveaux inférieurs » (Fadili et Chakiri 2014, 154) tels que l'espace, le temps, la matière, les objets, les événements, les actions. En tant que modèle générique, les ontologies de haut niveau ont un but d'interopérabilité entre toutes ontologies plus spécifiques. Nous présentons ci-dessous quelques ontologies de haut niveau :

- L'ontologie SUMO *Suggested Upper Merged Ontology* est l'exemple de plus haut niveau de ce type d'ontologie. Elle est développée dans le cadre du projet IEEE SUO (*Standard Upper Ontology*). Elle doit permettre l'interopérabilité sémantique entre tous les systèmes d'information. SUMO comporte plusieurs centaines de concepts et de relations généraux.
- L'ontologie YAGO (Suchanek, Kasneci, et Weikum 2008) a une couverture et une précision élevées et est dérivée de Wikipédia et de Wordnet.
- L'ontologie DOLCE, *Descriptive Ontology for Linguistic and Cognitive Engineering* (Borgo et *al.* 2002 ; Gangemi et *al.* 2003) développée dans les années 2000 en est également un exemple. Comme nous l'expliquent Gangemi, Guarino, Masolo et *al.*, cette ontologie a une orientation cognitive c'est-à-dire qu'elle vise à saisir les catégories ontologiques qui sous-tendent le langage

naturel et le sens commun humain (Gangemi et *al.* 2003)[55]. Les auteurs nous expliquent que cette ontologie fait par exemple la distinction entre des catégories ontologiques de très haut niveau telles que les *endurant* et les *perdurant*. La différence entre ces deux est faite par leur différent comportement dans le temps : les endurants ont besoin d'une parentalité indexée sur le temps alors que les perdurants non. Les personnes seront considérées comme des endurants tandis qu'une discussion comme un perdurant.

- Enfin, l'ontologie du CIDOC CRM utilisée dans nos travaux est également considérée comme une ontologie de haut-niveau, qui, de par sa conceptualisation très générique, peut être utilisée dans de nombreux domaines. Nous la plaçons en revanche entre les ontologies de haut niveau et les ontologies de domaine, car elle se concentre tout de même sur un domaine particulier qui est le patrimoine culturel (même si celui-ci est parfois considéré comme étant plus large qu'un domaine). La figure 14 est une représentation hiérarchique d'une partie du CIDOC CRM. La classe générique *E1 CRM Entity* regroupe les entités temporelles (*E2*), les entités persistantes (*E77*), les durées (*E52*), les lieux (*E53*), les dimensions (*E54*) et leurs classes spécifiques.

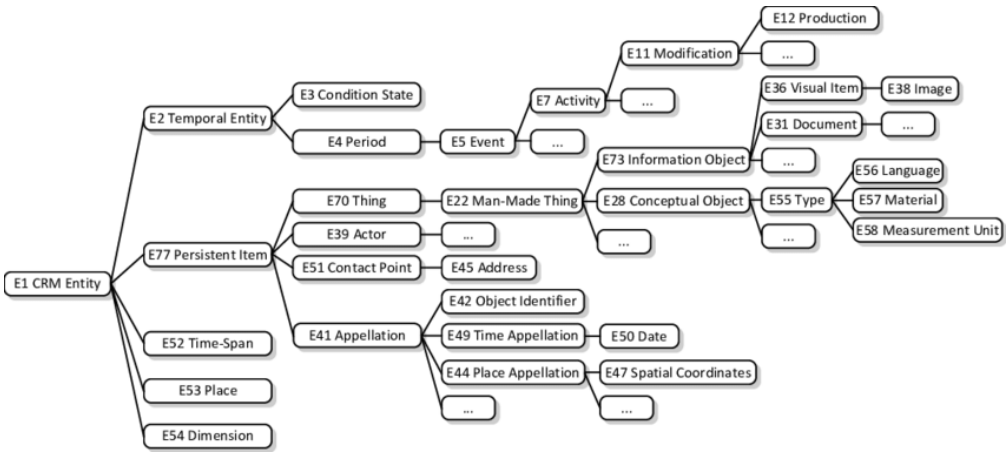


Figure 14 : "*Small section of the CIDOC CRM class hierarchy*" (version 4.2.1) par Bernhard Haslhofer

3.2.2 Ontologies de domaine ou de spécialité

Picard oppose les ontologies de domaine aux ontologies de noyau ou *core-ontologies* ou *top-domain ontologies*. D'après lui :

> « [e]lles s'intéressent aux termes servant à désigner les concepts dans les documents produits par les professionnels du domaine lors de leur activité. Chez ces dernières, il y a une prédominance de la dimension linguistique et terminologique. Par exemple, l'ontologie de domaine peut être plus ou moins riche de termes synonymes pour dénoter les concepts. Par ailleurs, la spécification peut être plus ou moins raffinée selon l'application à laquelle l'ontologie est destinée. Ainsi, pour un domaine donné, il peut y avoir plusieurs ontologies de domaine » (Picard 2014, 24).

[55] Traduit par l'auteure à partir de : « […] *DOLCE has a clear cognitive orientation, in the sense that it aims at capturing the ontological categories underlying natural language and human commonsense*. » (Gangemi et *al.* 2003, 17)

C’est ce type d’ontologie que nous souhaitons concevoir dans nos travaux.

- Un exemple de travail ayant eu pour objectif de concevoir une ontologie de domaine est celui présenté par Hernandez (2005). Son travail s’applique à modéliser le contexte en RI à travers deux ontologies dites « légères », c’est-à-dire d’un niveau formel moindre. Une des ontologies concerne « le thème abordé dans la collection sur laquelle l’utilisateur effectue sa recherche », l’autre est une ontologie de tâche et modélise « la tâche de recherche que l’utilisateur choisit pour combler son besoin ». Dans ce cadre, des questions de réutilisabilité de l’ontologie et de visualisation ont été traitées.

- Un autre exemple est l’ontologie GENE[56] qui vise à standardiser la formalisation des gènes. Les termes sont définis et sont reliés à des espaces de noms des termes (*term namespace*) qui définissent le processus biologique (*biological_process*), le composant cellulaire (*cellular_component*), la fonction moléculaire (*molecular_function*).

- OWL-Time[57] est un exemple d’ontologie de domaine temporel, décrivant le contenu temporel des pages Web et les propriétés temporelles des services Web. Un exemple de classes : *Day of week* (*Sunday, Monday*…) et de propriété : *has duration*.

- DEMO Discret Event Modeling Ontology est une ontologie de domaine de simulation à événements discrets. Les modèles dans les ontologies se concentrent sur l’évolution de l'état dans le temps[58].

Les ontologies de tâche ou d’application sont plus restreintes que les ontologies de domaine et s’appliquent à une tâche précise.

- Un exemple de ce type d’ontologie est RudOnto (Kolonja et *al.* 2016), qui est en fait un systèmes d’ontologies développé pour l’ingénierie minière et s’applique dans les domaines de l’équipement minier et de la sécurité dans les mines. Les ontologies RudOnto offrent la possibilité de générer des ressources terminologiques autonomes ou des ontologies dans des sous-domaines spécifiques, tels que les équipements miniers, la sécurité des mines et la géostatistique. Dans l’article, les auteurs montrent les résultats obtenus grâce à ce système d’ontologies sur les parties du corps blessées par heures de travail, les causes ou taux de gravité des blessures dans les mines. Cette ressource est développée à l’Université de Belgrade, à la *Faculty of Mining and Geology* et se veut être la ressource de référence pour la terminologie minière en Serbe dans un format digital. La version courante de RudOnto (en 2012) contenait environ 7000 concepts génériques utilisés dans la pratique de l’industrie minière, environ 1600 équivalents anglais, et quelques traductions vers d’autres langues (Obradović et *al.* 2012). Dans

[56] http://geneontology.org/
[57] https://www.w3.org/TR/owl-time/
[58] http://cobweb.cs.uga.edu/~jam/jsim/DeMO/

le domaine minier, un autre travail de l'Université de Stanford a étudié l'ontologie des catastrophes des eaux minières, conçue avec le logiciel Protégé (Sun, Yue, et Guan 2009).

L'objectif de nos travaux étant de représenter les connaissances d'un domaine donné, les ontologies que nous considérons dans la suite du mémoire sont des ontologies de domaine. « Ces ontologies permettent de normaliser la connaissance manipulée par le système par rapport à la connaissance qui lui est utile » (Hernandez 2005, 31).

3.3 Travaux actuels sur les ontologies dans le champ de l'OC

Récemment, une réelle volonté de définir les ontologies comme SOC dans le champ de l'OC a vu le jour et ceci internationalement grâce à l'ISKO. L'article d'encyclopédie publié en 2020 par Biagetti met particulièrement en valeur le rôle qu'ont les ontologies dans le web sémantique tel qu'il a été imaginé par Tim Berners Lee, c'est-à-dire un réseau de documents connecté à un réseau d'éléments de connaissance (Biagetti 2020). À ce propos, Biagetti écrit que les ontologies représentent le noyau dur de la transformation, permettant aux machines de requêter, de raisonner et de manipuler les sens et la connaissance.[59] Dans ce cadre, et depuis les années 2000 en France quelques thèses de doctorat s'intéressent aux ontologies du point de vue des sciences de l'information et de la communication et plus particulièrement :

-pour la recherche d'information (RI) avec une modélisation s'adaptant aux besoins et attentes des utilisateurs en termes de contenu, de navigation et de présentation des informations (Verlaet 2008),

-pour la description d'un genre littéraire, construction d'une théorie sémiotique, modélisation de représentation de la fiction et élaboration d'un système d'organisation de connaissances info-littéraires grâce à des analyses de corpus de textes (Archambault 2012),

-pour l'identification des ressources sur le Web, avec l'utilisation des technologies sémantiques dans le domaine de l'apprentissage et l'enseignement et la construction de plusieurs ontologies adaptées aux entités participant à l'apprentissage (Szilagyi 2014),

-pour la médiation culturelle, utilisation des technologies du Web sémantique pour la réalisation d'une ontologie étendue en base de connaissances avec instances issues d'un corpus dans le domaine culturel institutionnel des jeux pervasifs (Dufort 2016),

-pour l'interopérabilité entre différents systèmes et équipements dans le domaine de l'éducation avec une réflexion sur la mise en place des systèmes de gestion de corpus lexicaux homogènes, multilingues, partageables et interopérables et le caractère indispensable de la normalisation pour l'éducation et ses retombées bénéfiques en termes d'accessibilité, d'interopérabilité et d'adaptabilité (Boudhir 2016),

[59] Traduit par l'auteure de : « *Ontologies represent the "core" of the transformation, allowing machines to query, reason and manipulate meanings and knowledge.* » (Biagetti 2020, s.p.)

- pour l'interopérabilité et la réutilisation de données issues d'une recherche sur les qualités attendues d'un instrument de lecture sur écran issues d'entretiens compréhensifs et conception d'une ontologie standardisant quelques-unes de ces données (Tirole 2016),

-pour la représentation et l'organisation des connaissances spécialisées dans le domaine de la médecine avec l'apport de la Terminologie et les liens avec le web sémantique (Carvalho 2018),

-pour réduire les ambiguïtés communicationnelles entre plusieurs acteurs (experts de domaine, futurs experts et professionnels de la langue) dans le domaine de l'industrie du liège avec réalisation d'une micro-ontologie de domaine à partir d'une terminologie et d'un e-dictionnaire (Ramos 2020).

Ceci n'est pas exhaustif mais montre des enjeux forts et récents de l'utilisation des ontologies, qui, combinés, permettent de réaliser un système informationnel de transmission de connaissance complet. Suite à l'analyse des thèses, nous avons retenu des enjeux phares que nous représentons sous la forme d'un diagramme (figure 15). Nous plaçons nos travaux dans la lignée des travaux de Carvalho (2018), de représentation et d'organisation d'un domaine.

Figure 15 : Diagramme représentant les enjeux des ontologies dans les thèses françaises en Information-Communication. Réalisation par l'auteure.

3.4 Rôle des ontologies

Le rôle des ontologies dépend du cadre dans lequel elles sont conçues. Cette section développe quatre rôles qu'ont actuellement les ontologies qui sont la description, le partage et la réutilisation de connaissances, la recherche d'information, la construction de dictionnaires digitaux, puis l'inférence de connaissances.

3.4.1 Ontologies pour la description, le partage, l'échange et la réutilisation de connaissances

Comme le mentionne Messaoudi à propos de l'échange des connaissances d'un domaine :

> « Afin de faciliter l'échange des connaissances d'un domaine, l'idée principale aujourd'hui n'est plus de déplacer les bases de données et de les restructurer pour les unifier, mais d'exploiter d'autres technologies comme le web sémantique qui a pour stratégie d'améliorer la valeur de la connaissance collectée via des tâches rattachées au patrimoine culturel. La meilleure solution pour structurer les connaissances, via des technologies du web, est l'ontologie, qui fournit une compréhension partagée d'un domaine d'intérêt afin de soutenir la communication entre les Humains et les agents informatiques (ordinateur) étant typiquement représentée dans un langage de représentation compréhensible par les ordinateurs » (Messaoudi 2017, 128).

Pour le partage de connaissances, Gruber définit des principes pour la conception d'ontologie. Il propose cinq critères entrant en jeu dans ce processus : la clarté (*clarity*), la cohérence (*coherence*), l'extensibilité (*extendibility*), le biais d'encodage minimal (*minimal encoding bias*) et l'engagement ontologique minimal *(minimal ontological commitment*) (Gruber 1995). En 1997, Borst traite des ontologies d'ingénierie pour le partage de connaissance et pour sa réutilisation (Borst 1997). Il soulève des problématiques telles que la conception d'ontologies qui ne sont pas utilisées pour développer une application réelle, qui ne sont pas réutilisées pour différentes applications dans différents domaines et d'autres qui ne sont que des taxonomies de concepts de domaine et ne parviennent donc pas à capturer les connaissances de méta-niveau et les connaissances de fond tacites. À partir de cela, l'auteur étudie la facilité d'utilisation et de réutilisation des ontologies par la construction et la validation d'une ontologie pour un domaine large et complexe, utilisable pour le développement d'applications, réutilisable dans différentes applications et capturant les connaissances de méta-niveau et les connaissances de fond tacites.

La clarté fait référence à des définitions objectives indépendamment du contexte social et computationnel. La cohérence concerne les inférences qui doivent être cohérentes avec les définitions. L'extensibilité signifie qu'il faut être capable de définir de nouveaux termes pour des utilisations spéciales sur la base du vocabulaire existant, sans avoir à réviser les définitions existantes. Le biais d'encodage minimal signifie que la conceptualisation est spécifiée au niveau de la connaissance et ceci sans dépendre d'un encodage particulier. Enfin l'engagement ontologique minimal doit permettre aux parties engagées dans l'ontologie de se spécialiser et d'instancier selon les besoins en spécifiant la théorie la plus faible et en définissant seulement les termes qui sont essentiels à la communication de connaissances cohérentes avec cette théorie. Il signale que des compromis entre tous ces critères sont évidemment nécessaires, comme dans toute conception.

Du point de vue documentaire et des domaines associés, des ontologies ont été développées pour décrire, partager et réutiliser les connaissances issues soit de métadonnées des documents, soit du contenu issu de ces documents (tableau 5). Dans ce cadre, les ontologies mettent en relation des connaissances sur les œuvres en général. Le modèle du Dublin Core conçu pour enregistrer les métadonnées des documents est le plus simple d'utilisation et le plus souple, mais a rapidement montré ses limites : la catégorie

creator ou créateur ne différencie par exemple pas l'auteur originel et le traducteur de cette même œuvre. Le FRBR, centré sur les notices bibliographiques a été conçu pour spécifier ce type d'ambiguïtés. Elles permettent d'effectuer des inférences grâce à l'utilisation de relations sémantiques (si telle instance est fille de telle autre, alors inversement l'autre instance est père de la première, si telle instance est contenue dans tel document alors ce document contient cette instance, si telle instance est reliée à une autre et cette autre est reliée à une autre, alors il peut être intéressant d'étudier la première au regard de la dernière etc.). Pour ce faire, elles fournissent des langages qui permettent de conceptualiser les objets d'un domaine par leur identification pérenne et permet de les rendre manipulables, voire calculables (ceci avec toutes les questions que cela peut poser)[60].

Tableau 5 : Ontologies documentaires

Modèle	**Centré sur…**
Dublin Core	Le document
FRBR	Les notices bibliographiques
CIDOC CRM	Les objets du patrimoine culturel matériel ou immatériel
FRBRoo	FRBR+CIDOC CRM

Les ontologies peuvent servir dans toutes tâches de description de documents. Depuis les années 2000, des thèses se concentrent sur un type de document particulier : les documents audiovisuels. Ainsi, Bui traite de la structuration sémantique des contenus des documents audiovisuels selon les points de vue de la production (Bui 2003). La thèse de Troncy traite de la formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies pour l'application à la description de documents audiovisuels (Troncy 2004). Malaisé a défini une chaîne de traitement issue de méthodologies et outils linguistiques proposés pour l'aide à la construction d'ontologie à partir de vocabulaire conceptuel destiné à annoter des documents audiovisuels traitant de la petite enfance (Malaisé 2005). Diemert traite de la description sémantique de documents audiovisuels structurés (Diemert 2012).

Dans le domaine du patrimoine, les ontologies peuvent être utiles aux acteurs pour faciliter la gestion de l'information dans ce secteur. Dans ce cadre, elles peuvent aussi être considérées comme des outils de médiation (pédagogique). Notre travail s'appuie sur l'article de Doerr qui décrit comment les ontologies sont et pourraient être utilisées pour améliorer la gestion de l'information dans le secteur du patrimoine culturel (Doerr 2009, 463). Pour l'auteur, le patrimoine culturel est plus qu'un domaine : « *It comprises a broad spectrum of functions about the study and preservation of physical evidence of the past of all sorts of human activities* ». Il insiste sur la distinction à réaliser entre les bibliothèques et les objets de patrimoine culturel, pour lui ces objets « *are a rather mute evidence of past events that acquire relevance from understanding the context of their origin and history* » (Doerr 2009, 464). Ceci montre l'importance à donner au contexte de ces objets et son pouvoir.

Doer distingue ensuite quatre sortes de systèmes d'information du patrimoine culturel basés sur leur fonction : gestion des collections, conservation, recherche, présentation (Doerr 2009, 465) et précise qu'un des challenges majeurs de la gestion de l'information est l'interopérabilité de ces systèmes à

[60] D'après nous, créer une ontologie, c'est définir un concept comme un chiffre, lui appliquer des règles de calcul et lui attribuer un résultat qui sera toujours le même. Cela permet de manipuler les concepts et de les rendre calculables.

travers les fonctions et les disciplines (Doerr 2009, 465). Dans ce but d'interopérabilité et parmi les ontologies existantes, le modèle *International Committee for Documentation of the International Council of Museums (CIDOC) Conceptual Reference Model (CRM)* a été choisi pour le projet MémoMines.

Doerr présente la méthodologie utilisée par une équipe interdisciplinaire pour créer le CIDOC CRM qui est une ontologie de haut-niveau permettant l'intégration des informations sur le patrimoine culturel et leur corrélation avec les informations des bibliothèques et des archives (Doerr 2003). Le CRM CIDOC analyse les conceptualisations communes derrière les structures de données et de métadonnées pour soutenir la transformation, la médiation et la fusion des données. L'argument est que de telles ontologies sont centrées sur les propriétés, contrairement aux systèmes terminologiques et devraient être construites avec des méthodologies différentes. Il est démontré que les arguments ontologiques et épistémologiques sont tout aussi importants pour une conception efficace, en particulier lorsqu'il s'agit de traiter les connaissances du passé dans n'importe quel domaine. Il est supposé que la méthodologie présentée et le niveau supérieur de l'ontologie sont applicables dans un domaine beaucoup plus large, toujours d'après Doerr (2003).

Le CIDOC CRM est largement utilisé à l'international. En Allemagne, par exemple, le projet WissKi *Wissenschaftliche KommunikationsInfrastruktur* que nous traduisons « Infrastructure de communication scientifique » utilise le CIDOC CRM comme modèle commun à tous les systèmes du projet. En Angleterre, le *British Museum*, Musée de l'histoire et de la culture humaine à Londres a créé une interface, *Research Space*, dédiée à la manipulation des fonds patrimoniaux du musée. Aux États-Unis, le projet *ARCHES* qui est une plateforme logicielle *open-source* développée conjointement par le *Getty Conservation Institute* et le *World Monuments Fund* gère les données du patrimoine culturel. Comme nous pouvons enfin le lire sur le site du projet français symogig.org, celui-ci « a développé un modèle générique de stockage des données historiques permettant leur interopérabilité et leur publication sélective. À partir de ce modèle, une plateforme collaborative[61] pour la recherche en histoire a été mise en place, utilisée par plusieurs chercheurs et projets »[62].

Par ailleurs, le modèle est en constante évolution et il a été reconnu comme standard. La version que nous utilisons est la 6.2.1 car c'est une version publiée et disponible en RDF (permettant le traitement automatique). Depuis cette version, des évolutions de type ajout de classe et ajout de propriétés peuvent être notées. Par exemple, *E22 Man-Made Object* devient *E22 Human-Made Object* et toutes les classes

[61] http://symogih.org/?q=documentation
[62] http://symogih.org/?q=type-of-knowledge-unit-classes-tree et http://symogih.org/

commençant par *Man-* deviennent *Human-* à partir de la version 6.2.7. Quatre classes[63] et 19 relations[64] sont ainsi ajoutées.

Le CIDOC CRM possède des extensions (figure 16) ainsi que des propositions d'extensions que nous décrivons rapidement ci-dessous en donnant quelques exemples de classes qui ont été étudiées comme potentiellement intéressantes.

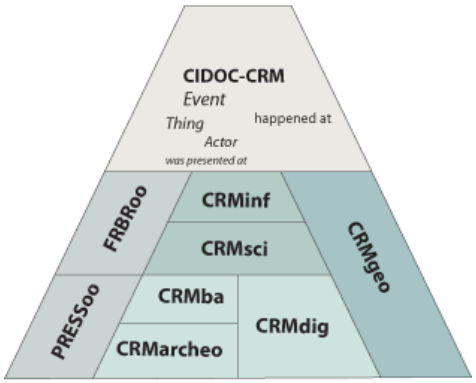


Figure 16 : Pyramide de la famille de modèle du CIDOC CRM (extensions) http://www.cidoc-crm.org/node/191

Parmi les extensions, CRMgeo intègre des propriétés spatio-temporelles d'entités temporelles et d'entités persistantes et CRMdig est consacrée aux œuvres numérisées. La classe *D1 Digital Object*, sous-classe de *E73 Information Object* est particulièrement intéressante. CRMsci intègre les métadonnées relatives aux observations scientifiques, aux mesures et aux données traitées dans les sciences descriptives et empiriques telles que la biodiversité, la géologie, la géographie, l'archéologie, la conservation du patrimoine culturel et autres, dans les environnements informatiques de recherche et les bibliothèques de données de recherche. Une des classes potentiellement intéressantes est la *S1 Matter Removal* qui permet de décrire l'enlèvement d'un composant ou d'un morceau d'un objet physique, l'enlèvement d'une couche archéologique ou géologique, le prélèvement d'un échantillon de tissu sur un corps ou d'un échantillon de fluide sur une masse d'eau. CRMinf soutient l'argumentation. Par exemple, la classe *l5 Inference making* permet de décrire les inférences et leur condition de vérité. CRMarcheo sert aux archéologues afin de palier le problème de collecte et de récupération de données commensurables sur le processus de fouille. Par exemple, la classe *A1 Excavation Process Unit* comprend les activités de fouilles au sens de l'archéologie, qui sont documentées comme un ensemble cohérent d'actions consistant à enregistrer et à retirer progressivement de la matière d'un emplacement prédéterminé selon des règles spécifiques.

Parmi les propositions d'extensions, CRMtext est destinée à modéliser les entités textuelles anciennes. Il y a par exemple des classes comme *TX2 Writing, TX6 Transcription, TX7 Written text* segment

[63] *E97 Monetary Amount, E96 Purchase, E97 Monetary Amount, E98 Currency, E99 Product Type*

[64] *P175 starts before or with the start of, P176 starts before the start of, P177 assigned property of type, P179 had sales price, P180 has currency, P182 ends before or with the start of, P183 ends before the start of, P184 ends before or with the end of, P185 ends before the end of, P186 produced thing of product type, P187 has production plan, P188 requires production tool, P189 approximates, P190 has symbolic content, P191 had duration, P195 was a presence of, P196 defines, P197 covered parts of, P198 holds or supports*.

(Felicetti et Murano 2021). CRMba sert à la documentation de bâtiments archéologiques et est actuellement (juin 2021) une proposition d'approbation par le CIDOC CRM-SIG. Un des buts de ce modèle est par exemple de reconnaître l'utilisation d'un bâtiment et son évolution au fil des ans. La classe *A2 Stratigraphic Volume Unit* comprend les parties reliées entre elles du terrain ou d'une autre structure solide sur, dans ou sous la surface de la terre ou du fond marin présentant une certaine homogénéité de structure ou de substance et complètement délimitées par des surfaces ou des discontinuités en substance ou en structure par rapport aux autres parties du terrain ou aux surfaces des objets/trouvailles.

Les extensions et propositions d'extensions ne seront cependant pas utilisées dans ce travail car la version de base du CIDOC CRM a été jugée suffisante pour l'utilisation qui en est faite.

3.4.2 Ontologies pour la recherche d'information

Certaines ontologies servent à la recherche d'information. Dans ce cadre, Aussenac-Gilles, Baziz, et Hernandez montrent l'importance de la terminologie (Aussenac-Gilles, Baziz, et Hernandez 2006). Pour certains, la terminologie représente cette charnière entre les connaissances et leur expression dans la langue (Reymonet, Thomas, et Aussenac-Gilles 2007). Dans ce sens, les ontologies « doivent à la fois être un support à la formulation de recherches et une ressource pour définir des métadonnées destinées à annoter des documents. » (Reymonet, Thomas, et Aussenac-Gilles 2007, 171). Elles sont également utilisées pour préciser les ambiguïtés du langage naturel. Dans cette lignée, les ontologies améliorent le fonctionnement du web. Au début des années 2000, dans le cadre des travaux du W3C et des réflexions sur le web sémantique par Tim Berners Lee, les ontologies deviennent le noyau des transformations du web vers le web sémantique (Charlet, Bachimont, et Troncy 2004).

3.4.3 Construction de dictionnaires électroniques

La combinaison des approches terminologiques et ontologiques propose de nouvelles méthodes pour la construction des dictionnaires électroniques (*e-dictionary*). Ramos, Costa, Roche et *al.* (2017) et Ramos (2020) proposent un dictionnaire électronique conçu comme un produit multilingue et multimodal dans lequel plusieurs ressources, notamment linguistiques et conceptuelles, sont associées. Ramos conçoit d'un côté une micro-ontologie de domaine, OntoCork, où les concepts sont enrichis de définitions en langage naturel et complétés par des images, soit annotées de méta-informations, soit enrichies d'hyperliens vers des informations supplémentaires, comme une ressource lexicographique et de l'autre TermCork, qui est un projet qui découle de l'organisation de la connaissance dans le domaine spécialisé du liège. Ce projet propose une base de données de connaissances terminologiques pour alimenter le dictionnaire électronique. Pour l'auteur, ce type de dictionnaire électronique est vu comme un outil terminologique utile dans la société de l'information numérique actuelle : la comptabilité pour ses principales caractéristiques ainsi qu'un format électronique pouvant être intégré au Web sémantique grâce à son format de données d'interopérabilité. Le but est de réduire les ambiguïtés et d'accroître l'efficacité de la communication entre les experts du domaine, les futurs experts et les professionnels de la langue.

3.4.4 Ontologies pour inférer de la connaissance

Les ontologies sont utilisées depuis les années 90 dans le cadre du web pour fournir des systèmes capable d'inférences automatiques et d'interopérabilité pour réaliser le web sémantique (Machado 2021). L'auteur mentionne néanmoins les limites des inférences dues au langage utilisé (RDF, puis par extension RDFS). L'auteur développe le fait que RDF n'est pas un langage de web sémantique mais n'est toujours qu'un langage orienté web ce qui signifie qu'il traite indifféremment les individus (instances), des classes (types), et des propriétés (attributs ou relations), en les visualisant tous comme des « *resources* », ceci étant différent de la logique de description.

3.5 Éléments d'un modèle ontologique

Comme le mentionnent Gandon et Passant « une ontologie inclut couramment une organisation hiérarchique des concepts pertinents et des relations qui existent entre ces concepts, ainsi que des règles et axiomes qui les contraignent » (Gandon et Passant 2015, 8).

3.5.1 Concept

« Un concept peut être décrit par un système symbolique (Granger 1960, 1979) : notamment une équation, une représentation visuelle, une icône, une ou plusieurs unités linguistiques. » (Béjoint et Thoiron 2000, 94). Si le dernier aspect intéresse la terminologie et ces auteurs, la représentation visuelle intéresse la modélisation ontologique. En effet, élaborer ou utiliser un modèle ontologique, c'est conceptualiser un monde, c'est faire du domaine étudié, un concept ou un ensemble de concepts. Il est intéressant de revenir ici sur les différentes définitions du concept.

Le concept est défini dans des ouvrages sur la recherche en sciences humaines et sociales comme « une idée, plus ou moins abstraite, un symbole qui désigne ou représente une réalité plus ou moins vaste » par Tremblay (1968, 72) ou par Grawitz (2000, 18) comme « une abstraction, une pensée, un moyen de connaissance », ou encore par Gauthier (2010, 514) comme « une abstraction recouvrant une variété de faits ; [une] structure mentale réunissant les attributs d'une réalité permettant aussi de la reconnaître et de la distinguer des autres ; [une] catégorie classificatoire première d'une systématisation théorique ; [une] nomination d'un phénomène complexe, mais non réductible au cœur de la réalité à expliquer », et enfin pour Dépelteau (2010, 175) le concept est « un mot ou un ensemble de mots désignant un ensemble de phénomènes » ou « une définition conventionnelle d'un phénomène ». De toute ces définitions nous retenons le caractère abstrait et la possibilité de nommer cette abstraction par une nomination. Néanmoins, ce ne sont pas ces définitions qui s'appliquent aux concepts des modèles ontologiques.

Dans le domaine plus spécifique de la représentation et la modélisation des connaissances, Paquette définit un concept comme une « connaissance qui décrit la nature des objets d'un domaine (le quoi). Ainsi, un concept décrit une classe d'objets par leurs propriétés communes ; chaque objet représente le concept et se distingue des autres par des traits spécifiques (valeurs) que prennent ses

propriétés » (Paquette 2002, 338). Cette définition ne mentionne plus l'abstraction et s'applique donc mieux à la représentation et modélisation des connaissances.

C'est ce que Borst précise dans sa définition des ontologies (Borst 1997, 12), en modifiant légèrement la définition de Gruber (1993, 199) « *an explicit specification of a conceptualization* » par « *an ontology is a formal specification of a shared conceptualization* », où la conceptualisation, l'aspect formel et partagé sont expliqués dans (Studer, Benjamins, et Fensel 1998, 25) :

> « *A 'conceptualization' refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machinereadable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group* ».

La conceptualisation implique donc le fait d'abstraire, mais aussi le raisonnement par le fait de devoir identifier des concepts pertinents d'un phénomène. Les modèles ontologiques sont le résultat formel d'une conceptualisation et la représentation mentale de ces modèles sont des concepts.

La stabilité d'un concept peut être évaluée en fonction de son ancienneté et de son utilisation, en fonction de la réussite des expérimentations effectuées à partir de celui-ci. Il existe des invariants conceptuels dont la sémantique est stable : le concept de l'eau par exemple, pour laquelle les expérimentations conservent toutes ses propriétés. Nous l'avons vu, dans les travaux en linguistique et en terminologie, le concept peut être représenté par un terme. Dans le domaine de la documentation, l'utilisation d'un terme-concept permet de représenter le contenu d'un document. Dans les modèles ontologiques, les concepts désignent la façon dont les acteurs (ou experts) conceptualisent leurs pratiques dans la langue.

En modélisation conceptuelle, le terme de classe est employé pour désigner un type abstrait caractérisé par des propriétés, des attributs et des méthodes, communes à un ensemble d'objets et permettant de créer des instances de ces objets, ayant ces propriétés. La classe est, dans ce sens, comme le concept, caractérisée par des propriétés et des relations avec d'autres classes. C'est ce terme qui est employé dans la terminologie du modèle ontologique du CIDOC CRM (*class*). Ainsi, *E22 Man-Made Object* est une classe. Dans le domaine minier, cette classe permet par exemple de créer l'instance *salle des douches*. Nous l'avons vu, les instances peuvent être nommées *individus* et forment le niveau de granularité le plus fin au sein de l'ontologie.

3.5.2 Relation ou propriétés

Comme Arsenault et Noruzi, par relation, « nous entendons l'association, la connexion et l'interaction entre différentes entités » (Arsenault et Noruzi 2012, 105). En ce qui concerne les ontologies, les relations sont sémantiques et comme le rappellent Hubert, Mothe et Will, « une relation sémantique lie des concepts » (Hubert, Mothe, et Will 2009, 5). Les auteurs listent les relations les plus courantes dans

la littérature : « les relations d'équivalence, taxonomiques, patronymiques, de dépendance, topologique, causale, fonctionnelle, chronologique (Gómez et *al.* 2000) » (Hubert, Mothe, et Will 2009, 5). Ils rappellent enfin que si les noms de relations sont communs avec celles des thésaurus, « elles ne portent pas sur les mêmes objets (concepts pour les ontologies *vs* termes pour les thésaurus) » (Hubert, Mothe, et Will 2009, 5).

Des types de relations spécifiques comme celles liant une ressource sujet et une ressource objet (attributs ou associations de classe de l'ontologie) sont représentées par des propriétés (*properties*) (Djurić, Gašević, et Devedžić 2005, 120). Le modèle ontologique du CIDOC CRM a ainsi des propriétés symbolisées par un P. Par exemple : *P7 took place at (witnessed)* relie une ressource sujet *E4 Period* et une ressource objet *E53 Place*.

3.5.3 Axiomes, inférences et règles de raisonnement

Messaoudi définit un axiome comme « "une vérité indémontrable qui doit être admise", c'est-à-dire quelque chose d'évident » (Messaoudi 2017, 149). C'est un postulat, une loi liée à un domaine. Les axiomes s'expriment en langage logique (sur une base épistémologique ou ontologique), ils définissent des connaissances qui n'ont pas un caractère strictement terminologique (Staab et Mädche 2000). En d'autres termes, ce sont les règles appliquées aux concepts ou aux relations qui sont rendues exploitables par des langages d'encodage pour réaliser des inférences (Picard 2014, 26). Par exemple, déclarer les deux axiomes suivants : « Si X ancêtre de Y, alors X *ancestor* de Y », « "Si X *ancestor* de Y, alors X ancêtre de Y" » permettraient à un agent informatique de traiter les informations concernant la relation génétique d'ancêtre en français et en anglais (Paquette et *al.* 2003). Les règles de raisonnement permettent de réaliser des inférences.

3.6 Différentes approches de conception d'une ontologie de domaine

Nous présentons dans cette partie quelques approches pour la conception d'ontologie de domaine exemplifiées par des travaux concrets : à partir de glossaires, à partir de textes et à partir de thésaurus. Nous présentons enfin une approche hybride puisant dans ces différentes méthodologies.

3.6.1 À partir d'un glossaire et ses définitions

Cette méthode est employée et décrite par Navigli et Velardi (2008). Les auteurs utilisent l'annotation sémantique automatique et la population dans le domaine du patrimoine culturel. Les définitions en langage naturel provenant des glossaires disponibles dans un domaine donné sont traitées et des expressions régulières sont appliquées pour identifier les relations générales et spécifiques au domaine. Ceci permet ensuite d'annoter automatiquement et de formaliser les gloses d'un thésaurus (Art & Architecture Thesaurus (AAT)) (que les auteurs appellent *glossary*), avec les propriétés (relations conceptuelles) d'une ontologie centrale, le CIDOC CRM (une dizaine de relations, plus la relation hyperonymique). Les glossaires annotés sont convertis en descriptions de concepts OWL et utilisés pour

enrichir le CIDOC. Les auteurs montrent que les méthodes automatiques demandent autant d'efforts que les méthodes non-automatiques.

De son côté, le groupe LAFARGE interrogé par Florence Gicquel (Gicquel et Lafarge 2007) décrit le processus de migration d'un glossaire industriel multilingue vers le web et le développement d'une ontologie appelée Mondeca (dizaine de classes et dizaine d'attributs), utilisée dans la modélisation des données. Dans ce cadre, elle décrit le schéma qui a été retenu qui est de type « *terme pivot égale terme en anglais plus métadonnées et traductions multiples autour de ce terme pivot « (*relation de 1 à n) » (Gicquel et Lafarge 2007). Il est malheureusement difficile de se représenter ceci sans un exemple (absent de l'article), mais nous comprenons le sens de « terme pivot » comme représentation d'un concept lié à ses différentes réalisations dans la langue formalisées (termes français/anglais). Les métadonnées sont probablement les sources dans lesquels apparaissent les termes. Le tout est représenté sous la forme d'une ontologie, base sur laquelle repose l'application qui avait pour but de faire circuler l'information et la rendre disponible pour les utilisateurs du groupe multinational (experts, ingénieurs, techniciens, chefs de produits etc.).

Si les ontologies peuvent être créées à partir de glossaires et de leurs définitions, elles peuvent également être réalisées à partir de textes.

3.6.2 À partir de textes

Les méthodes à partir de textes utilisent principalement le traitement automatique de langage. La première étape est de constituer un corpus de texte représentatif. Bourigault, Aussenac-Gilles et Charlet utilisent cette méthode dans le cadre de la construction de ressources terminologiques et ontologiques (Bourigault, Aussenac-Gilles, et Charlet 2004). Les auteurs nous rappellent, à propos des tâches de modélisation des connaissances que celles-ci « doi[vent] coordonner une approche ascendante basée sur l'exploitation des résultats des outils d'analyse de corpus et une approche descendante guidée par l'application cible et l'organisation générale du domaine. » Concernant les outils d'analyse de corpus, les auteurs utilisent deux logiciels de TAL pour identifier les candidats termes : SYNTEX, un analyseur syntaxique de corpus, dont le résultat est un réseau de mots et de syntagmes puis UPERY, logiciel d'analyse distributionnelle qui « exploite l'ensemble des données présentes dans le réseau de mots et syntagmes construits par SYNTEX pour effectuer un calcul des proximités distributionnelles entre ces unités » (Bourigault, Aussenac-Gilles, et Charlet 2004, 102). Ensuite, « il s'avère que les rapprochements effectués par UPERY sont extrêmement utiles et pertinents pour la construction de classes conceptuelles. » Les outils YAKWA (concordancier) et CAMELEON (logiciel de recherche de relations lexicales à partir de marqueurs linguistiques) sont utilisés pour extraire les relations. Enfin, l'interface TermOnto permet la visualisation des résultats de SYNTEX et UPERY et le logiciel Terminae est utilisée pour modéliser les RTO. Les auteurs concluent sur le fait que construire des RTO est chronophage mais qu'il faut réussir à montrer que ce n'est pas du temps perdu : « Il faudra alors bien arriver à montrer que le coût en temps expert est largement compensé par une augmentation importante

de la qualité et de l'efficacité des systèmes exploitant les ressources. Tel est l'un des enjeux des recherches en acquisition des connaissances à partir de textes. » (Bourigault, Aussenac-Gilles, et Charlet 2004, 108).

Malaisé montre quant à elle dans ses travaux de thèse « l'intérêt des ontologies par rapport aux systèmes à base de mots-clés pour la gestion documentaire » et fait « une proposition pour l'aide à la modélisation d'ontologies différentielles à partir de corpus textuels » (Malaisé 2005, 133). Elle définit l'ontologie différentielle comme « une terminologie organisée de manière strictement arborescente, autorisant toutefois des relations transversales entre termes de différentes branches (qui sont tous organisés selon une hiérarchie stricte), et dans laquelle la position de chaque terme est définie selon le rapport sémantique qu'il entretient avec les autres termes » (Malaisé 2005, 71-72). Elle utilise une méthodologie en traitement automatique du langage (TAL) et structure les termes à partir de patrons lexico-syntaxiques. Par exemple, le « patron DET SN, *entre autres* SN […] permet d'extraire une phrase contenant Les méningites, entre autres pathologies..., et de mettre en relation méningites et pathologies » (Malaisé 2005, 65).

Si nous construisons notre ontologie de domaine sur la base de textes comme ces auteurs, notre méthodologie ne s'appuie en revanche pas complètement sur du TAL.

3.6.3 À partir d'un thésaurus

Les travaux de Mothe et Hernandez présentent une méthodologie de construction d'ontologie de domaine qui prolonge TERMINAE (utilisation d'un corpus de référence) et intègrent les ressources terminologiques de type thésaurus (Mothe et Hernandez 2006). L'ontologie réalisée est appelée TtoO. La transformation repose sur quatre étapes : « l'extraction d'informations du corpus, l'identification des concepts issus du thésaurus, la construction de la structure de l'ontologie (hiérarchie de concepts et relations associatives entre concepts) » (Mothe et Hernandez 2006, 41). La méthodologie employée permet de ne pas trop solliciter les experts et enrichit le domaine de connaissances (les thésaurus sont parfois anciens et non mis à jour). La méthodologie est adaptée pour des thésaurus qui sont construit en « respectant la sémantique de la relation "est un" ». Les auteures précisent que si cela n'est pas le cas « une étape supplémentaire doit être ajoutée afin de distinguer les différentes relations telles que "est une partie de" ou "est une instance de" » (Mothe et Hernandez 2006, 41).

Hubert, Mothe et Will présentent également une méthodologie pour produire une ontologie légère de domaine à partir d'un thésaurus (Hubert, Mothe, et Will 2009). Les auteurs se focalisent sur la structure d'un thésaurus et en retirent des règles qui permettent de le transformer en OWL. En ce qui concerne les termes, les auteurs mentionnent qu'ils « deviennent des labels dans l'ontologie » (Hubert, Mothe, et Will 2009, 9). Si ce sont des non-descripteurs, ils seront des labels de termes. Si ce sont des descripteurs, ils seront des labels de concepts. Les auteurs proposent que les non-descripteurs soient considérés comme des labels d'instances et que les descripteurs comme des labels de classes étant donné qu'en

OWL, « les classes représentent des ensembles d'éléments (instances) possédant les mêmes propriétés » (Hubert, Mothe, et Will 2009, 9).

3.6.4 Approche hybride

L'approche que nous utilisons est hybride : d'une part, elle utilise des ressources lexicales (de plusieurs types : lexique, dictionnaire, terminologie, glossaire) pour identifier les unités linguistiques du domaine dans des corpus documentaires, d'autre part elle s'aide du sens porté par les définitions et de ressources classificatoires (SOC) pour structurer le domaine et réaliser un thésaurus. Enfin, les concepts portés par les termes du thésaurus sont alignés avec les classes du modèle ontologique du CIDOC CRM qui est au préalable réduit par une identification des classes et des propriétés potentiellement représentatives du domaine étudié. L'analyse des documents dans lesquels les unités linguistiques ont été identifiées permet de rendre compte du lien entre celles-ci (au niveau linguistique et extralinguistique par interprétation) et la possibilité de les formaliser grâce aux propriétés du CIDOC CRM.

CONCLUSION DE LA PARTIE 1

Nous l'avons vu dans le chapitre 1, le champ d'étude dans lequel s'insère nos travaux, l'organisation des connaissances, est issu de plusieurs courants d'investigation portés par des grands noms tels qu'Otlet, Bliss, Hjørland, Dahlberg et bien d'autre encore, ayant tous eu comme volonté d'étudier le continuum des connaissances afin de les rendre manipulables et fructueuses. Nous avons vu les limites de la structuration des connaissances à l'aide des systèmes traditionnels par rapport à la représentation visuelle graphique d'un domaine. À partir des années 2000, des domaines qui ne se formalisaient peu ou pas commencent à utiliser la représentation des connaissances (Bourigault, Aussenac-Gilles, et Charlet 2004). Ceci engendrant des questionnements concernant la nature des données, de l'information, de la connaissance et remettant en question les standards, les normes et les modes de représentation. Aujourd'hui, elle fait partie intégrante de nombreux projets en sciences humaines et sociales et promet l'échange de connaissances entre disciplines éloignées, notamment grâce à la standardisation des langages, le passage des données sur le web et la volonté de communiquer entre l'homme et la machine grâce au web sémantique. Pour pouvoir représenter graphiquement des connaissances, nous avons vu qu'il fallait tout d'abord définir ce que l'on entendait par représentation mentale et bien distinguer le « concept », de la « notion » et du « sens ». Une fois la distinction effectuée, la problématique de les projeter sur des supports sensibles et manipulables apparaît. Une autre question, de représentation globale (dans le cadre de l'étude d'un domaine de spécialité), est également posée. La modélisation des connaissances comme représentation structurée des connaissances étudie les ontologies, dernière solution stable proposée. Nous avons vu les trois principaux points qui posaient problème à ce propos il y a vingt ans puis les questionnements au niveau documentaire, il y a dix ans. Aujourd'hui, le web sémantique semble être la solution aux problèmes évoqués alors. Le chapitre 2 consistait à revenir sur les approches documentaires de l'organisation des connaissances en présentant l'hétérogénéité des SOC

d’une part et leur passage vers le web sémantique d’autre part et soulevait dans ce cadre, la question de l’interopérabilité puis faisait mention d’importants projets culturels nécessitant des SOC spécifiques. Le chapitre 3 consistait à présenter les approches ontologiques de l’organisation des connaissances. Nous avons vu que les ontologies étaient étudiées dans différents domaines tels que la philosophie, l’informatique, la linguistique, les sciences de l’information et de la communication et que la généricité conceptuelle opposait les ontologies de haut niveau à celles de domaine ou celles d’application. Les travaux actuels définissent les ontologies comme des SOC notamment dans le cadre de l’ISKO. Pour illustrer cela, nous revenons sur quelques travaux de doctorat récents en sciences de l’information et de la communication traitant des ontologies et dégageons sept grands axes pour lesquels les ontologies sont considérées comme SOC : la recherche d’information, l’identification de ressources sur le web, la réutilisation de données, la réduction des ambiguïtés communicationnelles, la description de typologies, la médiation et la représentation et l’organisation d’un domaine. Nous plaçons nos travaux dans ce dernier axe. Nous développons ensuite les différents rôles des ontologies : pour la description, le partage, l’échange et la réutilisation de connaissances, pour la recherche d’information, pour la construction de dictionnaires électroniques, pour inférer de la connaissance. Ceci nécessite de réfléchir aux composants des modèles ontologiques : les concepts, les relations, les axiomes. Enfin, nous avons montré différentes approches de construction d’une ontologie : à partir de glossaire, à partir de textes, à partir de thésaurus. La théorie nous a permis d’imaginer une approche hybride (ascendante et descendante) puisant dans ces trois approches afin de concevoir nos trois SOC.

PARTIE 2. Le patrimoine minier comme domaine de connaissance

Définir le patrimoine minier comme un domaine de connaissance nécessite de revenir sur ce qu'est un domaine et sur ce qui fait la particularité d'un patrimoine dit « minier ». L'enjeu, nous l'avons vu, est de réussir à structurer celui-ci pour pouvoir le représenter et pouvoir l'analyser.

D'une part, en lexicologie, le *domaine* est une catégorisation lexicale d'après Mathieu-Colas (2007), au même titre qu'une classe sémantique. D'après ce auteur[65], il est possible de rattacher les mots de la langue générale à un domaine, de la même manière que cela est effectué pour les termes scientifiques et techniques en terminologie. Selon Béjoint et Thoiron (2000, 182), qui étudient le sens en terminologie, « le domaine permet d'indiquer le système conceptuel auquel appartient le concept ». Dans notre cas, le domaine se définit comme système conceptuel auquel appartiennent un ensemble de concepts définis ou à définir, l'ensemble des termes scientifiques et techniques utilisés par des personnes qui co-construisent le domaine, l'ensemble des mots de la langue générale permettant de créer du sens dans les écrits et les oraux réalisés par ces personnes identifiées, l'ensemble des catégorisations lexicales et des classes sémantiques.

D'autre part, en terminologie et d'après le GDT, le terme « minier » relève d'un seul domaine, celui de l'« industrie minière ». Quant au terme « patrimoine » il relève des domaines de la comptabilité, du droit, de l'économie, de l'éducation et de la gestion. Au vu des définitions, le domaine de l'éducation correspond le mieux au terme que nous utilisons. Son sous-domaine : « interprétation du patrimoine » insère le terme au sens d'« ensemble des biens matériels et immatériels, qu'ils soient d'ordre naturel, culturel ou historique, appartenant à une communauté et transmissibles d'une génération à une autre. »

Plus précisément, sur Wikipédia, un article en ébauche définit le patrimoine minier comme désignant « l'héritage matériel et immatériel généré par les activités minières, et auquel des valeurs patrimoniales sont attribuées par les sociétés qui en sont responsables. Le patrimoine minier constitue une catégorie du patrimoine culturel, et en son sein, du patrimoine industriel »[66]. Le caractère d'ébauche avec dernière modification le 18 juin 2021 signifie, pour nous, que ce domaine est récent. Le patrimoine minier peut donc être considéré comme un domaine de connaissance à part entière et nous le définissons dans le chapitre 4. Comme tout domaine patrimonial, de nombreux enjeux sont posés dans le cadre de sa sauvegarde et de sa représentation et notamment la prise en compte de sa variété d'objets aussi bien matériels qu'immatériels. Une partie des éléments de ce patrimoine a été inscrite à l'UNESCO en 2012 mais le support de celui-ci (la langue) reste en danger avec un nombre toujours plus faible de locuteurs

[65] Dans son article, Mathieu-Colas montre en quoi le travail sur les classes et les domaines au niveau de la langue permet de désambiguïser le langage naturel (Mathieu-Colas 2007). La désambiguïsation du langage naturel sert à la recherche d'information, donc à l'identification de sources pertinentes et par expérience circulaire, à la définition précise des domaines.
[66] https://fr.wikipedia.org/wiki/Patrimoine_minier

et une variation difficile à représenter. Enfin, si la documentation est importante et aide à créer d'intéressantes expositions et nouveaux contenus, son hétérogénéité de forme, de langue, d'accès (dispersion dans plusieurs lieux de dépôt, accès en ligne ou en physique, diversité des langages documentaires) rend difficile la représentation globale et la richesse du domaine.

La méthodologie employée pour concevoir les trois SOC, décrite dans le chapitre 5, prend en compte ces différents enjeux. Tout d'abord, la constitution d'un corpus hétérogène représentatif du domaine permet de montrer sa richesse et sa variété. Au sein de ce premier corpus, des ressources lexicales sont identifiées comme pertinentes à la création d'une terminologie. Des ressources classificatoires, ensuite, sont analysées entre elles, puis au regard de la terminologie pour structurer les concepts de celles-ci. Ces ressources permettent en effet de créer une structure cohérente à enrichir sémantiquement grâce à l'analyse des autres ressources documentaires pour obtenir un thésaurus, en vue d'indexer le contenu conceptuel (porté par des termes) de toute ressource du domaine. Le résultat est enfin confronté à un modèle conceptuel de référence, le CIDOC CRM pour enregistrer et décrire les objets du patrimoine minier grâce à un formalisme sémantique riche et adéquat. Le résultat se compose de trois SOC, TerminoMines, ThesoMines, OntoMines, présentés d'une manière factuelle dans le chapitre 6 puis sont analysés et discutés en partie 3.

CHAPITRE 4. Définir le patrimoine minier

Ce chapitre présente le domaine d'application (le domaine houiller, plus précisément celui façonné autour de l'activité d'extraction houillère) et définit son patrimoine qui répond à des enjeux actuels forts, ceux de la disparition des anciens mineurs porteurs d'une mémoire sociale et technique ayant forgé tout un territoire qui est celui du bassin minier situé sur la région actuelle des Hauts-de-France. Il s'agit d'effectuer plusieurs focalisations : patrimoniale pour identifier les objets (au sens large) patrimonialisés ou à patrimonialiser, notionnelle (sur le domaine des mines de charbon) et une territoriale enfin (le bassin minier tel qu'il est inscrit à l'UNESCO). Les instances internationales du patrimoine telles que l'UNESCO et le TICCIH ont fourni de 1954 à 2003 des définitions d'un niveau de granularité assez haut sur lesquelles nous nous appuyons pour identifier les documents traitant du patrimoine dans son ensemble (culturel, naturel et industriel). Dans une première partie est évoquée la diversité des objets du patrimoine à partir des définitions de ces grandes instances (4.1). Ces définitions servent à identifier les objets du patrimoine minier. Dans une deuxième partie, nous revenons sur l'inscription du bassin minier au patrimoine mondial (4.2.1) et les enjeux liés à une première période marquée par l'activité minière et une deuxième marquée par l'arrêt de cette activité (4.2.2). Une troisième partie traite de ces enjeux : linguistiques tout d'abord avec la langue comme vecteur du patrimoine immatériel (4.3), puis la nécessité de reconstituer un lieu de travail souterrain qui n'existe plus (4.4) et nous soulevons enfin des questions posées par la documentation de ce patrimoine, dispersée et hétérogène de par ses formats et son accès (4.5) et la nécessité de concevoir des outils pour pouvoir enregistrer et mettre en visibilité les connaissances issues de celle-ci, traitée dans le chapitre 5.

4.1 Un patrimoine à plusieurs facettes

4.1.1 Patrimoine culturel, naturel, matériel et immatériel

La notion de patrimoine a été définie par plusieurs acteurs. Nous nous concentrons sur les définitions fournies par les grandes instances (UNESCO, TICCIH) destinées à cadrer sa sauvegarde. Tout ceci forme une riche matière pouvant être étudiée par les scientifiques. Chacune des conventions définit le patrimoine avec précision mais également une certaine souplesse permettant à toute communauté d'enregistrer son propre patrimoine.

De 1954, date de la première convention de l'UNESCO, jusqu'à 2003, date de la convention actuelle, l'UNESCO redéfinit ce que l'organisation considère comme patrimoine des biens culturels en (1954) au patrimoine mondial culturel et naturel en (1972), au patrimoine culturel étant aussi bien matériel qu'immatériel en (1982) avec des précisions sur le patrimoine culturel immatériel et ses domaines de manifestations en (2003). La même année, la Charte Nizhny Tagil du TICCIH définit le patrimoine industriel. Dans les encadrés ci-dessous, nous reprenons chacune des définitions. Les termes que nous avons mis en gras deviennent des catégories pertinentes pour décrire le patrimoine houiller.

Définition :
« Aux fins de la présente Convention, sont considérés comme biens culturels, quels que soient leur origine ou leur propriétaire:
a. Les biens, **meubles ou immeubles**, qui présentent une grande importance pour le patrimoine culturel des peuples, tels que les **monuments d'architecture, d'art** ou **d'histoire, religieux** ou **laïques,** les **sites archéologiques,** les **ensembles de constructions qui, en tant que tels, présentent un intérêt historique ou artistique,** les **œuvres d'art,** les **manuscrits, livres** et autres **objets d'intérêt artistique, historique** ou **archéologique,** ainsi que les **collections scientifiques** et les **collections importantes de livres**, **d'archives** ou de **reproductions** des biens définis ci-dessus;

b. Les **édifices** dont la destination principale et effective est de conserver ou d'exposer les biens culturels meubles définis à l'alinéa a, tels que les **musées**, les **grandes bibliothèques**, les **dépôts d'archives**, ainsi que les refuges destinés à abriter, en cas de conflit armé, les biens culturels meubles définis à l'alinéa a.;

c. Les centres comprenant un nombre considérable de biens culturels qui sont définis aux alinéas a. et b., dits « **centres monumentaux** ». » (UNESCO 1954)

Définition :
« Sont considérés comme "patrimoine culturel" :
− les **monuments**: **œuvres architecturales**, de **sculpture** ou de **peinture monumentales**, **éléments ou structures de caractère archéologique**, **inscriptions**, grottes et **groupes d'éléments**, qui ont une valeur universelle exceptionnelle du point de vue de l'histoire, de l'art ou de la science,
− les ensembles: **groupes de constructions isolées ou réunies**, qui, en raison de leur architecture, de leur unité, ou de leur intégration dans le paysage, ont une valeur universelle exceptionnelle du point de vue de l'histoire, de l'art ou de la science,
− les **sites**: œuvres de l'homme ou œuvres conjuguées de l'homme et de la nature, ainsi que les **zones** y compris les **sites archéologiques** qui ont une valeur universelle exceptionnelle du point de vue historique, esthétique, ethnologique ou anthropologique » (UNESCO 1972, 2)

« Sont considérés comme "patrimoine naturel"[67] :
− les **monuments naturels** constitués par des formations physiques et biologiques ou par des groupes de telles formations qui ont une valeur universelle exceptionnelle du point de vue esthétique ou scientifique,
− les **formations géologiques** et **physiographiques** et les zones strictement délimitées constituant **l'habitat d'espèces animale et végétale menacées**, qui ont une valeur universelle exceptionnelle du point de vue de la science ou de la conservation,
− les **sites naturels** ou **les zones naturelles strictement délimitées**, qui ont une valeur universelle exceptionnelle du point de vue de la science, de la conservation ou de la beauté naturelle »
(UNESCO 1972, 2)

Définition :
« 23. Le **patrimoine culturel** d'un peuple s'étend aux **œuvres** de ses artistes, de ses architectes, de ses musiciens, de ses écrivains, de ses savants, aussi bien qu'aux **créations anonymes**, surgies de l'âme populaire, et à l'ensemble des valeurs qui donnent un sens à la vie. Il comprend les **œuvres matérielles** et **non matérielles** qui expriment la **créativité** de ce peuple : **langue**, **rites**, **croyances**, **lieux** et **monuments historiques**, **littérature**, **œuvres d'art**, **archives** et **bibliothèques** » (UNESCO 1982, 3)

Définition :
« 1. On entend par "**patrimoine culturel immatériel**[68]" les **pratiques, représentations, expressions, connaissances et savoir-faire** - ainsi que les **instruments**, **objets**, **artefacts** et **espaces culturels** qui leur sont associés - que les communautés, les groupes et, le cas échéant, les individus reconnaissent comme faisant partie de leur patrimoine culturel » (UNESCO 2003, 2)
« Le "**patrimoine culturel immatériel**", tel qu'il est défini au paragraphe 1 ci-dessus, se manifeste notamment dans les domaines suivants :
(a) les **traditions** et **expressions orales**, y compris la **langue** comme vecteur du patrimoine culturel immatériel ;
(b) les **arts du spectacle** ;
(c) les **pratiques sociales**, **rituels** et **événements festifs** ;
(d) les **connaissances** et **pratiques concernant la nature et l'univers** ;
(e) les **savoir-faire liés à l'artisanat traditionnel** » (UNESCO 2003a, 2-3)

[67] D'après le GDT, le patrimoine naturel est un « patrimoine qui concerne les composantes écosystémiques caractéristiques d'un lieu, et qui font généralement l'objet de mesures de conservation », la source mentionne en note que « les composantes écosystémiques de ce patrimoine sont principalement **animales** et **végétales**, mais aussi **géologiques** (**géomorphologie**, **fossiles**, **composition des sols**, etc.), hydrographiques (geysers, eaux sulfureuses, chutes, lacs salés, etc.) et climatologiques (climat désertique, microclimat, etc.). Elles peuvent se trouver en milieu terrestre ou marin.
[68] D'après FranceTerm, le patrimoine culturel immatériel, (en anglais intangible cultural heritage) est l'« ensemble des connaissances, des représentations, des savoir-faire, des expressions et des pratiques culturelles propres à un groupe humain et susceptibles d'être transmis ».

4.1.2 Patrimoine industriel

Définition :
Le Comité international pour la conservation du patrimoine industriel définit celui-ci : « Le patrimoine industriel comprend les **vestiges** de la culture industrielle qui sont de valeur historique, sociale, architecturale ou scientifique. Ces vestiges englobent : des **bâtiments** et des **machines**, des **ateliers**, des moulins et des usines, des **mines** et des **sites de traitement** et de raffinage, des **entrepôts** et des **magasins**, des **centres de production**, de **transmission et d'utilisation de l'énergie**, des **structures et infrastructures de transport** aussi bien que des **lieux utilisés pour des activités sociales en rapport avec l'industrie** (**habitations, lieux de culte ou d'éducation**) » (TICCIH 2003, 1).[69]

Ces définitions ont pour vocation à être appliquées, et pour cela, nous partons du principe que chaque terme que nous avons mis en valeur (en gras) représente une classe d'objet au sens de (Paquette 2002, 338). Nous identifions ensuite les instances de ces classes, elles-mêmes représentées par des termes, dans le corpus documentaire. Une méthode qui permet de repérer l'information plus rapidement est d'utiliser la forme des termes à chercher dans les corpus. Dans le corpus de ressources lexicales, l'information sera à chercher dans le champ des définitions. Ainsi, et pour exemple, le terme générique *machine*, que nous prenons comme une classe, permettra d'identifier le terme *haveusse* par sens prochain, car il est défini comme une 'machine pour abattre les blocs de charbon' d'après (B. Turpin et Goudaillier 2004). *Haveusse* est une variante orthographique dialectale du terme *haveuse*. Il suffit ensuite d'étudier les occurrences de ce terme spécifique dans le corpus documentaire pour en retirer ce qui en fait sa spécificité dans le domaine de la mine et de mesurer l'importance qu'il a aux yeux des communautés concernées. Le concept doit en effet, s'il est pertinent, porter un ensemble de représentations mentales partagées entre anciens mineurs et permettre d'indexer les ressources représentatives du patrimoine documentaire minier.

4.2 Un patrimoine inscrit à l'UNESCO

4.2.1 Un « paysage culturel »

Le 30 juin 2012, c'est le 38ème site français inscrit à l'UNESCO. Son inscription est faite au titre de « paysage culturel », « œuvre conjuguée de l'homme et de la nature » d'après les termes de la Convention. Au niveau quantitatif, 353 éléments répartis en 109 sites sont inscrits[70]. 17 fosses et vestiges significatifs de fosse, 21 chevalements, 51 terrils, 54 kilomètres d'anciennes voies de chemin de fer, 3 gares ferroviaires, 124 cités ouvrières, 45 écoles et équipements de loisirs, 17 églises et chapelles, 21 équipements de santé, 3 sièges de compagnies minières, 4 000 hectares de paysage[71]. 25% de la totalité du patrimoine minier a été inscrit au Patrimoine mondial.

La carte suivante (figure 17) représente la zone tampon des biens inscrits au patrimoine. Celle-ci a une surface de 18 804 hectares ; elle concerne 124 communes et comprend environ 712 000 habitants. Elle englobe tous les éléments formant le bien et elle inclut systématiquement les périmètres prévus au titre des monuments historiques et, le cas échéant, des sites naturels protégés. Elle comprend des éléments

[69] Pour un regard scientifique plus approfondi sur le patrimoine industriel, nous renvoyons aux articles de Marina Gasnier (Gasnier 2011, 2019) ou Edith Fagnoni (Fagnoni 2015). Pour le patrimoine industriel minier numérique, nous renvoyons à Anthony Pamart (Pamart 2013).
[70] https://fr.wikipedia.org/wiki/Bassin_minier_du_Nord-Pas-de-Calais
[71] Données de la brochure (Mission Bassin minier 2015)

du patrimoine industriel minier de moindre valeur mais ceux-ci renforcent les significations d'ensemble et ils expriment la continuité territoriale et paysagère du bassin minier. Elle permet de protéger des angles de vision sur le bien à partir de ses principales voies d'accès. Elle est aussi appelée : « zone de cohérence paysagère »[72].

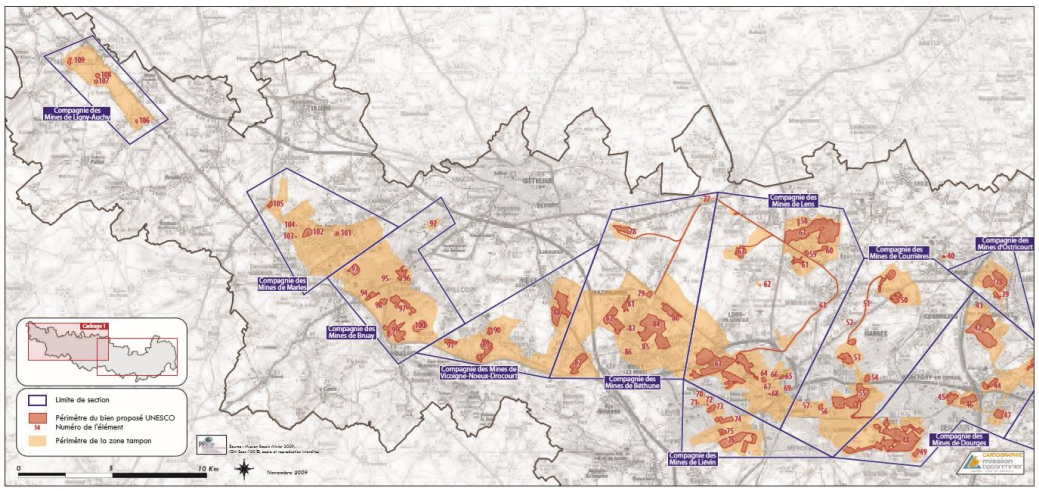


Figure 17 : Zone tampon des biens inscrits au patrimoine mondial de l'UNESCO disponible en ligne à https://whc.unesco.org/fr/list/1360/multiple=1&unique_number=1776

Comme nous pouvons le constater, la zone tampon (en orange clair) dessine un périmètre plus large que la seule limite spatiale des biens inscrits (en orange foncé sur la carte). Cela permet de représenter la richesse du patrimoine minier et notamment le patrimoine immatériel, qui n'a de territoire que la situation des entités matérielles.

4.2.2 Un espace marqué par deux temporalités

Nous nous intéressons dans notre travail à toute documentation, tout objet, toute trace révélatrice de ce qui a touché ce territoire du début des exploitations à aujourd'hui, c'est-à-dire à partir du 3 février 1720, date à laquelle la première veine de houille du bassin minier du Nord–Pas-de-Calais est découverte à la fosse Jeanne Colard située à Fresnes sur Escaut. Nous prenons l'événement de la remonte du dernier morceau de charbon à la fosse 9 à Oignies le 21 décembre 1990 (figure 18) comme marqueur de la fin définitive des mines. La période d'analyse s'étend donc sur environ trois siècles. Ceux-ci peuvent se décomposer en deux parties : 1. la période de l'exploitation (1720-1990) et 2. la période de l'après-mine (1990 à aujourd'hui), la première ayant une influence forte sur la deuxième. L'ambition n'est pas d'étudier tous les documents produits lors de ces périodes (ce qui serait probablement impossible), mais de n'exclure aucun nouveau document identifié par la sélection d'un premier corpus et s'insérant dans ces périodes et dans la thématique, seul moyen de réaliser une véritable ontologie de domaine.

[72] http://www.missionbassinminier.org/nos-chantiers/patrimoine/le-perimetre-du-bien-inscrit.html

Figure 18 : Photographie d'une berline marquant l'activité minière à la fosse 9 d'Oignies de 1930 au 21 décembre 1990, fosse 2 d'Oignies

L'extraction minière a marqué le territoire économiquement et socialement. Depuis le 1er janvier 2018, 150 communes du bassin minier (75 dans le Nord, 75 dans le Pas-de-Calais) bénéficient « d'une mesure exceptionnelle de soutien à la création d'entreprises, par un allègement d'impôts »[73]. La zone appelée « zone franche » bénéficie donc d'un dispositif « bassin urbain à dynamiser (BUD) ». Le territoire est en effet aujourd'hui caractérisé par un taux de chômage et de pauvreté élevé, ce qui n'a pas toujours été le cas et qui nécessite de ce fait, d'importantes actions de redynamisation. Au niveau humain, la première période a permis la création de syndicats mobilisant les ouvriers à faire la grève pour obtenir de meilleures conditions de travail, a fait se déplacer de la main d'œuvre venant de l'étranger qui a aussi été victime d'accidents liés aux risques du travail ou de maladie créant de fortes valeurs humaines de solidarité mais avec une mémoire fragile, les anciens mineurs ne voulant pas forcément garder en mémoire des événements tragiques. Enfin, l'arrêt des exploitations a continué de transformer le territoire ce qui ne permet pas de le figer et rend difficile sa représentation. Des enjeux écologiques et culturels sont venus s'ajouter aux enjeux sociaux et politiques. La documentation et les archives de ces périodes permettent aux professionnels de réfléchir au présent et à l'avenir et de renouveler les questions : quels efforts sont à fournir et dans quelle direction pour vivre ensemble dans un monde meilleur ? Que doit-on à la mémoire de ces hommes « morts pour la France » ?

Ainsi, au niveau de la sauvegarde, plusieurs problématiques ont été soulevées :

- La progressive disparition des personnes parlant le langage de la mine a mis en danger tout ce riche patrimoine immatériel,
- Une importante documentation subsiste et témoigne des actions passées : que doit-on archiver et comment ?
- Depuis le remblayage des puits, une partie du patrimoine souterrain est invisible. Comment rendre visible l'invisible, que doit-on raconter ? Que ne devons-nous pas oublier ?

[73] https://www.prefectures-regions.gouv.fr/hauts-de-france/Actualites/Nord-et-Pas-de-Calais-le-bassin-minier-bassin-urbain-a-dynamiser

Du point de vue des sciences de l'information et de la représentation des connaissances, l'intérêt est de réfléchir à cet « écosystème » à l'échelle du web, de l'archivage numérique et le fait de pouvoir décrire à l'aide de systèmes d'organisation des connaissances toute cette réalité qui s'inscrit dans un territoire marqué spatialement et temporellement. Cet écosystème fait interagir des personnes que nous appellerons acteurs car, ensemble, ils ont façonné et continuent pour certains de façonner le patrimoine minier. Ce patrimoine est à la fois industriel, culturel matériel et immatériel. Grâce à la représentation des liens entre les choses, il est possible de noter les effets produits sur un territoire par l'arrêt d'une activité industrielle. À partir du corpus représentatif de document sur l'activité et sur l'après-mine, il doit être possible de montrer en quoi et comment l'arrêt de l'exploitation a modifié le territoire. Il s'agit de pouvoir répondre à des questions telles que : quels sont les résultats des impacts sur l'environnement dus à l'exploitation minière ? Quelles sont les actions réalisées dans ce sens ? Quelles évolutions a subi le territoire du Bassin minier ? Dans quels secteurs ? Quel rapport peut-on faire entre histoire sociale et économie ? Quelles ont été les luttes pour de meilleures conditions de travail ? Que peut-on dire des valeurs portées par l'activité industrielle ? Quels systèmes politiques sont apparus ? Quelles sont les sources de tensions entre les différents acteurs, faut-il les apaiser ?

4.3 Un patrimoine langagier en danger

La Constitution française reconnaît depuis la loi constitutionnelle du 23 juillet 2008, que « les langues régionales appartiennent au patrimoine de la France ».[74] D'après l'Atlas de l'UNESCO des langues du monde en danger, le picard est considéré comme « *seriously endangered* ».[75] D'après (Moseley 2010), le picard[76], également appelé ch'ti, ch'timi, rouchi a une vitalité considérée comme sévèrement en danger « *severely endangered* » avec un total de 700 000 locuteurs dont 500 000 pour la France et 200 000 pour la Belgique. Les régions concernées sont la Picardie et le Nord–Pas-de-Calais en France ainsi que l'aire spatiale de Tournai à Mons dans l'ouest de la province du Hainaut en Belgique.

4.3.1 Picard ou ch'ti, chtimi, cht'i mi, rouchi ?

Le picard est le terme utilisé par les linguistes pour désigner la langue parlée dans le Nord–Pas-de-Calais et la Picardie, et également dans l'ouest de la Belgique romane. Le ch'ti ou « ch'ti mi » est la variante dialectale de ce picard parlée dans les départements du Nord et du Pas-de-Calais. Le rouchi est une autre variante du ch'ti et est appelée patois de Valenciennes. Ces « langues » sont largement intercompréhensibles et elles partagent des caractéristiques morphosyntaxiques fondamentalement communes. Est considérée comme dialecte une langue régionale de France n'ayant pas eu accès au statut de langue à part entière. Plusieurs auteurs étudient ces dialectes tel que Jean Dauby avec un ouvrage sur

[74] https://fr.wikipedia.org/wiki/Langues_en_voie_de_disparition#France
[75] http://www.unesco.org/languages-atlas/index.php?hl=en&page=atlasmap&lid=398
[76] L'identification de cette langue est détaillée dans la norme ISO 639-3 « Codes pour la représentation des noms de langues - Partie 3 : Code alpha-3 pour un traitement exhaustif des langues » qui décrit les langues avec un niveau de détail linguistique très fin.

le rouchi (Dauby 1979) et (Dauby 1983), ou Marius Lateur qui a écrit sur le parler des mineurs dans la région de l'Artois (Lateur et Roques 1951). Dans l'avant-propos de ce lexique, l'auteur note :

> « Après avoir écrit plusieurs recueils de poésies patoises, de dictons de chez nous, des chansons, des ouvrages folkloriques et littéraires, dont une douzaine ont été publiés, j'ai écrit ce « Lexique du Parler Populaire et Ouvrier des Régions Minières d'Artois », (avec 60 figures), sans aucun signe phonétique, aussi proche que possible de l'orthographe de notre langue nationale (tels : « Ch'est » = c'est, « Arbomd » — rebond, « Baisse » = baiser). Ceci, non pas pour enseigner à parler le langage de notre Pays noir, ni à l'écrire, tout simplement pour faire connaître ses principales expressions aux savants, à ceux qui les ignorent et voudraient s'y intéresser, et par là-même, faire connaître, également, une partie de la vie des vaillantes populations ouvrières de nos Régions jusque dans les travaux miniers. […] » Le lexique « compte, à présent, près de 6000 [termes], parmi lesquels figurent la plupart des termes techniques du métier de mineur. Notre langage populaire est un mélange de patois picard de nos campagnes artésiennes et minières, de français — parfois démarqué —, de « rouchi » et de mots nés au fond de la mine. Le « rouchi » fut introduit chez nous, d'abord : par les mineurs spécialistes de Belgique et surtout du département du Nord, venus pour ouvrir les premiers puits des mines artésiennes, après avoir été sollicités par les dirigeants intéressés, ensuite par d'autres ouvriers, des mêmes endroits, qui vinrent y travailler pour gagner de meilleurs salaires. Ils furent suivis plus tard par un nombre imposant de « Carbonniers » des Mines d'Anzin (main-d'œuvre d'élite) congédiés pour simples faits de grève, principalement de 1884, mais qui furent reçus, à bras ouverts, par leurs nouveaux directeurs, heureux de cette précieuse aubaine. (Une grande partie des belges retournèrent dans leur pays surtout pendant la grève de 1893. Voir « grèfe ». Les familles de ces travailleurs se multiplièrent et leur patois se mêla à celui du pays. C'est pourquoi vous trouverez, dans ce lexique, certains mots écrits et prononcés de différentes façons, étant utilisés aussi bien les uns que les autres. En général, les familles de nos Régions Minières parlent toujours, à peu de chose près, le même langage d'avant la guerre de 1914-18 (qu'elles aiment plus que tout autre) à part un certain nombre de leurs enfants devenus employés d'Etat ou d'Administration » (Lateur et Roques 1951, 11).

Nous ne pouvions pas ne pas mentionner et nous ne pouvions pas paraphraser cet extrait d'avant-propos si bien écrit et à nos yeux, si représentatif des particularités langagières de l'histoire minière et des enjeux du patrimoine langagier de ses régions. Nous notons notamment la richesse du vocabulaire et l'évolution du langage courant mêlé aux termes spécifiques à la mine. Cet avant-propos met également en avant la production d'écrits (poésies patoises, dictons, chansons, ouvrages folkloriques et littéraires) qui entrent tout à fait dans les définitions du patrimoine culturel. Ces écrits, stabilisants en quelque sorte le langage oral sans en supprimer leur richesse offre pour nous une opportunité de travailler à partir de ses formes. Il est alors intéressant d'identifier ces écrits ainsi que d'autres du même type et de les analyser afin d'en retirer les termes portant les concepts représentatifs de ce patrimoine. Une terminologie ainsi constituée permet d'indexer plusieurs ressources, dans le but de représenter un patrimoine documentaire sur la mine.

4.3.2 Documentation et archivage du matériel linguistique

Malgré le foisonnement de la littérature écrite en picard et ses variantes, la langue reste en danger. Pour pallier ce problème, l'UNESCO rappelle l'importance de la documentation et de l'archivage du matériel linguistique et notamment : « [L]a création de centres de recherche où l'on enseignera aux locuteurs des

langues en danger à étudier, documenter et archiver leur propre matériel linguistique est une bonne mesure » (UNESCO 2003b)[77].

4.3.3 Vocabulaire de la mine : enjeux linguistiques

Cette expérience que nous racontons ici est issue des premiers pas sur notre terrain, enrichie par d'autres ensuite. Elle permet de nous concentrer sur les enjeux linguistiques émergeant directement du terrain et orientant notre méthodologie. Lors de notre première visite au Centre Historique Minier de Lewarde en juin 2017, un ancien mineur nous a fait la visite du site, avec un fort accent du Nord. Nous ne connaissions rien alors au monde de la mine et avions noté tous les termes qui nous étaient inconnus, parfois avec une orthographe incertaine. Nous avions par exemple noté *borion* au lieu de *porion*, terme employé pour désigner un agent de maîtrise responsable d'un chantier d'exploitation au fond. Par la suite, nous avons confronté ces résultats aux termes contenus dans nos ressources lexicales. 35 termes y figuraient, 17 non. Comme exemple de terme n'y figurant pas : *CCFC* pour *Congés Charbonniers de Fin de Carrière* puis *pit ponies*, terme issu de l'anglais qui désigne les types de poneys descendus dans les mines ou encore *garde lavabo*. Nous avons donc cherché des occurrences des termes non mentionnés dans les lexiques. Nous avons trouvé quelques occurrences de *garde-lavabo* sur le site web *Histoires de ch'tis* dans une discussion sur les lavabos entre anciens mineurs[78]. Ces derniers critiquaient l'utilisation par certains musées du terme *salle des pendus* (ou *salle des pindus*, *salle ed pindus* en ch'ti) pour désigner le vestiaire ou les bains douches. Cette expression aurait en effet été introduite par des journalistes mais n'était pas du tout utilisée par les mineurs. Nous voulons montrer par là que la défense du vocabulaire réellement utilisé est primordiale pour ces anciens mineurs. Henri, un fils de mineur interrogé nous déclare d'ailleurs en montrant les locaux sur la maquette réalisée par Martial Ansart (figure 19), ancien mineur : « *ça c'était ch'lavabo et non la salle des pendus, ils disaient, soit les mineurs ne disaient jamais ça soit qu'ils disaient ch'lavabo ou bien les bains douche* ». De même, Jean-Marie Minot, un passioné très actif dans la sauvegarde de la mine nous souffle : « *jamais sur un plan on a vu salle des pendus* ». De ceci, nous comprenons l'importance d'utilisation des bons termes, *salle des pendus* pouvant véhiculer une représentation mentale connotée et erronée. La difficulté réside lorsque plusieurs anciens mineurs utilisent des termes différents pour désigner une même chose et sont convaincus d'employer la bonne forme.

[77] https://ich.unesco.org/doc/src/00120-FR.pdf
[78] http://www.histoires-de-chtis.com/index-news-23998-lavabos.php

Figure 19 : Maquette de la mine réalisée par Martial Ansart, ancien mineur interrogé à Givenchy-en-Gohelle en juin 2018

4.4 Un patrimoine souterrain à rendre visible

L'activité minière arrêtée, tout ce qui se situait en-dessous de la terre a été, soit remonté, soit remblayé. En surface, des bâtis ont été détruit ou sauvegardés et des panneaux de mémoire, symboles de l'activité passée ont remplacé certains objets. Aujourd'hui, les associations, les musées et les archives travaillent parfois en collaboration pour mettre en lumière les objets témoins de cette vie passée.

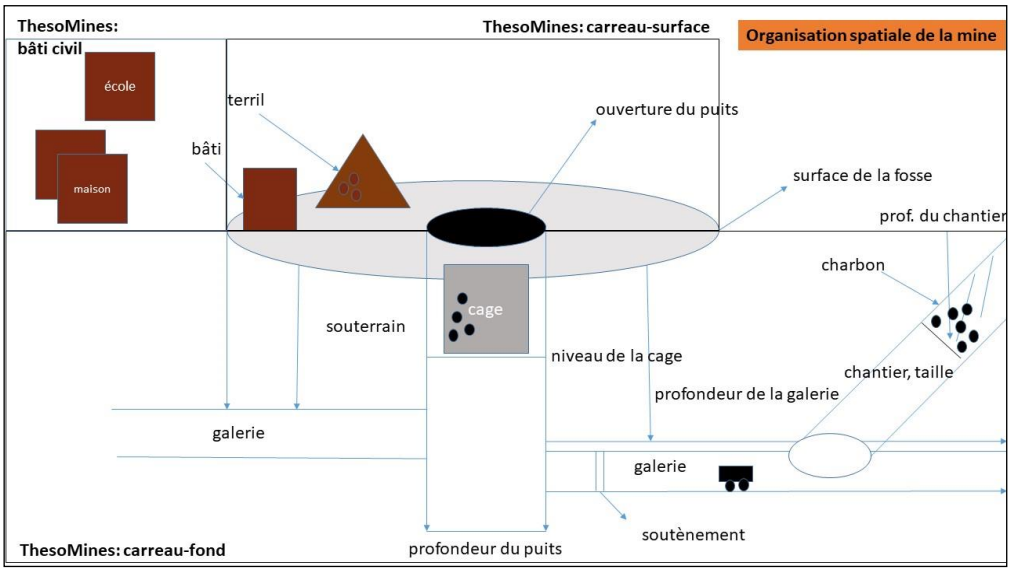


Figure 20 : Schéma représentant une partie du patrimoine minier en surface et en souterrain lors de la première temporalité

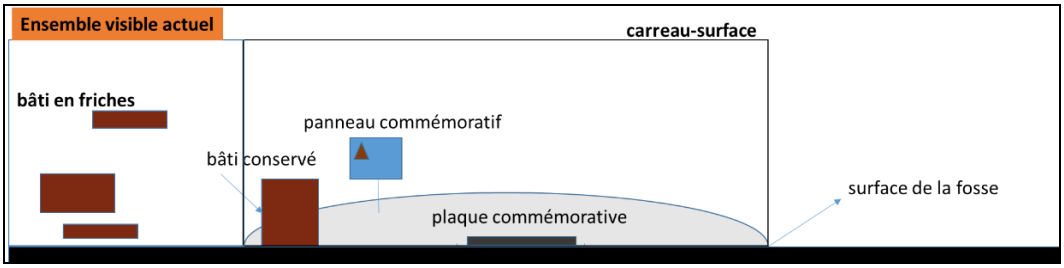


Figure 21 : Schéma représentant la visibilité actuelle, avec un souterrain remblayé

Comme nous pouvons le constater sur ces deux schémas que nous avons réalisés (figure 20) et (figure 21), ce qui pouvait être visible auparavant lors de l'activité minière (uniquement pour les personnes travaillant ou pouvant exceptionnellement descendre au fond) ne se voit plus depuis que le fond a été remblayé (c'est-à-dire rebouché). Dans le bassin houiller du Nord–Pas-de-Calais (contrairement au site de Blegny-Mine en Belgique où les galeries souterraines sont accessibles aux visiteurs via le puits d'origine) tous les puits de mines ont été remblayés. Ceci rend le patrimoine souterrain invisible et implique la reconstitution des lieux comme dans les musées de la mine à Harnes[79], ou à la fosse 2 d'Oignies, mais aussi au Centre Historique Minier (CHM), par exemple, où des galeries ont été entièrement reconstituées. Des objets du fond ont été remontés pour permettre aux visiteurs de ces musées de se replonger dans l'univers de la mine et lui permettre de se représenter ce qu'était le travail de mineur. Dans cette optique, tous documents (iconographiques, textuels, audiovisuels) ainsi que le témoignage des personnes ayant vécu la mine peuvent permettre aux personnels des musées de reconstituer le plus précisément possible des scènes, des espaces passés. Nous réalisons ainsi l'importance de l'exploitation d'un maximum de sources pour reconstituer des ambiances au plus proche de la réalité. C'est notamment dans cette optique que se situent les travaux en réalité virtuelle ou réalité augmentée de Nicolas Lissarrague. Membre du projet MémoMines du laboratoire DeVisu à Valenciennes, Nicolas a par exemple réalisé des numérisations par photogrammétrie permettant de visualiser des objets de la mine en 3D[80].

4.5 Un patrimoine documentaire hétérogène et dispersé

Sur le territoire des Hauts-de-France, plusieurs sites sont liés à l'histoire minière. Depuis la fin de l'exploitation, des questions se posent concernant :

- le dépôt ou la destruction d'objets (certains étant parfois assez encombrants),
- la sauvegarde ou la destruction de bâtiments,
- l'organisation des ressources issue de l'activité minière, celles documentant le passé minier et celles produites lors de l'après-mine et actuellement,
- les actions générales à réaliser pour préserver la mémoire.

[79] https://memomines.hypotheses.org/1763
[80] https://memomines.hypotheses.org/1609 et https://sketchfab.com/search?q=memomines&sort_by=-pertinence&type=models

Pour réfléchir à ces questions et agir en conséquence, des institutions, des organisations, des associations se sont créées, parfois sur la base même de sites miniers comme par exemple le Centre Historique Minier de Lewarde (CHM)[81] installé sur l'ancienne fosse Delloye. Le musée a été ouvert au public en 1984. Le centre a bénéficié d'un apport conséquent de matériels et documents des autres fosses du bassin au fur et à mesure qu'elles fermaient[82].

La méthodologie consiste donc à identifier les acteurs détenteurs de documentation et à les interroger sur celle-ci. Nous verrons comment le CHM organise ses ressources en 5.1.1.2.

L'illustration que nous avons réalisée suite à nos visites de sites miniers (figure 22) montre une partie de cette hétérogénéité qui peut être étudiée (et donc indexée) selon au moins deux points de vue : ce qui est représenté et le mode de représentation. Le mode de représentation est le même pour tous : la photographie. En revanche, ces photographies représentent une diversité de choses : une photographie (représentant la lampisterie avec des lampes et des personnes), une affiche (avec un message de prévention et représentant une berline de charbon pleine qui déraille), une statuette (qui représente Sainte-Barbe), un homme (certainement un mineur, il porte en effet un casque et c'est un objet symbolique), un site minier (nous reconnaissons un chevalement), un buste (représentant Henriette de Clercq, figure liée à la découverte du charbon), un morceau de charbon (et plus précisément du coke comme nous pouvons le lire sur le petit cartel), un paysage (représentant un terril), des livres (sur le mineur, sur la colombophilie), une photographie (représentant le château du directeur des mines), deux berlines, un article de presse (sur le chanteur Pierre Bachelet)… L'un des objectifs de nos travaux est de pouvoir indexer ces types de documents avec un vocabulaire contrôlé à un niveau de granularité fin, puis, par exemple, de pouvoir relier la photographie du château du directeur des mines à un article de presse dans lequel il est mentionné.

Figure 22 : Illustration de la variété des objets du patrimoine, photographies issues de visites de sites miniers, du fond photographique du CHM de Lewarde et de la documentation Mission Bassin minier Patrimoine mondial.

Si les bibliothèques, centres de documentation, archives, associations ou tout organisme culturel ne construisent pas de discours patrimonial contrairement aux musées, leurs documents font référence à

[81] https://www.chm-lewarde.com/fr/
[82] https://www.chm-lewarde.com/fr/le-centre-historique-minier/de-la-fosse-delloye-au-centre-historique-minier/

des objets, des lieux, des personnes faisant partie de ce patrimoine. Toute cette documentation associée aide en effet les professionnels des musées à recréer du contexte, à créer des expositions au plus proche de la réalité passée, à alimenter leur discours de médiation. À cela s'ajoutent les nombreuses informations disponibles en accès libre sur internet et sur le web.

Nous pouvons effectuer une première typologie distinguant ce qui provient de la première période et ce qui provient de la deuxième.

Ressources matérielles : première période (patrimoine matériel)

- Matériel minier (collection d'engin, outils, accessoires)
- Ressources géologiques
  - Fossiles
  - Minéraux
- Ressources pédagogiques de formation
  - Maquettes
  - Supports pédagogiques

Ressources représentatives documentaires : deuxième période

- Ressources textuelles
  - Ressources lexicales (dictionnaires, lexiques, glossaires, vocabulaires)
  - Ressources terminologiques (terminologie, thésaurus)
  - Ressources littéraires (poèmes, folklore, fiction, littérature populaire (BD), prose (biographie, essai littéraire, nouvelle, roman))
  - Ressources journalistiques
- Ressources audiovisuelles
  - Ressources iconographiques et art visuel (croquis, photographies, gravures, peintures, dessins, sculptures)
  - Ressources cartographiques (plans)
  - Ressources musicales (médiathèque)
  - Ressources vidéo (témoignage, films)

Le patrimoine détenu par les communautés l'ayant vécu et inscrit dans des documents reste enfin parfois difficile d'accès. En effet, les derniers anciens mineurs sont âgés et sont pour la plupart malades (ils souffrent souvent de silicose ou d'autres maladies professionnelles liées à la mine) et sont très fatigués. L'accès à leur mémoire en devient de ce fait très fragile.

Quant à l'accès à la documentation du patrimoine minier, elle est empêchée par un certain nombre de facteurs tels que la crise sanitaire, la numérisation, la disposition et la dispersion des différents lieux possédant des ressources. La crise sanitaire survenue entre 2019 et 2021 n'a en effet pas facilité l'accès à la documentation, ni à ces personnes devenant d'autant plus fragiles. Celles-ci n'utilisent par ailleurs pas ou peu le digital. Au centre de documentation et archives du musée de Lewarde, nous verrons que les ressources disponibles ne sont pour la grande majorité pas numérisées. Il n'y a pas non plus de catalogue utilisateur. Pour travailler sur les ressources il est donc nécessaire de rester travailler sur place.

CHAPITRE 5. Vers une méthodologie de conception des SOC du domaine minier

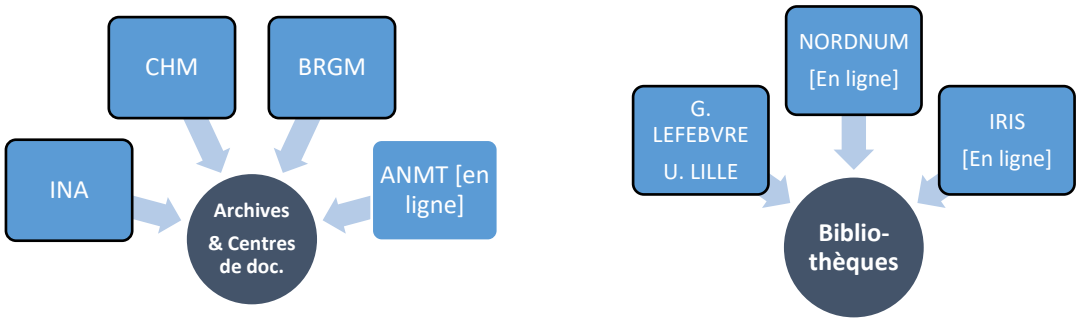
Ce chapitre présente une méthodologie hybride pour la conception des SOC du domaine minier ou plutôt comme nous l'avons vu dans le chapitre 4, du patrimoine houiller. Une première partie (5.1) met en lumière l'accès et l'organisation de la documentation du domaine grâce à des échanges avec certains acteurs principaux, permettant l'identification de ressources lexicales et/ou de ressources classificatoires pertinentes. Un corpus d'étude composé de documents textuels et de documents audiovisuels est ensuite réalisé. Le choix des outils utilisés est justifié en (5.1.5). La deuxième partie (5.2) présente le passage des ressources lexicales à la terminologie de la mine, la troisième (5.3) le passage de la terminologie au thésaurus du patrimoine minier, la quatrième (5.4) du thésaurus au modèle ontologique et la cinquième (5.5) du modèle ontologique à l'ontologie de domaine instanciée.

5.1 Constitution d'un corpus documentaire hétérogène

La constitution d'un corpus documentaire hétérogène passe par une première rencontre avec les acteurs du domaine. Même si les acteurs ont tous des objectifs qui leurs sont propres et ont de ce fait des pratiques hétérogènes, ils partagent la production, la préservation et parfois la mise à disposition du public du savoir du domaine. De nos échanges avec les acteurs, nous avons pu retirer des ressources classificatoires, des ressources lexicales et des documents. Grâce à leur étude systématique, nous avons ensuite pu constituer un corpus d'étude, non exhaustif mais visant la représentativité du domaine. Nous revenons sur ces différentes rencontres et sur leur apport dans notre travail.

5.1.1 Rencontres avec les acteurs détenteurs de documentation

Les quatre diagrammes ci-dessous (figure 23) représentent une partie des acteurs « physiques » rencontrés (archives et centres de documentation, bibliothèques, musées, associations) ainsi que quelques ressources consultées en ligne. Les rencontres avec les acteurs dont le contour est noir sont développées dans cette partie[83].



[83] Chaque entité possède sur le diagramme un lien hypertexte cliquable renvoyant vers son site web.

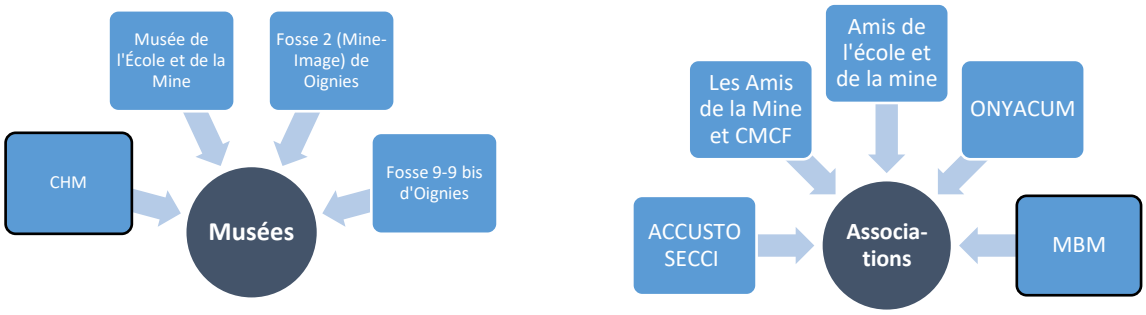


Figure 23 : Diagrammes de quelques acteurs du patrimoine minier

5.1.1.1 Archives audiovisuelles : INA

Le 5 décembre 2017, nous avons rencontré Jean Carrive, responsable documentation à l'Institut National de l'Audiovisuel de la région nord (INA Nord)[84] afin de parler de potentielles ressources classificatoires disponibles sur le domaine de la mine. L'INA met à disposition du public une fresque chronologique (figure 24) composée d'archives audiovisuelles intitulée *Mémoires de Mines*[85] que nous avions identifiée en ligne.

Figure 24 : Capture de la fresque chronologique sur le site : *https://fresques.ina.fr/memoires-de-mines/fresque*

Nous revenons sur cinq points évoqués lors de cette rencontre : les notices descriptives, le type de documentation, la base d'archive, l'accès au document et leur réutilisation, les ressources d'indexation. Nous finissons par mentionner notre utilisation réelle des ressources. Nous considérons l'INA comme l'acteur le plus avancé au niveau des efforts de numérisation et richesse des informations en ligne.

La sélection des vidéos de cette fresque est faite par l'INA à partir d'un plus gros corpus vidéo et selon une approche grand public. Des notices descriptives et des métadonnées accompagnent chaque vidéo : les fichiers vidéos sont initialement accompagnés de documents textes qui servent à l'élaboration des notices et contiennent des informations telles que les conducteurs d'émission, le temps (date de diffusion, heure de diffusion, durée), les thématiques etc. Ces notices sont adaptées et standardisées au cadre plus global de l'INA. Les notices réalisées contiennent le résumé des séquences, des mots-clés, le nom des intervenants, le temps de diffusion ou de rediffusion. Toutes les ressources ne sont en revanche pas indexées. Au niveau du type de document audiovisuel, des vidéos de télévision comme les émissions « *Magazine du mineur* », des enregistrements radio et du web sont disponibles. Au niveau technique, le logiciel *Totem* est utilisé comme base d'archive patrimoniale. L'application est développée par les

[84] L'INA Nord est créé en 1977 à Lille.
[85] Fresque chronologique vidéo sur l'histoire de la mine et des mineurs du Nord et du Pas de Calais (consultable à l'adresse http://fresques.ina.fr/memoires-de-mines/)

équipes de l'INA pour les besoins spécifiques des archives et réunit à la fois la production de données et la recherche documentaire. *Totem* pilote une base *Oracle* associée à un moteur de recherche *Exalead*.[86] Concernant la consultation, deux postes sont disponibles pour un accès à l'ensemble des fonds TV, radio, web médias, collections partenaires et sources écrites et un accès aux outils de consultation et d'analyse. Les médias de la vidéothèque sont disponibles en consultation sur rendez-vous. Concernant la réutilisation des médias, étant soumis au dépôt légal, ils ne peuvent être utilisés en *open source* en raison de la question des droits d'auteur mais peuvent, s'ils sont disponibles sur internet, être référencés par un lien vers le site de l'INA. Concernant les ressources d'indexation, aucune ressource n'est récupérable et la granularité d'indexation est trop élevée. Concernant d'autres acteurs du domaine, Jean Carrive mentionne l'importance des ressources documentaires et filmiques provenant des Houillères du NPC disponible au CHM de Lewarde.[87] Nous avons finalement utilisé la fresque en ligne pour comprendre et vérifier l'utilisation de certains termes dans les documents audiovisuels (figure 25) et nous nous sommes appuyée sur l'organisation thématique des documents pour structurer notre connaissance du domaine (figure 26).

Figure 25 : Utilisation du moteur de recherche pour identifier des termes dans les notices descriptives des vidéos de la fresque Mémoires de Mines, ici requête du terme « lavage » présent dans quatre médias.

[86] Données issues de : http://arcmc-corpus.huma-num.fr/sites/default/files/atoms/files/arcmc_ddcol_20160503_vf.pdf
[87] Depuis 1996, le Centre Historique Minier conserve l'ensemble de la production audiovisuelle des Houillères du Bassin du Nord-Pas-de-Calais. 501 films en 16 et 35 mm et 300 enregistrements sonores constituent l'important fonds de la cinémathèque. (http://www.chm-lewarde.com/fr/ressources-et-collections/la-cinematheque/)

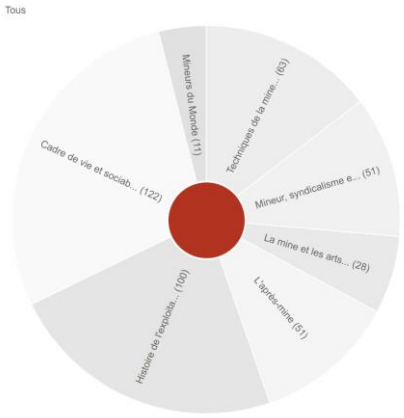


Figure 26 : Organisation thématique des documents vidéo de la fresque https://fresques.ina.fr/memoires-de-mines/pie

5.1.1.2 Centre de documentation : CHM de Lewarde

Suite à un premier entretien avec Madame Frédérique Delforge le 9 janvier 2018, assistante de documentation au CHM, nous avons rencontré Virginie Malolepszy, directrice des archives au CHM le 15 novembre 2018 à Lewarde afin de parler des ressources disponibles ainsi que de leur organisation. Nous revenons sur neuf points évoqués lors de ces rencontres : l'organisation de la documentation, l'état du classement, la ressource classificatoire d'un fond scientifique, la numérisation des ressources, la potentielle utilisation d'un thésaurus, le contact vers d'autres acteurs détenteurs d'archives, le logiciel d'archives, le catalogue d'accès, les ressources audiovisuelles et les objets.

Dès les années 1970, le CHM a récupéré des documents d'archives. Les HBNPC (qui n'étaient pas des professionnels de la documentation) ont créé un document pour pouvoir retrouver tous les documents sauvegardés mais aussi les objets. Présenté comme *cadre de classement* par Madame Delforge, le document est plutôt un *plan de recherche* pour Madame Malolepszy (figure 27). Pour rappel, un cadre de classement est défini « dans un fonds ou un service d'archives, [comme un] schéma rationnel prédéterminé servant de guide à la mise en ordre et souvent à la cotation des documents » (Boulogne et Institut national des techniques de la documentation 2004, 44). Nous pensons que l'utilisation de *plan de recherche* vient du fait que le document est utilisé pour aider à la recherche des documents tandis que le terme « classement » désigne la numérotation utilisée sur les fiches signalétiques des documents avec l'identification du titre des pages, ainsi FOND est identifiée par DF, JOUR par DJ, ORGANISATION par OP, DIVERS par DD.

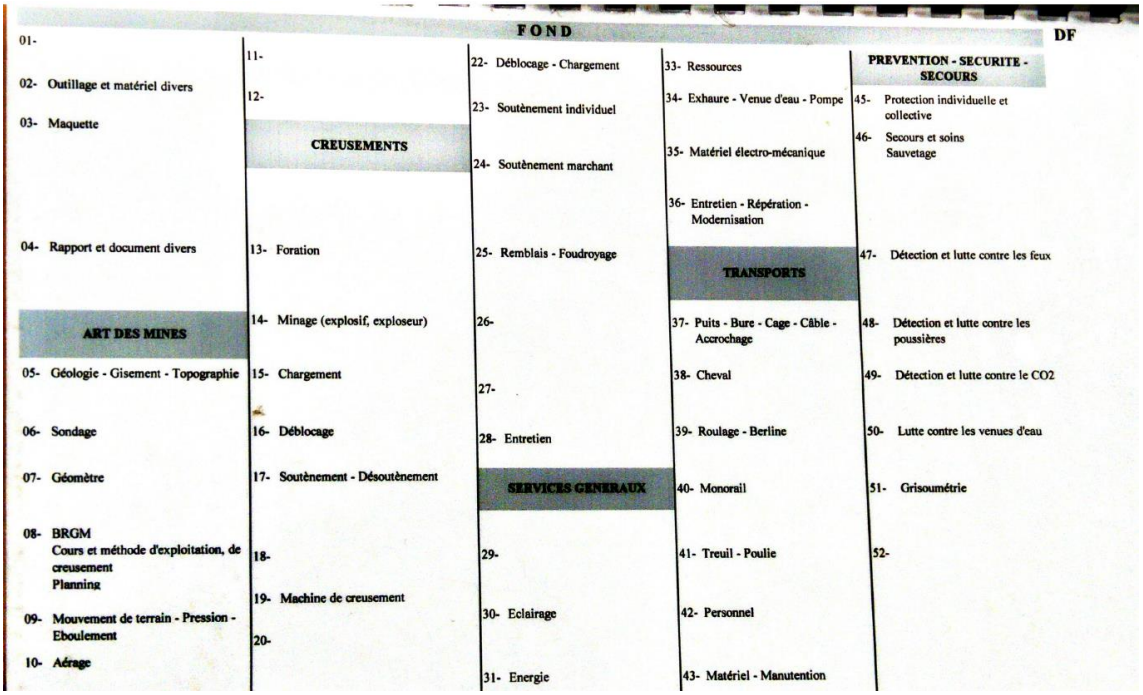


Figure 27 : Première page du plan de recherche des archives du CHM

Le premier plan de recherche possède donc quatre pages : Fond, Jour, Organisation, et Divers. Sur chacune de ces pages, des mots-clés ou thématiques sont numérotées. Certains numéros n'ont pas de thématique (Cf. 29, 20, 18 etc. sur figure 27) pour une raison qui reste inconnue. Le deuxième plan de recherche liste par groupes les compagnies et concessions (tableau 9).

Tableau 6 : Plan de recherche du CHM par groupes, compagnies et concessions

GROUPES	COMPAGNIES	CONCESSIONS	
AUCHEL			1
Le groupe d'Auchel fusionne avec Bruay en 1961	LIGNY-LES-AIRES *(1894)*	FLECHINELLE *(Cie de la Lys Superieur 29 déc 1855-31 août*	2
			3
		AUCHY-AU-BOIS	4
	LA CLARENCE (De Wendel) *(13 août 1895)*		5
		CAMBLAIN-CHATELAIN *(mai 1859)*	6
	FERFAY *(29 déc 1855 - 31 mars 1870)*		7
		CAUCHY-A-LA-TOUR *(31 mars 1870)*	8
	MARLES-LES-MINES *(absorbe la Cie de Ferfay-Gauchy le 1er juil 1925)*	MARLES *(29 déc 1855)*	9
BETHUNE			10
	BETHUNE		11

Le premier plan permet donc de retrouver des fiches signalétiques (papier) (ou notice bibliographique) sur des ouvrages et des archives (papiers). Ces fiches contiennent la référence bibliographique de l'ouvrage ou de l'archive et une côte numérique. Sur la figure 34, *9468* est la côte numérique. *DD12* correspond au plan de recherche page Divers, numéro 12 et ceci correspond au thème « vocabulaire minier ». L'ajout au crayon de DD11 sous la référence indique que l'ouvrage appartient également au thème « Écrit sur la mine et sur la Région, Bibliographie – Littérature – Coutume – Chant ».

DD 12

9468

TIT LOUIS alias BALLE (André). - Le Patois du Pas-de-Calais : Traduction et anecdotes / Tit Louis ; Pierre Coolos, ill. - (S.l.) : (S.n), 1980 : (Calais : Offset Service). - 155 p. : ill. ; 21 cm.

DD11

Figure 32 : Exemple de fiche signalétique du CHM appartenant à la thématique « vocabulaire minier »

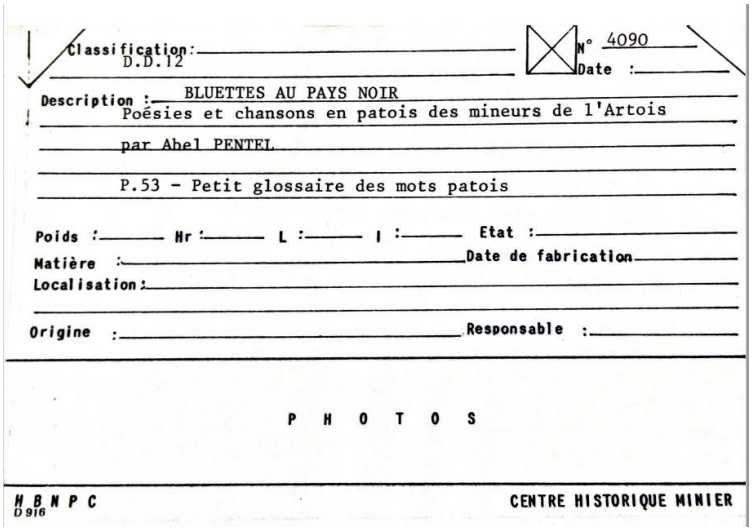


Figure 33 : Exemple de fiche signalétique du CHM appartenant à la thématique « vocabulaire minier »

Tous les fonds sont classés ou en instance de classement dans le cas des dons récents. Le CHM conserve également les fonds du CERCHAR (Centre d'étude et de recherches des charbonnages de France) pour les HBNPC et l'accès s'y fait grâce à une classification. La classification présentée par la directrice est une version papier de juillet 1962. Une version plus récente (1976) de cette classification nous a été transmise lors d'échanges avec le musée de Blegny-mine. Nous verrons ci-après comment nous l'avons utilisée. Concernant la numérisation, VM a son opinion. « *Numériser tout, à quoi ça sert ? Ça coûte cher. Quelles sont les garanties pour la sauvegarde ? Un chercheur d'Amiens qui a besoin d'un document nous appelle et si on l'a, il vient au centre* [...] ». Il n'y a pas de campagne pour numériser massivement et cela est fait au cas par cas, pour des expositions, des publications ou des demandes de l'extérieur. Il n'y a pas de thésaurus et l'indexation par mot-matière (ou vedettes-matière) serait, d'après Madame Malolepszy, plus pertinente. Nous apprenons que Les Archives du Monde du Travail (ANMT) à Roubaix possèdent également des fonds et que le fonctionnement diffère des archives de Lewarde. Au niveau des archives, le CHM utilise le logiciel Avenio, qui n'est alors pas encore utilisé dans sa globalité. La recherche s'effectue par mot-clé. Il n'y a pas de catalogue (pour les utilisateurs). Il y a cependant des bases de données où tous les ouvrages sont regroupés (sous Excel) et sont utilisés par les professionnels. Les données sont transférées sur le logiciel Avenio. Ainsi, si un chercheur veut trouver le lexique minier le plus récent, il doit soit passer par les professionnels pour accéder à la ressource, soit rechercher dans

les fiches. Quant aux films, ils sont inventoriés dans un gros livre relié. « 501 films et 300 enregistrements sonores constituent l'important fonds de la cinémathèque »[88]. Nous avons, pour notre part, eu accès à un fichier Excel inventoriant 4 018 photographies, avec une courte description du sujet de la photographie, le lieu et la date. Par exemple : « Mineur en taille, treuil, fosse 2 d'Oignies, 20/10/1975, Carpentier » ou « carte de vœux de Sainte-Barbe » ou « Photographie d'identité d'ouvriers marocains arrivés au centre de Noyelles sous Lens, 12/01/1976, Druelle » (figure 28). Nous avons pu relever des entités tels que des noms de fosses, de tailles, de sièges, des dates et des noms de villes) ainsi que les termes utilisés pour décrire celles-ci.

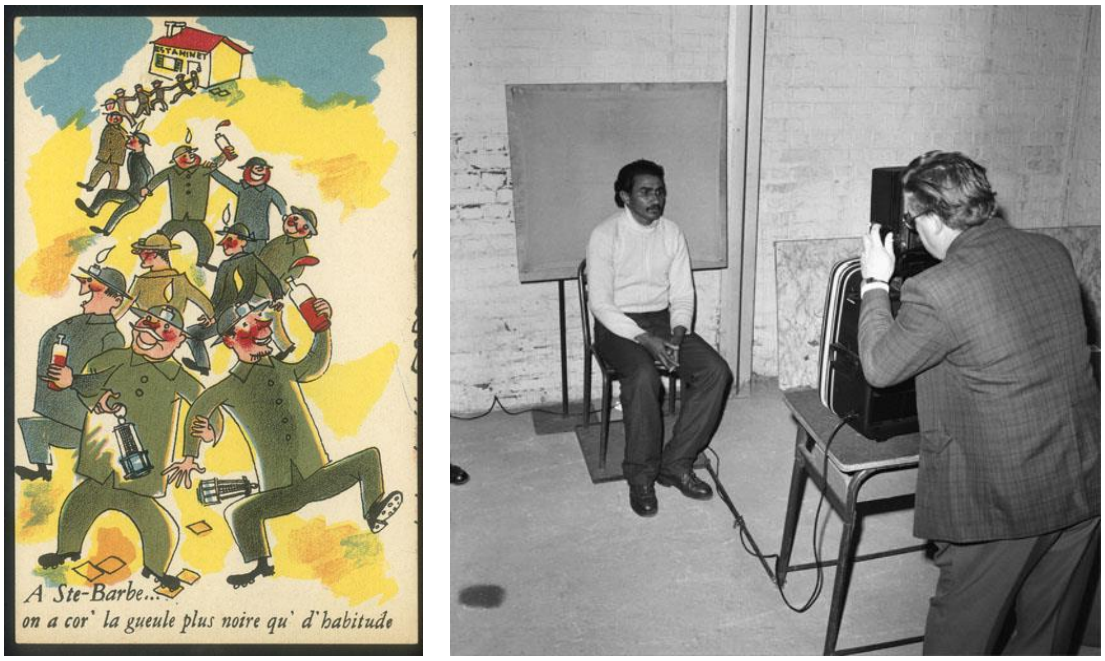


Figure 28 : À gauche, carte de vœux de Sainte-Barbe. Image CHM. À gauche, Photographie d'identité d'ouvriers marocains arrivés au centre de Noyelles sous Lens, 12/01/1976, Druelle. Image CHM.

Les objets comportent quant à eux une fiche d'analyse et un numéro d'inventaire.

Nous nous intéressons plus particulièrement à la catégorie *DD12* du premier plan de recherche qui correspond à la thématique « vocabulaire minier ». Les ouvrages et éléments d'archives sont signalisés sur un ensemble de fiches papier. 74 fiches correspondent à cette thématique. Les références bibliographiques sont présentées sur deux types de fiche dont certaines ont une structure prédéfinie et dactylographiées, d'autres imprimées ou encore d'autres écrites à la main. Nous verrons en (5.1.2) quelle sélection nous avons fait parmi ces ressources et comment nous les avons analysées.

5.1.1.3 Organisme scientifique : BRGM Hauts-de-France

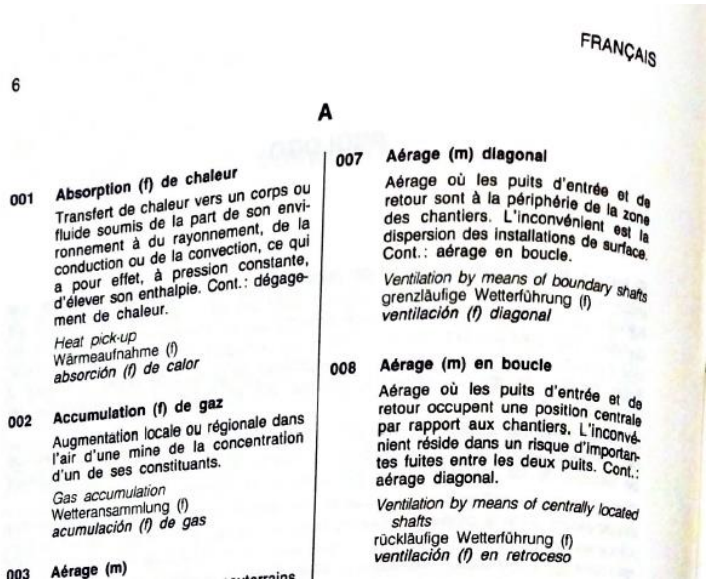
Le 5 avril 2018, nous avons rencontré Philippe Andrzejewski, responsable des archives au Bureau de Recherches Géologiques et Minières (BRGM) Hauts-de-France, dans les locaux de l'UTAM (Unité Territoriale Après-Mine) Nord de Billy-Montigny. Lors de cette rencontre, des informations sur la tenue des archives ont été données et nous avons notamment eu accès aux archives papier de l'ancien

[88] https://www.chm-lewarde.com/fr/ressources-et-collections/la-cinematheque/

exploitant minier Charbonnages de France (CdF) constituées des dossiers de déclaration d'arrêt des travaux (DADT) relatifs aux anciennes exploitations et installations minières de Charbonnages de France, ainsi qu'une notice sur la normalisation des plans de détail par veine à l'échelle 1/1000 (Avril 1948). Jusqu'en 2022, une convention entre le BRGM, l'état et les charbonnages est en place. Les archives sont soumises au « contrôle scientifique et technique » du directeur des Archives départementales. Au niveau des ressources documentaires, le BRGM possède deux kilomètres 300 d'archives linéaires. L'ancien service d'archives n'étant pas structuré, des dossiers sont parfois reproduits en plusieurs fois et une restructuration de l'information est effectuée pour l'ensemble des agents du BRGM, avec un tri et un dédoublonnage du matériel de conservation (ex : photographies, inventaire classique) selon les normes archivistiques et selon des thématiques. À cette époque, un contractuel en archives de l'université de Lille, des alternants et des stagiaires avaient cette mission. Parmi les documents se trouvent des plans d'exploitations dont 50 000 cartographies dessinées avec 38 typologies de plan (les géomètres des mines étudient en effet le souterrain à différentes échelles). Un travail sur trois ans a été mené avec une équipe de géologues, de dessinateurs projeteurs (pour considérer l'intérêt dans le futur). En 2013, les fonds ont été analysés par le Service interministériel des archives de France (SIAF) et le sort final des archives a été questionné. Il en est ressorti que le service public territorial était compétent. Des enjeux territoriaux et environnementaux sont posés et notamment au niveau hydrodynamique : après trois cents ans, des remontées d'eau ou de gaz ont été constatées sur le territoire. Certains documents n'ont jamais été numérisés et font l'objet de surveillance pérenne. La question de la place que prennent les archives est également posée. Au niveau de l'héritage, ce sont des archives techniques qui sont utiles pour la surveillance et pour les travaux sur le bassin minier. Les dossiers ont été numérisés et OCRisés puis transmis en mairie (pour chaque commune). Une base de données faite « maison », Auressia (Archives Techniques Intermédiaires Minières) va être améliorée pour répondre aux normes[89]. En 2006, une interface de saisie a été mise en place avec la possibilité de requêter. Le domaine des archives est plutôt géographique avec le périmètre minier des concessions par région (recherche libre par mot-clé). L'ambition est d'avoir sur un serveur commun les documents, la vidéothèque, la photothèque. Concernant les sources scientifiques, il existe un portail HAL spécialisé BRGM[90]. Lors de cette rencontre, Monsieur Andrzejewski nous a conseillé de contacter Jean-Marie Minot, passionné de la mine et travaillant avec des anciens mineurs de la fosse 9 d'Oignies, ainsi que Guy Dubois, un auteur patoisant du Nord. Il nous a également conseillé d'aller aux archives départementales du Nord[91]. Nous avons donc par la suite rencontré Jean-Marie Minot et identifié quelques écrits de Guy Dubois sur le parler ch'ti et les mines (Dubois 1981, 2019, 1988, 1993 ; Dubois et Minot 1991, 1992). Au BRGM, nous avons également identifié quelques ressources lexicales dont deux multilingues et techniques : (Chorosz et Société de l'industrie minérale 1987) (figure 29) et

[89] La structuration de la base de données en 2011 est disponible dans ce document en ligne : https://infoterre.brgm.fr/rapports/RP-61186-FR.pdf
[90] https://hal-brgm.archives-ouvertes.fr/
[91] https://archivesdepartementales.lenord.fr/

(Chalon 1909) et le *Thésaurus de l'exploitation minière – souterraine et à ciel ouvert* – de Florence Auclaire (Auclaire 1990) décisif pour notre approche et jouant en partie le rôle de thésaurus source.



FRANÇAIS

6

A

001 **Absorption (f) de chaleur**
Transfert de chaleur vers un corps ou fluide soumis de la part de son environnement à du rayonnement, de la conduction ou de la convection, ce qui a pour effet, à pression constante, d'élever son enthalpie. Cont.: dégagement de chaleur.

Heat pick-up
Wärmeaufnahme (f)
absorción (f) de calor

002 **Accumulation (f) de gaz**
Augmentation locale ou régionale dans l'air d'une mine de la concentration d'un de ses constituants.

Gas accumulation
Wetteransammlung (f)
acumulación (f) de gas

003 **Aérage (m)**

007 **Aérage (m) diagonal**
Aérage où les puits d'entrée et de retour sont à la périphérie de la zone des chantiers. L'inconvénient est la dispersion des installations de surface. Cont.: aérage en boucle.

Ventilation by means of boundary shafts
grenzläufige Wetterführung (f)
ventilación (f) diagonal

008 **Aérage (m) en boucle**
Aérage où les puits d'entrée et de retour occupent une position centrale par rapport aux chantiers. L'inconvénient réside dans un risque d'importantes fuites entre les deux puits. Cont.: aérage diagonal.

Ventilation by means of centrally located shafts
rückläufige Wetterführung (f)
ventilación (f) en retroceso

Figure 29 : Exemples de termes contenu dans Chorosz et Société de l'industrie minérale (1987)

5.1.1.4 Organisme associatif : Mission Bassin minier (MBM)

La Mission Bassin minier est un outil d'ingénierie, de développement et d'aménagement du territoire. Le 26 février 2019, nous avons rencontré Virginie Lapierre responsable de communication à la Mission Bassin minier, Vianney Haeuw, chargé de Mission Information au CAUE (Conseil d'Architecture, d'Urbanisme et de l'Environnement**)** et Christophe Grandjacques, architecte-urbaniste, délégué territorial Douaisis-Valenciennois chez CAUE du Nord. Lors de cette rencontre, nous avons appris que la Mission Bassin minier ne possède pas de documents concernant le mémoriel mais plutôt sur l'urbanisme et l'aménagement du territoire. La Mission Bassin minier a été créée le 12 mai 2000 et est issue d'une conférence permanente en 1996 qui a donné lieu au livre Blanc (premier acte) publié en novembre 1997 (Conférence permanente du bassin minier 1997). Elle s'installe en 2002 sur le site du 9-9bis à Oignies (figure 30).

Figure 30 : Locaux de la Mission Bassin minier sur le site du 9-9bis à Oignies

L'activité minière a provoqué des séquelles au niveau environnemental et a forcé le développement de projets d'urbanisme. Un travail de recensement (des cités minières, des fosses, des terrils…) a été effectué en lien avec Bassin Minier Uni (BMU)[92] en 2012 pour la candidature à l'UNESCO. BMU et MBM ont fusionné en janvier 2013 pour devenir un outil de promotion et de médiation (local et international). 80% des activités tournent autour de l'inscription UNESCO et MBM gère 353 éléments. De ces 353 éléments, une dizaine est en danger et une souscription a été envoyée. MBM participe par exemple à l'organisation d'événements telle que l'inauguration de la cité des électriciens qui a eu lieu les 18 et 19 mai 2019[93]. L'objectif serait de fusionner les ressources MBM et BM patrimoine mondial. Lors de cette rencontre, Virginie Lapierre nous a donné une vingtaine de brochures et publications de la MBM traitant des problématiques d'aménagement du territoire et du patrimoine.

Concernant la CAUE (société civile immobilière), une plateforme numérique permet la diffusion de ses ressources[94] et les fonds documentaires sont accessibles grâce à un thésaurus (Architecture, Urbanisme, Environnement, actuellement à sa 6ème édition[95]). Une partie des ressources est disponible localement (livres, audio) et une partie en observatoire (éléments qui forgent l'identité du territoire). Il y a 3000 observations/fiches. Le centre de ressources numériques n'est pas encore au W3C et il n'y a pas de documentaliste. Sur la plateforme numérique, nous trouvons des publications, des logos et des études. Si les ressources ne concernent pas directement notre domaine, l'expertise documentaire de Vianney Haeuw (qui a participé à la réalisation du thésaurus de la CAUE) a tout de même été mobilisée pour l'évaluation du thésaurus.

5.1.1.5 Bibliothèques de recherche physiques ou numériques

Trois bibliothèques de recherche sont jugées pertinentes pour notre domaine d'étude : George Lefebvre, IRIS et Nordnum.

La bibliothèque George Lefebvre est une bibliothèque de recherche de l'Institut de Recherches Historiques du Septentrion (IRHiS)[96] spécialisée en histoire régionale. Concernant le domaine « MINES », la bibliothèque recense 77 documents[97] (tableau 7). Ces documents sont tous disponibles à partir du catalogue du SUDOC et sont donc indexés avec le langage RAMEAU. Par exemple, l'ouvrage n°7 dans le Tableau 7 traite de la mine générale et a pour indexation : Mines (sites d'extraction) – Constructions. Ceci permet d'identifier un certain nombre de vedettes-matières plus ou moins génériques qui indexent le domaine (annexe 5) et de repérer d'autres documents dans les catalogues les utilisant comme celui de la bibliothèque IRIS[98], bibliothèque numérique dédiée au passé des sciences modernes que nous présentons ci-dessous.

[92] http://www.bassinminier-patrimoinemondial.org/bassin-minier-uni/
[93] Mentionné dans le dossier de presse disponible en ligne via ce lien : https://citedeselectriciens.fr/sites/default/files/2020-01/DOSSIER%20DE%20PRESSE%20Cite%CC%81%20des%20e%CC%81lectriciens%20MAJ230120.pdf
[94] https://www.ressources-caue.fr/
[95] Information renseignée par Vianney Haeuw le 15 avril 2021
[96] UMR 8529 CNRS-Université de Lille Sciences Humaines et Sociales
[97] Nous avons obtenu cette liste auprès d'un personnel de la bibliothèque.
[98] IRIS n'est pas un sigle.

Tableau 7 : Extrait de la liste des ouvrages sur la thématique MINES bibliothèque G. Lefebvre Université de Lille

MINES – SI 12			
1	DAMIEN René	Acier français 1953 1953	FSI 202
2	CALMETTE A. BRETON M.	L'ankylostomiase : maladie sociale (anémie des mineurs) : biologie, clinique, traitement, prophylaxie 1905	FSI 203
3	LEVEQUE J.	Les ardoisières du bassin de Fumay : description de leur exploitation et de la fabrication de l'ardoise 1905	FSI 204
4	MOLINIE M. DIETZ H.	L'argent et les métaux de la mine de platine 1912	FSI 205
5	DROUX Henri	Les câbles de mines 1884	FSI 206
6	BLEIBTREU H.	Le charbon pulvérisé, le poussier de charbon et leurs applications 1932	FSI 207
7	LECOMTE-DENIS M.	Comment on crée une mine 1913	FSI 208

L'analyse des titres de ceux-ci permet d'identifier 17 documents traitant spécifiquement des mines de charbon. D'autre part, le domaine MINES ne prend pas en compte les ouvrages sur le vocabulaire minier tel que (B. Turpin et Goudaillier 2004) ou (Lebon 1976) pourtant disponibles dans cette même bibliothèque, ni un certain nombre d'ouvrages de littérature patoisante. La bibliothèque possède en effet quelques recueils de poésie de Jules Mousseron, célèbre poète mineur écrivant en rouchi, le parler picard de Valenciennes (Mousseron et Jonas 1933). Au niveau de l'accès, la plupart de ces ressources sont en magasin et ne peuvent malheureusement pas être prêtées. Dans tous les cas, les ouvrages ne peuvent être empruntées et doivent être consultés sur place. Seules quelques numérisations ont donc pu être effectuées et les ouvrages pertinents ont pu être identifiés dans d'autres bibliothèques.

La bibliothèque IRIS est une bibliothèque numérique spécialisée en histoire des sciences. IRIS possède des entrées thématiques dont l'une s'intitule « Patrimoine scientifique, technique et industriel – Nord de la France et Wallonie ». Sous cette entrée le classement se décompose en six catégories, dont la catégorie *Le bassin minier du Nord–Pas-de-Calais* qui comporte 37 documents provenant de différents sites tels que Lilliad (Université de Lille 1), le Musée d'histoire naturelle de Lille, la Bibliothèque recherche UMR 8198 Evo-Eco-Paléo CNRS de l'Université Lille 1, le Centre d'histoire des sciences et des techniques (CHST) de l'Université de Liège et l'Institut Industriel du Nord (IDN) et la Centrale Lille.

Enfin la bibliothèque Nordnum est une bibliothèque numérique d'histoire régionale du Nord et du Pas-de-Calais. Une recherche par sujet « Mines » permet de retrouver 23 documents. Contrairement aux documents d'IRIS, ceux-ci ne sont pas téléchargeables et ils ne sont pas océrisés. La bibliothèque possède néanmoins une fonctionnalité que nous trouvons intéressante au niveau de la visualisation des données qui est une frise chronologique (figure 31). Si les documents ne peuvent être téléchargés, cette visualisation permet de se représenter une partie de l'histoire des mines et de placer des événements dans le temps.

Figure 31 : Frise temporelle des documents correspondant au sujet « Mines » dans Nordnum

5.1.2 Identification des ressources lexicales pour définir le domaine

5.1.2.1 Création du corpus de référence

Le corpus lexical de référence (Rastier 2005) est composé des 74 ressources papiers identifiées lors de la consultation des fonds du CHM (rangées sous la thématique DD12 « vocabulaire minier »). Ces premières ressources ont pu servir à l'identification de nouvelles ressources en ligne. En effet, grâce aux moteurs de recherche, nous avons testé les termes des entrées identifiées dans les ressources puis nous les avons combinés (ou non) avec des mots-clés désignant le type de document : lexique, glossaire, dictionnaire, vocabulaire ou terminologie et des mots-clés désignant le domaine : mine, minier, houiller. Un total de 44 autres ressources lexicales a pu être identifié dans des catalogues de bibliothèques, des sites Web, des blogs et des wikis.

Les ressources lexicales du corpus de référence ont tout d'abord été enregistrées au regard de plusieurs critères. Un identifiant a été attribué à chaque ressource, puis nous avons relevé : le format de la ressource (numérique ou papier), le volume (nombre de pages ou le cas échéant de termes), la langue (ou le dialecte), la zone géographique (couverture géographique), le domaine d'application (mine générale ou de charbon), le lien internet (si ressource numérique), le type de ressource (glossaire, lexique, dictionnaire, vocabulaire, terminologie), le type de langage (langage courant ou terminologie de spécialité), un commentaire ou une description sur l'œuvre, la pertinence (évaluée selon des critères définis), la date (publication ou création pour le site), le ou les auteur(s) ou concepteur(s), une note sur l'auteur (activité ou type), la référence bibliographique, la provenance, la structure, l'organisme titulaire de la ressource, les coordonnées (adresse, téléphone, courriel), les sources (utilisées pour l'élaboration de la ressource) et la localisation (pour la numérisation, en bibliothèque ou ailleurs).

Ce premier travail nous a permis de créer plusieurs typologies des ressources en fonction de ces critères. Nous avons jugé intéressant de les regrouper par statut de leur « auteur » :

- Experts en terminologie (spécialistes de la description linguistique de la/des langues étudiée(s)) : les dictionnaires de Turpin et Goudaillier (2004, 872 entrées), de Dauby, spécialiste du rouchi (1979, 417 pages), le lexique d'Auger et Rousseau (1973, plus de 735 entrées),

- Anciens mineurs (experts de la pratique des langues étudiées) : les glossaires de Mousseron, poète patoisant écrivant sur la mine (1994, 640 entrées) et de Mongaudon (1988, 176 entrées), le lexique de Lateur, poète picard et Roques (préfacier) (1951, plus de 6000 entrées) ou le vocabulaire technique des mineurs du Nord et du Pas-de-Calais de Jean Bovio, ingénieur des mines (1906, 451 entrées),

- Passionnés actifs (spécialistes dans le domaine étudié par leur connaissance sur le domaine) :
    - créateurs de blog : André de Marles (147 entrées)[99], *Au pays des chtis* (37 entrées)[100]
    - les associations de sauvegarde du patrimoine : le site *Mineur de fond*[101] créé par l'APPHIM (Association pour la Pérennisation du Patrimoine Historique Industriel et Minier, 141 entrées) ou le site *Chaîne des terrils*[102] de l'association du même nom (54 entrées), le glossaire du Haut-Fourneau de l'association *Le savoir ... fer* (Corbion 2016, 82121 entrées dont un nombre indéfini d'entrées correspondant au vocabulaire minier),
    - les proches des mineurs : le site *Les Gueules Noires* créé par un fils et un petit fils de mineur polonais immigré à Liévin (136 entrées)[103], le site *Le Bassin Minier Nord Pas-de-Calais* créé par un arrière-petit-fils de mineur et descendant d'une grande famille de mineurs (29 entrées)[104].

Ils peuvent également être regroupés en fonction du type de ressource lexicale. Nous présentons ci-dessous (tableau 8) les critères de distinction des ressources lexicales à partir de l'étude des définitions de chaque type de ressources retirées du TLFi.

Tableau 8 : Critères de distinction des ressources lexicales

Critères de distinction des ressources lexicales		**Ressources lexicale**
Unités linguistiques	-mots	dictionnaire, vocabulaire, lexique, glossaire
	-termes	terminologie
Unités linguistiques représentent	-langue	dictionnaire, glossaire
	-langue spéciale, technique	vocabulaire
	-une œuvre	glossaire
	-un domaine	dictionnaire
	-un dialecte, un patois	glossaire
Type de locuteur	-un groupe social, professionnel	lexique
Utilisation des unités linguistique propre à	-technique, science, corporation	lexique
Contenu	-définitions, explications, références	glossaire
	-informations relatives au sens et à l'emploi	dictionnaire
Organisation des unités linguistiques	-alphabétique	dictionnaire
	-système notionnel	terminologie

[99] Site : http://andredemarles.skyrock.com/ Lexique à : http://andredemarles.skyrock.com/3113954741-Lexique-des-termes-miniers-1ere-partie-de-A-a-D.html
[100] Site : http://www.aupaysdeschtis.com Lexique à : http://www.aupaysdeschtis.com/le-lexique-de-la-mine/
[101] Site : http://www.mineurdefond.fr/ Lexique à : http://www.mineurdefond.fr/articles.php?lng=fr&pg=671&mnuid=662&tconfig=0
[102] Site : http://www.chainedesterrils.eu/ Lexique à : https://web.archive.org/web/20070801023056/http://chaine.des.terrils.free.fr/lexique.htm
[103] Site : https://centriris.fr/ La page sur le vocabulaire, auparavant : http://centriris.fr/vocabulaire/page19.php n'existe plus et a été remplacée par : https://centriris.fr/les-termes-miniers Les termes et leurs définitions n'ont en revanche pas changé.
[104] Site : https://bassin-minier-regis.jimdofree.com/
Lexique à : https://bassin-minier-regis.jimdofree.com/un-peu-d-histoire/les-m%C3%A9tiers-de-la-mine/

Ainsi, nous avons par exemple :

- Le dictionnaire du TLFi pour lequel il est possible de sélectionner la discipline « industrie : mines » (258 termes).
- Le glossaire des principales expressions patoises employées dans les poésies de Jules Mousseron (Mousseron 1994), le glossaire du PPRM Plan de prévention des risques miniers (Direction Départementale Des Territoires Et De La Mer Du Nord 2017) ou encore le *Glossary of mining terminology* des Affaire indiennes et du Nord Canada[105]
- Lexique : *Le parler des mineurs d'Artois. — Lexique du parler populaire et ouvrier des Régions Mineures d'Artois* de Lateur Marius (Lateur et Roques 1951) ou *Les mots de la mine : lexique dialectal de la profession de mineur au début du 20ème siècle dans la région de Denain* d'André Lebon (Lebon 1976)
- Le vocabulaire professionnel du houilleur borain de Ruelle et Jadon (Ruelle et Jadon 1953) également appelé glossaire, le *vocabulaire technique des mineurs du Nord et du Pas-de-Calais* (Bovio 1906), le vocabulaire de l'atmosphère minière (Chorosz et Société de l'industrie minérale 1987)
- La terminologie d'Auger et Rousseau intitulée *Lexique Anglais-Francais de l'industrie minière* (Auger, Rousseau, et Québec. Office de la langue française 1973)
- Et nous avons également un abécédaire : *Les mots de la mine: petit abécédaire pour mineur du XXIe siècle* de Benneteu et *al.* (Benneteu et *al.* 2010)

5.1.2.2 Création du corpus d'étude

Face à l'hétérogénéité des ressources lexicales, il est nécessaire de choisir des critères pour restreindre le corpus de référence et analyser le corpus d'étude sur une base commune. Le premier critère de sélection est celui de la langue. Les entrées[106] doivent être en français ou en dialecte parlé sur le territoire de l'exploitation des mines de charbon du Nord–Pas-de-Calais. Enfin, les termes doivent être définis afin que nous puissions travailler à partir de leurs définitions.

Parmi les 74 ressources du corpus de référence, les critères mentionnés ci-dessus ont permis d'éliminer 54 ressources, dont certaines en anglais, d'autres portant sur le vocabulaire d'autres bassins houillers français ou étrangers ou encore sur d'autres industries minières telles que les mines de potasse en Alsace (Weissenberger et Tugler 1984). Nous avons également dû supprimer les ressources difficiles à numériser et océriser. L'analyse comparative porte finalement sur un sous-corpus de travail composé de 20 ressources correspondant aux ressources numériques, numérisées et/ou océrisées du corpus d'étude.

[105] https://www.rcaanc-cirnac.gc.ca/eng/1100100028056/1614019384552
[106] Nous utilisons le terme « entrée » pour désigner le mot mis en vedette au début de chaque article d'un dictionnaire.

Le corpus d'étude ainsi réalisé est analysé selon cinq critères permettant de sélectionner et de définir la nature des termes et de leur relation dans le cadre de l'élaboration d'une terminologie pour un thésaurus. Cette analyse permet également d'étudier les particularités de la terminologie minière et de ses ressources relatives.

5.1.3 Identification des ressources classificatoires pertinentes

Dans cette section, nous présentons une typologie des outils classificatoires du domaine minier identifiés dans le cadre de cette recherche dans leur contexte de création. Ils constituent le matériau à partir duquel nous avons conçu le thésaurus. Il s'agit de systèmes d'organisation des connaissances de nature extrêmement diverses : plans ou cadres de classement, classifications, index thématiques, terminologies etc. Ces SOC correspondent aux SOC épistémiques généralistes et aux SOC vernaculaires de Zacklad (Zacklad 2018).

5.1.3.1 Plans et cadres de classement

Un plan de classement est « un document qui présente une classification de manière ordonnée, en faisant apparaître la signification donnée à chaque indice » (Boulogne et Institut national des techniques de la documentation 2004, 190) tandis qu'un cadre de classement est un « schéma rationnel prédéterminé servant de guide à la mise en ordre et souvent à la cotations des documents » (Boulogne et Institut national des techniques de la documentation 2004, 44).

En 1907, suite à la plus grande catastrophe minière connue, celle de Courrières qui, le 10 mars 1906, fit plus de 1 100 victimes, est créé l'ancêtre du Centre d'Études et Recherches des Charbonnages de France (CERCHAR)(M. Turpin 1988, 112), qui prendra ce nom en 1947. Le groupement de la documentation des Industries Extractives oblige le CERCHAR à créer un plan de classement (tableau 9), révisé en 1973 puis en 1976 (CERCHAR 1976). En décembre 1990, peu avant la fermeture de la dernière mine dans le Nord–Pas-de-Calais, il deviendra l'Institut National de l'Environnement Industriel et des Risques (INERIS) dont la mission sera d'évaluer et prévenir les risques accidentels ou chroniques pour l'homme et l'environnement liés aux installations industrielles, aux substances chimiques et aux exploitations souterraines. Le plan de classement est accessible au Centre Historique Minier (CHM) de Lewarde, qui conserve les archives du CERCHAR associées au Bassin minier du Nord–Pas-de-Calais. En 1965, l'Institut National de l'Industrie Charbonnière en Belgique (INICHAR), homologue belge du CERCHAR, publie une nouvelle édition de sa documentation par fiches avec un tableau détaillé des index (tableau 10), dans lequel il est écrit que « les progrès de la science et de la technique exigent avant tout une forte documentation bien classée » (INICHAR 1965, 1).

Tableau 9 : Domaines du plan de classement du CERCHAR, édition de 1976

A - GEOLOGIE - GISEMENTS - PROSPECTION - SONDAGES
B - AMENAGEMENT - METHODES D'EXPLOITATION
C - ABATTAGE - CHARGEMENT
D - PRESSION DE TERRAINS - SOUTENEMENT
E - TRANSPORTS - EXTRACTION
F - AERAGE - GRISOU - POUSSIERES - FEUX - ECLAIRAGE
G - EPUISEMENT
H - ENERGIE - AIR COMPRIME - VAPEUR - ELECTRICITE
I - CONSTITUTION - PREPARATION - AGGLOMERATION DES CHARBONS
K - CARBONISATION
L - GAZEIFICATION
M - COMBUSTION - CHAUFFAGE
N - NUISANCES
O - SYNTHESES ET INDUSTRIES CHIMIQUES
P - MAIN-D'ŒUVRE - SECURITE - HYGIENE
Q - ETUDES D'ENSEMBLE
R - RECHERCHES, SCIENCES ET QUESTIONS SCIENTIFIQUES
S - DIVERS
T - TELETRANSMISSIONS - AUTOMATISATION - APPLICATIONS DE L'ELECTRONIQUE
U - MATIERES MINERALES

Tableau 10 : Domaines du plan de classement de l'INICHAR, édition de 1965

A. — Géologie - Gisements - Prospection - Sondages.
B. — Accès au gisement - Méthodes d'exploitation.
C. — Abattage et chargement.
D. — Pressions et mouvements de terrains - Soutènemer
E. — Transports souterrains.
F. — Aérage - Eclairage - Hygiène du fond.
G. — Epuisement.
H. — Energie.
I. — Préparation et agglomération des combustibles.
J. — Autres dépendances de surface.
Y. — Constitution, propriétés et analyse des combustibl
solides fossiles.
K. — Carbonisation.
L. — Gazéification.
N. —
M. — Combustion et chauffage.
O. — Valorisations diverses et industries chimiques dérivées
l'industrie charbonnière.
P. — Main-d'œuvre - Santé - Sécurité - Questions sociales
Q. — Etudes d'ensemble.
R. — Recherches - Documentation.
S. — Sujets divers techniques et scientifiques.

Les deux plans de classement se ressemblent à quelques exceptions près : sept domaines reprennent les mêmes termes, les autres diffèrent plus ou moins légèrement. Si leurs structures ne sont pas utilisées comme source pivot, les différents termes et leurs associations (précoordination) sont néanmoins analysés pour enrichir le thésaurus.

De leur côté, vers la fin de l'exploitation minière, nous l'avons vu, les Houillères mettent en place un cadre de classement, actuellement disponible au centre de documentation du CHM, constitué d'un ensemble de 169 entrées thématiques. Le cadre de classement organise les documents de la cinémathèque comportant l'ensemble de la production audiovisuelle des Houillères du Bassin du Nord–Pas-de-Calais, des archives (de 1720 à 1990), d'une bibliothèque contenant près de 7 000 ouvrages provenant des collections des compagnies et de campagnes d'acquisitions depuis 1985 et des images de l'iconothèque. De la même manière que la ressource scientifique du CERCHAR, les termes de cette ressource sont analysés.

D'autre part, en parallèle de sa fresque interactive[107] qui retrace l'histoire de la mine et des mineurs du Nord et du Pas-de-Calais en vidéo, l'INA a élaboré une classification hiérarchique thématique des vidéos qui est évolutive (un nouveau thème « Mineurs du monde » a été ajouté depuis notre travail de 2019 (Daloz et Chaudiron 2019)). Celle-ci possède quatre niveaux de spécificité.

[107] La fresque est accessible au public à l'adresse https://fresques.ina.fr/memoires-de-mines/fresque

Ces trois outils ont en commun le fait qu'ils relient tous des thématiques à des notices bibliographiques : signalétiques pour le CHM, analytiques pour le CERCHAR et l'INICHAR et descriptives pour l'INA[108].

5.1.3.2 Index thématiques

Certaines ressources sur le domaine minier possèdent une indexation thématique des unités qui les constituent (des articles de site internet, des séquences vidéo, des références bibliographiques, des termes). C'est le cas des vidéos témoignages d'anciens mineurs datant de 2013 (Cf. Ressources audiovisuelles en 5.1.4.1) qui sont indexées librement à l'aide de 77 thèmes ou l'index du site *Mineur de fond*[109] site de l'association APPHIM avec 37 thèmes repartis en 7 index libres. En 2004, Béatrice Turpin, linguiste, publie son glossaire *Les Mots de la Mine* bilingue français-chti, « réalisé principalement à partir d'un corpus constitué d'œuvres de mineurs écrivains du Nord et du Pas-de-Calais (poésies ou récits de vie) » et à visée « des linguistes par ce qu'il révèle de l'histoire et du fonctionnement de la langue ; il peut aussi intéresser tous les esprits curieux : curieux du monde d'abord, mais aussi de la langue et du langage »[110]. L'ouvrage contient un index classant le vocabulaire sous 6 grands thèmes déclinés en sous-parties. Cette ressource étant notre ressource pivot, nous adaptons son indexation à nos termes pour réaliser l'indexation thématique de notre terminologie (annexe 6). Enfin, en 2010, le directeur du Centre historique minier de Lewarde (Nord), André Dubuc, écrit un ouvrage dans lequel il range les références bibliographiques sous sept thématiques : histoire générale, histoire des mines et des compagnies, habitat minier, vie quotidienne, syndicalisme, les catastrophes minières et jeunesse (Dubuc 2010).

5.1.3.3 Terminologies

En 1974, suite à la publication par L'Office de la langue française (aujourd'hui et depuis 2002, Office québécois de la langue française) du premier fascicule de son Lexique anglais-français de l'industrie minière, Jean-Louis Rousseau, terminologue, revient sur sa méthodologie et présente la fiche terminologique associée à son travail (L.-J. Rousseau 1974). Il mentionne le fait d'avoir préféré « faire un lexique plutôt qu'un vocabulaire (avec définitions), puisque [leur] clientèle est constituée de spécialistes (ingénieurs, techniciens et mineurs) qui connaissent bien la matière et pour qui la définition des termes ne pose pas de problème » (L.-J. Rousseau 1974, 194). En 1978, sur la base de leurs travaux sur la terminologie minière[111] Auger et Rousseau publient un ouvrage méthodologique plus général sur la terminologie (Auger et *al.* 1978) dans lequel ils mentionnent, à propos du choix des parties du domaine à traiter, « [a]insi, dans la terminologie minière, on éliminera tout ce qui touche les charbonnages, car ce type d'exploitation n'existe pas au Québec » (Auger et *al.* 1978, 19). Si la

[108] La notice descriptive contient un ensemble d'indications sur les particularités de l'exemplaire du document décrit, la notice signalétique contient la description bibliographique d'un document et des points d'accès à celle-ci et la notice analytique contient un commentaire plus ou moins développé et éventuellement une critique du contenu du document. (Boulogne et Institut national des techniques de la documentation 2004)

[109] http://mineurdefond.fr/

[110] https://motsdelamine.pagesperso-orange.fr/informations.html

[111] De nombreuses éditions des fascicules sur les thèmes de l'exploitation (première édition en 1973), du traitement des minerais (première édition en 1974) et de la minéralurgie (première édition en 1981), seront publiés entre temps. Les termes sont aujourd'hui accessibles dans la banque de fiches terminologiques du Grand Dictionnaire Terminologique http://gdt.oqlf.gouv.qc.ca/index.aspx (Consulté le : 26/04/20).

ressource ne couvre pas notre domaine, les publications méthodologiques sont néanmoins intéressantes pour la conception de notre terminologie.

D'autre part, le Grand Dictionnaire Terminologique (GDT) « est une banque de fiches terminologiques rédigées par l'Office québécois de la langue française ou des partenaires de l'Office[112] ». 168 domaines d'emploi spécialisés permettent d'accéder aux concepts et aux termes qui les désignent en français et en anglais. La ressource enregistre un certain nombre de termes sous le domaine de l'*industrie minière*. Il n'est en revanche pas possible d'obtenir tous les termes du domaine car la saisie d'un mot-clé est obligatoire. Un moyen simple d'obtenir une grande quantité de termes est de repérer dans les lexiques une terminaison récurrente et de rechercher cette terminaison en sélectionnant « terme se termine par » dans le domaine recherché. C'est ainsi qu'en notant la terminaison –eur pour les noms de métier dans les mines, en sélectionnant *industrie minière* et *appellation de personne* nous trouvons 82 fiches que nous pouvons ensuite télécharger. Nous trouvons ainsi *abatteur*, *houilleur*, *aide-mineur*, *conducteur d'haveuse* etc. Des sous-domaines à l'industrie minière sont disponibles tels que *exploitation minière*, *disposition générale de la mine*, *exhaure de la mine*, *transport dans la mine*, *travaux d'exploitation minière*[113] etc. Les termes peuvent être indexés sont plusieurs domaines : par exemple, le terme *préposé au puits* relève du domaine de l'industrie minière mais aussi du domaine de l'appellation de personne. La richesse de la ressource nous sert à vérifier l'acception de certains termes tandis que les sous-domaines nous servent à réaliser une partie de l'indexation thématique de notre terminologie (annexe 6).

5.1.3.4 Thésaurus documentaires

En France, en 1990, le Bureau de Recherches Géologiques et Minières (BRGM) publie la troisième édition du *Thésaurus de l'exploitation minière souterraine – souterraine et à ciel ouvert* – (Auclaire 1990) – qualifié de « répertoire de termes codés, utilisé pour l'indexation des documents relatifs à l'exploitation des mines et carrières sur la base de données MINMOND du département Projets Miniers ». Le thésaurus est en parti réalisé à partir des travaux de thèse de Florence Auclaire en linguistique, soutenue en 1985 sur *L'Élaboration d'un dictionnaire technique trilingue français-allemand-anglais : domaine minier* (Auclaire 1985). Elle y traite notamment la problématique du multilinguisme. La figure 32 représente le classement conceptuel du thésaurus.

[112] http://gdt.oqlf.gouv.qc.ca/index.aspx

[113] Ou encore « aérage de la mine », « abattage du minerai », « soutènement dans la mine », « forage minier », « prospection minière », « minéralurgie », « Tir dans la mine », « Méthode d'exploitation minière », « service auxiliaire dans la mine », « Pyrominéralurgie », « Classification du minerai ou mineralurgie », « Concentration gravimétrique du minerai », « Fragmentation du minerai », « Contrôle minéralurgique »

Première Partie : classement conceptuel 4
0. Recherche minière 5
1. Travaux préparatoires 8
2. Opérations d'exploitation 10
3. Matériel 13
4. Méthodes d'exploitation 20
5. Annexes 23
6. Evaluation économique 25
7. Thèmes connexes 26
7.0. Economie 26
7.1. Environnement 26
7.2. Informatique appliquée 26
7.3. Infrastructure 26
7.4. Ingénierie 26
7.5. Législation 27
7.6. Minéralurgie 27
7.7. Transport 27

Figure 32 : classement conceptuel du thésaurus d'Auclaire (Auclaire 1990)

En 1997, le Service National de Géologie et Minière du Chili publie le *Tesauro de la mineria y Materias Afines*[114]bilingue espagnol-anglais (Avilés-F. et *al.* 1997). Le thésaurus est utilisé pour indexer les documents de la base de données REDIMIN (*Revista Digital Minera*), une revue électronique sur les mines pour un domaine encore en activité. Le Tableau 11 reprend les catégories thématiques du thésaurus que nous traduisons en français.

Tableau 11 : Catégories thématiques du thésaurus du Service National de Géologie et Minière du Chili

Catégories originales en espagnol	**Traduction française proposée par l'auteur**
Aleaciones	Alliages
Ciencias Básicas	Sciences fondamentales
Ciencias de la Tierra	Sciences de la Terre
Ciencias Sociales	Sciences sociales
Equipos	Equipement
Industria Minera	Industrie Minière
Materiales	Matériaux
Medio Ambiente	Environnement
Metalurgia Física	Métallurgie physique
Usos de Metales	Utilisations des métaux

En 2018, le Centre liégeois d'archives et de documentation de l'industrie charbonnière (CLADIC) a mis à notre disposition son thésaurus construit sur la base de la liste d'autorités RAMEAU et de l'indexation INICHAR destiné à indexer « un fonds d'environ quarante-sept mille documents (monographies, tirés à part et articles) et à un catalogue de plus de mille titres de périodiques (belges, français, hollandais, allemands, anglais et même russes), le tout paru entre 1525 et 2019 »[115]. Nous l'appellerons « thésaurus de Blegny-Mine ». Il comporte 23 domaines en octobre 2018 dont : archives et documentation, écologie, enseignement, industries, poids et mesures, santé, sciences naturelles, transports, culture, économie politique, entreprises, mathématiques, questions sociales, sauvetage, urbanisme, droit, énergie, génie civil, philosophie, recherches, sciences humaines, sécurité. La figure 33 présente un extrait du domaine sécurité du thésaurus du CLADIC. Comme nous pouvons le remarquer, la couverture du thésaurus dépasse le domaine des mines et du territoire de Blegny-Mine. Il permet néanmoins d'identifier de nouveaux termes candidats pour enrichir le thésaurus, d'étudier sa structure et de vérifier qu'aucun sous-domaine n'a été oublié.

[114] Traduction par l'auteur : Thésaurus de l'exploitation minière et des sujets connexes
[115] https://www.blegnymine.be/fr/centre-liegeois-d-archives-et-de-documentation-de-l-industrie-charbonniere-cladic (Consulté le : 26/04/20).

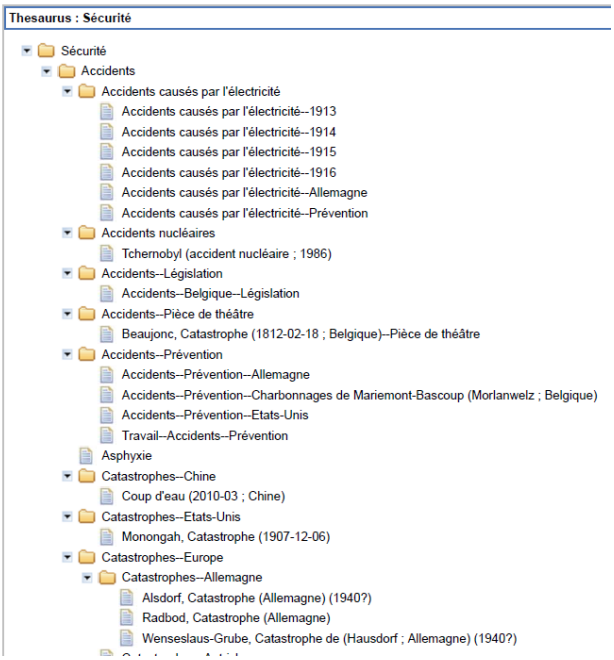


Figure 33 : Extrait du domaine sécurité du thésaurus du CLADIC en octobre 2018

Cette liste chronologique ne saurait être exhaustive mais présente un aperçu des ressources auxquelles nous avons eu accès et qui permettent de délimiter et d'organiser une partie conséquente du domaine. Cette vision est enrichie par d'autres ressources qui constituent les corpus d'enrichissement et de validation des SOC réalisés.

5.1.4 Autres ressources

5.1.4.1 Ressources audiovisuelles

Trois sous-corpus audiovisuels distincts sont utilisés dans nos travaux et servent pour des objectifs différents : un premier lot est destiné à l'enrichissement du thésaurus, un deuxième à sa validation et un troisième est destiné à être indexé (par ThesoMines, le thésaurus du patrimoine minier que nous avons conçu présenté en 6.2). Ces trois sous-corpus proviennent d'un corpus plus large d'environ 29 heures pour les vidéos/audios (la totalité des enregistrements et des vidéos n'a en effet pas été retranscrite) et de 315 photographies, qui a forgé une partie de notre représentation du domaine minier. Un premier lot de vidéo a été réalisé en 2013. Ces vidéos sont sous la forme de rushes et représentent les témoignages de 7 anciens mineurs ayant travaillé à Arenberg. Ils sont interrogés par Sylvie Leleu-Merviel, porteuse du projet MémoMines et ces vidéos appartiennent à la Communauté d'Agglomération de La Porte du Hainaut (CAPH). Un autre lot est réalisé en 2018 avec la participation de plusieurs membres du projet MémoMines, ce lot est destiné, suite à son traitement, à être déposé dans les archives ouvertes (HAL). Nous travaillons quant à nous sur les rushes. Enfin, des enregistrements sont effectués lors de différentes visites de sites miniers tout au long du projet, principalement en 2019 avant la crise sanitaire qui a rendu difficiles les captations et les rencontres avec les anciens mineurs.

Le premier sous-corpus est constitué d'un total d'environ douze heures de vidéo transcrites :

- Un témoignage d'une ancienne épicière à Arenberg (2018)

- Un témoignage d'un ancien mineur et un fils de mineur à Givenchy-en-Gohelle (2018)
- Une visite de site avec les membres de l'association *Les amis de la mine* de la fosse 2 à Oignies (2019)
- Un retour sur le parcours de visite du site d'Arenberg avec une guide touristique (2018)
- Des témoignages de vie et de travail de quatre anciens mineurs à Arenberg (2013)

Le deuxième sous-corpus est constitué d'environ 9 heures de vidéo transcrites :

- Une présentation d'une maquette représentant une fosse réalisée par un ancien mineur à Givenchy-en-Gohelle (2018)
- La suite du témoignage d'un ancien mineur et un fils de mineur à Givenchy-en-Gohelle (2018)
- La visite du musée de l'association d'histoire ONYACUM (2019)
- Des témoignages de vie et de travail de cinq autres anciens mineurs à Arenberg (2013)
- Une visite du site minier d'Arenberg par un ancien mineur (2018)

Le troisième sous-corpus est constitué de :

- 315 photographies prises lors des visites des musées de la mine (fosse 2 et fosse 9 à Oignies, musée de Lewarde, ONYACUM) dont 70 ont été indexées avec les descripteurs de ThesoMines.

5.1.4.2 Ressources supplémentaires

Pour compléter la représentation du domaine, un autre corpus hétérogène a été constitué avec des documents textuels :

- Un premier corpus presse océrisé sur le patrimoine minier provenant des journaux *Le Nord* et *La Voix du Nord* avec 695 fichiers organisés en une cinquantaine de thématiques (couvrant la période de 1990 à 2018)[116],
- Un deuxième corpus presse récupéré à partir d'Europresse provenant de *La Voix du Nord* avec 1313 articles (258 fichiers PDF) (couvrant la période de janvier 2004 à juin 2020)[117],
- 27 documents traitant d'aménagement du territoire publiés par la Mission Bassin minier,

[116] Ce corpus a été effectué avant notre arrivée sur le projet MémoMines et la méthodologie du choix des articles est inconnue.
[117] Nous ne sommes pas à l'origine de ce corpus non plus. La requête qui a été utilisée est la suivante : TIT_HEAD= (charbon | mines | mine | minier | mineur | fosse) & LG= court & LEAD= ((charbon | mines | mine | minier | mineur | fosse)! ("le fossé" | "au fossé")) ! TEXT= ("école des mines" | antipersonnel | "mine de rien" | "grise mine" | noeux | douchy | bully | auchy | halluin | chine | marles | enquin | carmaux | nomades | monoxyde). Une note est ajoutée concernant le bruit provoqué par cette requête large, qui est réduit par la lecture des titres et premières lignes des articles (Sur 5544 articles, 4231 n'ont pas été retenus).

- 11 articles scientifiques au format PDF en histoire, en linguistique, en information-communication, en art,
- Un document du BRGM sur la remontée des eaux dans les anciennes mines de charbon du NPC,
- Des documents d'archives de la bibliothèque IRIS de Lilliad (documentation technique sur l'exploitation des mines, sur les explosifs etc.),
- Un inventaire de l'association « les amis du musée de la mine »,
- Des numérisations d'écrits en ch'ti sur les mines.

Des ressources web d'autre part :

- Des URL vers des ressources sur les mines

Et des ressources iconographiques :

- Des cartes géographiques du territoire du Bassin minier

5.1.5 Outils utilisés

Les outils utilisés pour la construction du corpus sont : LibreOffice Calc pour l'enregistrement et la description des données, le logiciel Agent Ransack pour les requêtes dans les corpus (placés dans des dossiers sur notre disque dur), un logiciel d'océrisation de l'Atelier National de Reproduction des Thèses (ANRT à l'Université de Lille[118]) pour océriser le premier corpus presse. Les numérisations ont été réalisées avec l'application *ClearScanner* sur smartphone. Le moteur de recherche Google a été utilisé pour les recherches en ligne.

5.2 Des ressources lexicales à la terminologie de la mine

Comme le mentionnent (Bourigault, Aussenac-Gilles, et Charlet 2004, 89) « [l]e constat de la variabilité des terminologies s'impose : étant donné un domaine d'activité, il n'y a pas une terminologie, qui représenterait le savoir sur le domaine, mais autant de ressources terminologiques ou ontologiques que d'applications dans lesquelles ces ressources sont utilisées ». Ainsi, s'il existe déjà des terminologies sur le domaine de la mine, elles ne correspondent pas forcément au patrimoine minier. Il faut donc concevoir une ressource spécifique à partir de celles existantes.

5.2.1 Plusieurs terminologies, plusieurs vocabulaires miniers

Pour pouvoir définir le vocabulaire et la terminologie minière, il faut déjà définir ce que l'on entend par la « mine » et son domaine que l'on nomme « domaine de la mine », « domaine minier » ou « domaine des mines ». Le domaine des mines, comme la pluralité du terme « mines » le laisse entendre, désigne

[118] L'ANRT n'existe plus depuis 2017 et a été integré en 2018 au Service Commun de Documentation (SCD) de l'université de Lille

plusieurs mines. Il existe plusieurs terminologies minières et plusieurs vocabulaires miniers. Ceux-ci dépendent en effet du type de minerai extrait, et des pratiques qui s'adaptent au type d'exploitation, construisant des carrières à ciel ouvert ou des galeries souterraines. Dans notre cas, les exploitations dans le bassin du Nord – Pas-de-Calais étaient souterraines et le minerai extrait était le charbon. Des termes comme « à ciel ouvert » et tous les termes équivalents ou spécifiques qui en traitent tel que « carrière » ne rentreront donc pas dans notre champ d'étude.

5.2.2 Un langage spécifique à une activité et à un territoire

Les termes, spécifiques au travail à la mine, employés par les mineurs et leur entourage forment le vocabulaire minier.

Dans les entretiens avec les anciens mineurs, des expressions contenant le verbe « appeler » et précédé du pronom impersonnel « on » désignant très souvent la communauté des mineurs, permettent de distinguer des termes spécifiques au langage de la mine : « on appelait ça une *rallonge* », « on appelait ça *porion de quartier* » (entretien avec Raymond, ancien mineur ayant travaillé à Arenberg, corpus de 2013), « […] qu'on appelle le *généphone* » (visite avec l'association des amis de la mine de la fosse 2 d'Oignies, corpus de 2018).

Le langage peut varier au sein du même domaine d'activité comme le montre cet exemple tiré de l'article du journal *Le Nord* n° 138 de mars 2000 :

> « Les employés du "jour"[119] ne se mêlaient pas aux mineurs... Après mai 68, on a vu arriver les ingénieurs qui "avaient fait les barricades", comme ils disaient. Ceux-là étaient plus proches des mineurs... mais ça ne plaisait pas toujours à la direction. Je me souviens d'un ingénieur, fils de mineur, qui s'était fait réprimander parce qu'il parlait patois avec les ouvriers ! »

Cet extrait de notre entretien avec Martial, un ancien mineur, le montre également :

« Virginie: #1 Et le garde il parlait français, patois ? #
Martial: #2 ah non, ch'garde il parlait franç- euh # patois
Virginie: #1 ah, et par contre le directeur de la mine il parlait français lui ? #
Amélie: #2 tout le monde parlait patois #
Martial: euh il parlait français mais quand il était au fond avec certaines équipes ch'était un peu du paternalisme il faisait exprès de parler patois
Virginie: ouais c'était par euh, vous voyez je suis parmi vous quoi, je suis avec vous hein. Et les ingénieurs ?
Martial: bah il y en avait des bons mais
Virginie: il y en a qui parlaient un peu patois aussi quand même ?
Martial: ouais il y en avait des bons, il y avait des crapules qui bon souvent ils, c'est arrivé qu'ils changeaient un carreau, bon ça c'était de sa faute hein »

Un jugement de valeur à travers la langue est mis en évidence ici et ces extraits nous montrent bien que le parler des ouvriers du fond diffère de celui des ingénieurs ou du directeur au jour. Ceci permet donc de distinguer deux sous-domaines d'activités : le travail réalisé par la hiérarchie et le travail réalisé par les ouvriers.

[119] Note dans l'article : ceux qui travaillaient en surface, par opposition à ceux du « fond ».

De l'autre côté, le terme « patois » est employé par les anciens mineurs pour désigner leur dialecte. Ce terme permet d'identifier d'autres termes spécifiques au langage des mineurs du Nord – Pas-de-Calais. Comme Raymond nous l'affirme « *mes toiles bleues, mes loques ed fosse comme on dit en patois* ». Des expressions telles que « *comme on dit en patois* » permettent en effet d'identifier des termes qui étaient appelés d'une certaine façon par une communauté plus ou moins restreinte.

Enfin, les définitions comportent parfois des mentions spécifiques sur l'utilisation des termes sur le territoire : ainsi, l'emploi de *bouveau*, *car* et *barrou* est situé dans le département du Pas-de-Calais. Parfois, la mention n'est pas très précise « en usage dans quelques mines du Pas-de-Calais ».

5.2.3 Des noms, des verbes, des expressions ?

Les terminologies n'intègrent en général que des substantifs. D'après Cabré (2000, 26), les termes et les mots coïncident dans la classe du nom : « [s]i on restreint la catégorie de termes aux unités référentielles et si on exclut les autres classes fonctionnelles de signification spécialisée (adjectifs, verbes et adverbes), nous pouvons affirmer que les termes et les mots coïncident dans la classe du nom ». Mais peut-on représenter un domaine dans sa globalité en excluant ce que Cabré appelle les « classes fonctionnelles de signification spécialisée » ? Le langage minier est en effet rempli d'expressions et de verbes riches de sens. Par rapport à la productivité par exemple, le verbe *farcer* et son emploi *être farcé* ou *ête farcé* signifie le fait d'arriver en retard dans le langage des mineurs et de perdre sa journée de travail. Est-ce que cette expression et les autres doivent apparaître dans la terminologie ? Si nous choisissons de ne pas inclure les adjectifs et adverbes dans notre terminologie, nous gardons néanmoins des verbes tels que : *ardoubler* qui signifie '*faire des heures supplémentaires*', *abistoquer* qui signifie '*faire un boisage rapide et provisoire*' ou *s'ajouquer* qui signifie '*s'accroupir*'. Ceux-ci portent le sens des actions effectuées par les mineurs qui ne peuvent être nominalisées, ou bien que par l'emploi de termes associés comme *heure supplémentaire*, *boisage*, *position de travail*. Nous verrons néanmoins que seuls les substantifs seront inclus dans le thésaurus.

5.2.4 Terme vs concept et relations sémantiques

Le tableau 12 ci-dessous illustre la différence entre des termes représentés par des lettres et des concepts représentés par deux couleurs (ici rouge et noir). Dix termes sont représentés (A, B, C, D, E, F, G, H, I, J) et deux concepts (dont le premier, en rouge, est composé des termes A1, B, C, D, E, F, G, H et le deuxième, en noir, des termes A2, I, J). Nous considérons que toute relation de synonymie, de variation et d'antonymie lient les termes entre eux à un même concept. Ceci paraît simple avec des lettres et deux couleurs mais devient plus complexe avec des termes ainsi que des définitions basées sur un code qui est la langue et qui, par économie et respect de la mémoire humaine, n'autorise pas une infinité de forme pour ses termes. Dans notre travail, la différenciation conceptuelle ne sera d'ailleurs pas notée par une couleur comme ici (les couleurs n'étant pas infinies) mais par un identifiant c'est-à-dire une forme unique, qui peut être alphanumérique.

Tableau 12 : Termes vs Concepts					
Entrée	Acception	Définition	Synonyme	Variante	Antonyme
A	A1	-	B	F	
A	A2	-			
B		-	C		G
C		-	E	H	
D		-		A1	
E		-			
F		-			
G		-			B
H		-			
I		-	J		

Le premier terme A, possède deux acceptions que nous notons A1 et A2. La première acception A1 a pour synonyme B et comme variante orthographique F dont les définitions se situent dans la ligne de leur entrée. La deuxième acception n'a pas de synonyme, ni de variante. Le terme B a une seule acception et a pour synonyme C et pour antonyme G. G a réciproquement pour antonyme B. Les relations de synonymie, de variation et d'antonymie sont toutes réciproques c'est-à-dire que A1 est équivalent à B et à l'inverse B est équivalent à A, A pour variante F et F a pour variante A, B a pour antonyme G, et G a pour antonyme B. Même si A1 et B sont synonymes, ceci n'empêche pas A1 d'avoir une définition différente de B (sur la forme, mais pas dans le sens).

Le tableau 13 illustre l'identification d'un concept dans les ressources lexicales. Le concept sera noté par l'identifiant *barrou* (la fréquence de cette forme est supérieure aux autres) et est identifié à 16 entrées différentes. Nous n'en montrons ici que 10.

Tableau 13 : Le concept *barrou* et ses réalisations dans la langue							
Entrée	**source bibliographique**	**Acception**	**Équivalent français**	**Synonyme**	**Renvoi**	**Variante**	**Antonyme**
barrou	memo-mines_RL_11	Wagonnet vide dans le Pas-de-Calais.	barroud		balle | berline | car |	barou | barout	
car a vude	memo-mines_RL_11	Wagonnet vide.					
baroux	memo-mines_RL_16	Berline vide, en patois	berline vide				
barou	memo-mines_RL_11	Wagonnet vide.	berline vide		barrou		
barrou	memo-mines_RL_14	Berline vide destinée à être encagée pour retour au fond (contrairement à la balle)					balle
barroud	memo-mines_RL_18	Dans le nord une berline vide					
car a vide	memo-mines_RL_18	Berline vide.					
car à vute	memo-mines_RL_39	Berline vide					
barrou	memo-mines_RL_19	Berline vide					

Le travail consiste ensuite à relier un identifiant représentant le concept aux différents termes avec des relations sémantiques de thésaurus comme dans le tableau 14.

Tableau 14 : Concept *barrou*			
Concept	**Variantes (EP)**	**Termes associés (TA)**	**Terme générique (TG)**
barrou	barroud barou barout berline vide car a vide car a vude car a vute wagonnet vide	balle	berline [EP car]

Le concept est par la suite étudié au niveau de sa granularité sémantique et de sa fréquence d'apparition dans les corpus pour être intégré dans la hiérarchie du thésaurus ou non. Dans le cas du concept *barrou*, il sera rattaché en employé pour du descripteur *berline* car il a été jugé trop spécifique pour indexer les documents.

5.3 De la terminologie minière au thésaurus

5.3.1 Structuration : relation hiérarchique

La structuration permet d'organiser le lexique en domaines et sous-domaines et d'établir des relations hiérarchiques entre les termes. Dans notre sous-corpus de travail, B. Turpin et Goudaillier (2004) est la ressource la plus riche en vocabulaire à indexer ses unités linguistiques. Ceci nous permet de la choisir comme ressource pivot. Elle divise le lexique en six grands domaines : le travail du mineur, la mine, le transport, l'air et l'eau, le mineur, la productivité et appointements. La plupart du lexique se situe dans les deux premiers thèmes, le travail et la mine (figure 34). Le domaine du travail du mineur est le plus spécialisé, avec sept sous-thèmes qui sont les suivants : l'abattage à l'explosif, le boisage et le soutènement, le gros matériel d'abattage et ses accessoires, les actions, les cadres et piles de soutènement, les métiers et tâches, les outils et pièces diverses (figure 35).

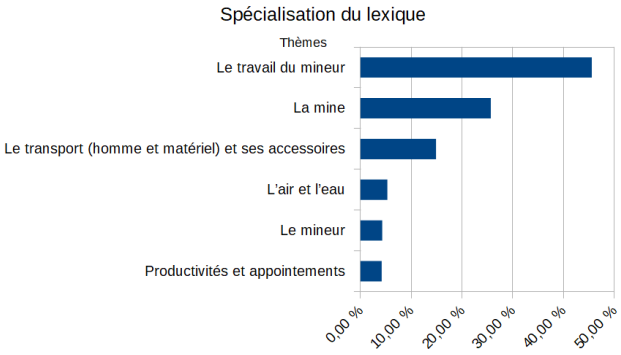


Figure 34 : Fréquence des termes indexés par le vocabulaire de (B. Turpin et Goudaillier 2004)

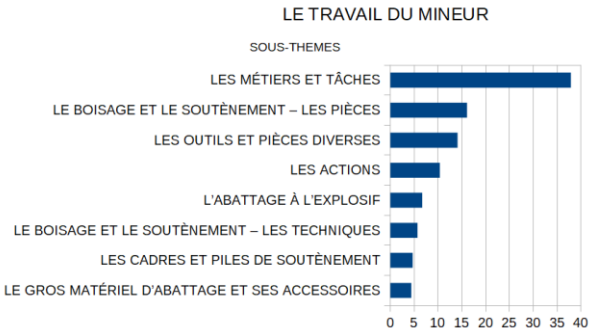


Figure 35 : Nombre de termes indexés par le vocabulaire de (B. Turpin et Goudaillier 2004) par sous-thématique : le travail du mineur

Un premier travail consiste donc à indexer avec les thématiques issues de Turpin et Goudaillier les mêmes termes ou les synonymes identifiés dans les autres ressources. Par exemple : le terme *arcette* est rangé sous la thématique *La mine* et sous-thématique *Les lieux*. Nous marquerons *La mine – Les lieux* dans un premier temps pour toutes les variantes de ce terme donc également *recette* dans toutes les ressources lexicales (ici au nombre de neuf).

Cette première structuration est enrichie par l'ajout de catégories sémantiques. Ainsi, si la catégorie est *le transport (homme, matériel) et ses accessoires*, nous ajoutons *moyen de transport* si le terme relève de cette catégorie. Si la catégorie est *l'air et l'eau*, nous ajoutons *eau* ou *air* en fonction de l'élément associé. Si le terme présente une ambiguïté sémantique, une notation chiffrée est utilisée pour dénoter les différents concepts. Ainsi le terme *clichage* est indexé : *1 ; le transport (homme, matériel) et ses accessoires_appareil_retenue ; 2 ; les lieux_recette*. Le lecteur retrouve l'ensemble de l'indexation thématique en annexe 6.

Tableau 15 : Indexation thématique non contrôlée du vocabulaire[120]

accident
animal
animal_transport
appareil de communication
appareil_air comprime
appareil_forage
appareil_gps
appareil_mesure_inclinaison
appareil_mesure_inclinaison de galerie
appareil_sécurité_grisou
appareil_séparation gravimétrique
appareil_tri
appellation_mineur
appellation_mineur_blessé
appellation_mineur_creusement
appellation_mineur_délégué
appellation_mineur_galibot
appellation_mineur_péjoratif
appellation_mineur_rescapé
appellation_mineur_syndicat
appellation_mineur_tâche
appellation_personne
appellation_personne ; appareil_sécurité
appellation_personne_culture
appellation_personnes

Une autre possibilité est de travailler au niveau des termes génériques contenus dans les définitions. Nous nous inspirons pour cela des travaux de Gheorghita (2011, 2014) qui se base sur l'analyse des définitions du TLFi pour l'indexation et la recherche d'images. Nous utilisons en revanche une méthode

[120] Le tiret du bas joue le rôle de séparateur entre les termes, avec suppression des mots vides. L'indexation APPAREIL DE COMMUNICATION est une erreur qui devrait apparaître comme telle : *appareil_communication*. L'indexation *appellation_personne ; appareil_sécurité* est également une erreur, elle devrait être notée « 1 ; *appellation_personne* : 2 ; *appareil_sécurité* » pour noter l'ambiguïté.

plus manuelle et qualitative. Il suffit pour cela de repérer des tendances dans les définitions afin d'extraire semi-automatiquement les termes. Par exemple, le terme *mesureux* est défini comme 'ouvrier chargé de mesurer le charbon destiné aux mineurs'. Nous recherchons toutes les définitions qui contiennent le terme générique, ici, *ouvrier* et nous les analysons. Pour cette requête, 228 définitions contiennent ce terme, dont 139 commencent par *ouvrier*. Les 139 entrées correspondantes seront automatiquement reconnues comme étant des termes spécifiques. L'analyse manuelle permet de repérer 21 entrées supplémentaires avec des définitions comme 'mauvais ouvrier', 'les mineurs d'about ou abouts sont les ouvriers affectés aux travaux du puits' ou encore 'jeune apprenti ouvrier-mineur'. Suite à l'analyse des formes des entrées par rapport à leur sens, 71 concepts sont en réalité spécifiques à la catégorie *ouvrier*.

Un troisième travail consiste à replacer ces termes génériques dans le domaine étudié au bon niveau de granularité par rapport aux autres et de réfléchir à la catégorie conceptuelle à laquelle appartient *ouvrier*. Pour cela, il est possible de rechercher le terme générique dans des thésaurus existants. Par exemple, dans le thésaurus DemoVoc[121], le descripteur *profession* est utilisé, le descripteur *statut social* dans le thésaurus INRAE[122] ou encore *CSP (catégorie socioprofessionnelle)* dans le thésaurus Motbis[123] (renommé THESAGRI 2020). On choisira finalement d'employer un synonyme de profession, *métier*, comme terme générique (avec les termes équivalents *personnel* et *main d'œuvre*) inséré dans un concept générique *activité* représentant l'activité minière au sens large. On ne choisira cependant pas d'inclure *ouvrier* dans la structure car nous avons souhaité structurer les descripteurs en fonction du type d'activité réalisée : abattage, administration, entretien, exhaure etc. Ceci donne les descripteurs métier-abattage, métier-administration et ainsi de suite. Le nombre de dénomination pour les métiers dans les mines est assez conséquent. Le regroupement par type d'activité n'est cependant pas possible pour tous car certains métiers sont destinés à effectuer plusieurs types d'activité. Nous avons choisi de ne pas regrouper ceux-ci et de les placer directement sous *métier*. Par exemple, mécanicien pourrait apparaître sous métier-transport et métier-entretien d'après l'étude des définitions, mais notre thésaurus n'étant pas polyhiérarchique, nous le plaçons donc sous métier.

[121] http://thesaurus.web.ined.fr/navigateur/demovoc/fr/
[122] https://consultation.vocabulaires-ouverts.inrae.fr/thesaurus-inrae/fr/
[123] http://www.cndp.fr/thesaurus-thesagri/site/

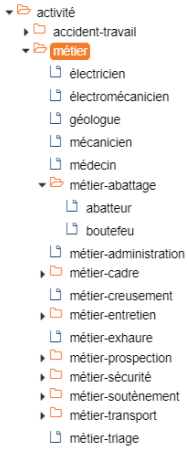


Figure 36 : Structure des descripteurs sous « métier » dans ThesoMines sur Opentheso[124]

Enfin, pour chacun des SOC sur le domaine identifié (ressources classificatoires), une analyse sémantique et thématique des unités constitutives minimales (tantôt linguistiques, documentaires, tantôt techniques) est effectuée. Chaque nouvelle thématique fait l'objet d'un nouveau concept approuvé ou non lors de l'étape de validation. La notation de ces concepts doit suivre les recommandations des normes des thésaurus. La figure 37 illustre la structuration de notre thésaurus à gauche (par niveaux) et l'alignement horizontal vers les thématiques dans les différentes ressources à droite. Ainsi, le concept *matériel-sécurité* dans le thésaurus est aligné avec les indexations des ressources suivantes : (Auger, Rousseau, et Québec. Office de la langue française 1973 ; Auclaire 1990 ; Servicio Nacional de Geología y Minería et Sociedad nacional de minería 1997 ; B. Turpin et Goudaillier 2004 ; CERCHAR 1976), les vedettes matières de RAMEAU et la terminologie.

Premier niveau	Deuxième niveau	Troisième niveau	Quatrième niveau	A= ROUSSEAU&AUGER	C = AUCLAIRE	D = TESAURO	E = TURPIN	H = CERCHAR	B = RAMEAU
	matériel-soutènement-marchant								
		vérin hydraulique							
	scie								
matériel-sécurité				18. Prévention. — Sécurité.	3.13. MATERIEL-SECURITE		la sécurité	S – DIVERS (machines, entretien,bâtiments du jours, sécurité incendie, procédés et matériaux de construction, emploi de matériaux)	Charbon -- Mines et extraction -- Mesures de sécurité
	antidéflagrant				matériel antidéflagrant				
	arrêt-barrage				arrêt barrage				
	barrière-sûreté								
	détecteur-gaz				détecteur de gaz			F - AERAGE - GRISOU - POUSSIERES - FEUX - ECLAIRAGE	Lampes de sûreté
	jeton								
	matériel-éclairage						l'éclairage et ses accessoires	F - AERAGE - GRISOU - POUSSIERES - FEUX - ECLAIRAGE	Électricité dans l'exploitation minière

Figure 37 : Alignement des concepts et construction des descripteurs du thésaurus

5.3.2 Relation associative

La terminologie est l'« ensemble des termes relatifs à un système notionnel élaboré par des constructions théoriques, par des classements ou des structurations de matériaux observés, de pratiques sociales ou

[124] Logiciel utilisé pour concevoir ThesoMines (Cf. présentation du logiciel en 5.3.5) https://opentheso.huma-num.fr/opentheso/

d'ensembles culturels » (TLFi). L'aspect relatif implique un certain nombre de liens entre les termes qui forment le système notionnel. Dans le cadre des thésaurus documentaires, les liens sont des relations sémantiques de type hiérarchique, équivalent et associatif. En tant qu'outil d'indexation, le thésaurus nécessite l'association pour orienter l'utilisateur dans sa recherche. Le concepteur de thésaurus doit donc anticiper les associations que l'utilisateur s'attendra à trouver. Pour établir ces relations d'association entre les termes, plusieurs choix sont possibles :

1. Identifier dans les ressources classificatoires les termes ou notions précoordonnés et la syntaxe ou les symboles reliant plusieurs termes entre eux

Par exemple, dans la classification du CERCHAR, *pression de terrains* est placé avec *soutènement*, et dans le plan de classement du CHM, *Machine d'extraction* avec *Chevalement* et *Treuil de secours*.

2. Identifier dans les ressources lexicales les renvois des termes vers d'autres (ou les antonymes)

Ceci ne fonctionne pas dans la mesure où les renvois des ressources lexicales correspondent souvent aux variantes orthographiques ou dialectales. Ces renvois sont donc plutôt des relations d'équivalence.

3. Relever les termes présents dans les définitions (hors termes spécifiques) d'une entrée et lier cette entrée aux termes de la définition.

Par exemple, *creusement* est défini : 'Voies principales tracées pour accéder au puits'. Il est alors possible de relier le descripteur creusement au descripteur puits.

5.3.3 Termes descripteurs et non-descripteurs : relation d'équivalence

Le taux de recouvrement permet de calculer le nombre de termes communs aux ressources afin de pouvoir définir les termes descripteurs et les non-descripteurs grâce au calcul de la fréquence de leur apparition dans les corpus. Par exemple, ci-dessous (tableau 16) les termes descripteurs seront *abattage* (qui apparaît dans seize lexiques) et *battroulle* (qui apparaît dans deux). Les termes rejetés seront *abatache, abattache, batteroulle et battroulle*, de par leur plus faible fréquence d'apparition. Si le terme *batroulle* et ses variantes apparaissent bien dans le thésaurus, ils ne seront finalement pas descripteurs car nous avons fait le choix de n'afficher que des termes en français (dans la mesure du possible). Ici, *battrouille* est l'équivalent français de *battroulle* d'après (B. Turpin et Goudaillier 2004) mais nous ne trouvons pas d'occurrence de celui-ci ni dans les dictionnaires de langue française ou terminologies spécialisées, ni dans les corpus, ce qui nous amène à choisir un autre terme équivalent *barre à mine* que nous formalisons en *barre-mine*.

Tableau 16 : Fréquence d'apparition de quelques termes dans le corpus	
Terme	**Fréquence d'apparition du terme dans le corpus**
abatache	1
abattache	1
abattage (fr.)	16
batteroulle	1
battrouille (fr. d'après (B. Turpin et Goudaillier 2004) mais nous ne trouvons pas d'occurrence dans les dictionnaires de la langue française)	1
battroulle	2

Les critères de sélection des descripteurs sont donc la fréquence dans les corpus puis si la forme est en dialecte et qu'un équivalent français existe, nous préférons le terme en français s'il apparaît dans plusieurs ressources lexicales ou terminologiques fiables, le terme en dialecte sera relié au descripteur par une relation d'équivalence.

5.3.4 Enrichissement du thésaurus

L'enrichissement du thésaurus se réalise grâce à au premier sous-corpus (Cf. 5.1.4.1) composé de douze heures de vidéo retranscrites. Une extraction manuelle des candidats termes permet premièrement d'enrichir la terminologie puis au regard de leur apparition fréquente dans les corpus ou leur pertinence pour l'indexation de les insérer dans la structure du thésaurus. Sur Opentheso (Cf. 5.1.5), un module permet d'enregistrer les candidats termes avant de les accepter ou non dans le thésaurus. Une trace est ainsi réalisée pour les choix effectués.

Figure 38 : Capture des candidats termes pour l'enrichissement du thésaurus sur Opentheso

5.3.5 Outil utilisé

Le logiciel utilisé pour l'organisation de ces connaissances est le logiciel OpenTheso[125] dans sa version 4.5.8. C'est un gestionnaire de thésaurus multilingue développé par Miled Rousset[126] et proposé sur l'infrastructure d'Huma-Num[127]. Le logiciel peut se définir selon 9 fonctionnalités principales dont les éléments sont décrits dans le tableau 17. L'état de notre utilisation de chaque fonctionnalité est mentionné.

Tableau 17 : Fonctionnalités et utilisation du logiciel Opentheso

Fonctionnalités	**Description des fonctionnalités**	**État de l'utilisation**
(1) Fonctionnalités de thésaurus documentaire	-relations de généricité, d'association, d'équivalence, définition de notes d'application -présentation hiérarchique des concepts -poly-hiérarchie	-utilisées -utilisées -non utilisées
(2) Fonctionnalités linguistiques	-définition -note historique -note éditoriale -ajouter une langue	-utilisées pour les termes techniques -non utilisées -non utilisées -non utilisées
(3) Fonctionnalités collaboratives	-possibilité de travailler en mode collaboratif et de proposer des candidats descripteurs	-utilisées
(4) Fonctionnalités illustratives et de spatialisation	-possibilité d'ajouter des images -possibilité d'ajouter des coordonnées GPS et une carte	-non utilisées -non utilisées
(5) Fonctionnalités web sémantique/interopérabilité	-possibilité d'exporter le thésaurus au format csv, pdf, skos, Json-Id, Turtle, Json -possibilité d'aligner son vocabulaire avec d'autres -possibilité d'exporter les identifiants et les alignements	-utilisées : skos, pdf, csv -non utilisées -non utilisées
(6) Fonctionnalités de structuration ontologique	-possibilité de distinguer un terme spécifique en « terme générique », « partitif », ou « instance » -possibilité d'ajouter des facettes	-non utilisées -non utilisées
(7) Fonctionnalités de gestion et d'analyse	-possibilité de générer des statistiques	-utilisées

[125] https://opentheso.huma-num.fr/opentheso/
[126] Miled Rousset est responsable Plateforme Web Sémantique&Thesauri (PSIR) et Directeur informatique du GDS-Frantiq, MOM-CNRS (FR 3747) Université Lyon2
[127] https://www.huma-num.fr/

(8) Fonctionnalités d'ouverture et de pérennité	-possibilité d'ouvrir le thésaurus au public	-utilisées
(9) Fonctionnalités de recherche	-possibilité d'effectuer des requêtes pour naviguer dans la structuration	-utilisées

Les fonctionnalités liées à l'ouverture (8) ainsi qu'au web sémantique (5) nous ont incitée à sélectionner ce logiciel. Notre utilisation est actuellement basique, dans le sens où notre intérêt se porte sur la structuration du thésaurus, et sur la gestion facilitée des descripteurs contrairement à la saisie de données dans un tableur. Les fonctionnalités liées à l'enregistrement d'autres types de données seront à approfondir dans la suite du travail. Pour une valorisation du patrimoine culturel immatériel, l'ajout d'une catégorie de langue de type « dialecte » serait intéressante.

5.4 Du thésaurus au modèle ontologique

L'étape décrite dans la précédente section a consisté à définir, structurer et relier les concepts propres au patrimoine minier dans un langage documentaire à base de descripteurs tel qu'un thésaurus. Ces tâches ont permis l'évolution du thésaurus jusqu'à sa version v1.0 disponible en ligne[128] depuis novembre 2021. Le résultat est une base pour le travail suivant. L'objectif final du travail est de réaliser une ontologie de domaine. L'approche utilisée pour passer du thésaurus à l'ontologie se compose de plusieurs étapes que nous décrivons dans cette partie. La méthodologie s'appuie sur des travaux tels que ceux de Szabados et Letricot (2012), sur l'utilisation du modèle CIDOC CRM pour les objets du patrimoine antique, sur ceux de Bourigault, Aussenac-Gilles et Charlet (2004) pour la construction d'ontologie sur la base de textes et Aussenac-Gilles, Baziz et Hernandez (2006) pour l'importance de la prise en compte de la terminologie dans la construction de ressources ontologiques.

Dans ce document, le terme *classe* désigne les éléments notés *E* dans le modèle ontologique du CIDOC CRM. Ainsi, *E18 Physical Thing* est une classe. Elle peut avoir une ou plusieurs sous-classe(s) et une ou plusieurs classe(s) générique(s). Le terme « propriété » désigne les relations qu'entretiennent les classes entre elles dans le modèle. Elles sont notées P. Ainsi *P131 is identified by*, *P2 has type* sont des propriétés. De la même manière que les classes, elles peuvent avoir des sous-propriétés et des propriétés génériques. Elles ne peuvent pas relier n'importe quelles classes entre elles. Dans le thésaurus, les relations sémantiques de type génériques, associatives et équivalentes peuvent être représentées par des propriétés contenues dans le modèle ontologique. Par exemple, la classe *E41 Appellation* peut être reliée à une autre classe *E41 Appellation* par la propriété *P139 has alternative form* qui pourrait représenter la relation équivalente. Dans le modèle, les équivalences peuvent par exemple se faire entre identifiants, entre nom d'acteur, entre dates, entre adresses etc. L'objectif du travail est d'aller plus loin que des relations de thésaurus (hiérarchie, association, équivalence) dans la sémantique de représentation des concepts du patrimoine minier. Les étapes ci-dessous retrace l'appropriation du modèle jusqu'au peuplement du modèle sélectionné.

[128] https://opentheso.huma-num.fr/opentheso/api/theso/MemoMines

5.4.1 Choix du modèle ontologique du CIDOC CRM

Le CIDOC CRM a été choisi dans le cadre du projet MémoMines, choix orienté et justifié par les expériences d'Éric Kergosien et de Bernard Jacquemin lors de leurs projets respectifs TECTONIQ[129] que nous citons plus bas (Cf. 5.4) et DOREMUS[130]. Nous reprenons ci-dessous les raisons que nous avons développées dans notre article (Daloz 2019) pour appuyer ce choix :

> « Premièrement, il s'agit d'une norme stable et mature, dont la version officielle (disponible en français et en anglais) est un standard ISO (ISO 21127), version ayant été déjà mise à jour une fois (publiée en 2006 puis mise à jour en 2014). [Depuis 2000, le CIDOC CRM est développé en standard ISO dans un effort collaboratif avec le groupe de travail du CIDOC CRM (SIG) et ISO/TC46/SC4/WG9]. Deuxièmement, le modèle permet une manipulation facilitée par la mise à disposition d'outils et d'une documentation détaillée. Troisièmement, le modèle a été construit sur la base d'une centaine de formats de métadonnées, y compris les informations patrimoniales UNESCO, […]. Quatrièmement, le modèle possède une implémentation déjà effective dans les langages en vigueur, notamment OWL pour sa version officielle, qui permet de faire des inférences, c'est-à-dire de déduire des nouvelles connaissances qui se réalisent par de nouvelles relations entre les objets. Le Bœuf (2013, 102) écrit qu'« il existe deux encodages du modèle CIDOC CRM dans des langages spécifiques permettant la publication d'information sous forme de données liées : l'encodage en RDF Schema – RDFS – disponible sur le site même du CIDOC CRM *Special Interest Group* (CIDOC CRM *Special Interest Group*, 2011a), et l'encodage en OWL établi par l'université d'Erlangen- Nuremberg (Friedrich- Alexander Universität Erlangen-Nürnberg, 2009- …) ». Cinquièmement, un *mapping* vers le modèle FRBRoo, son successeur, et pour l'instant toujours en cours d'évolution, est déjà effectué. Enfin et sixièmement, son développement est spécifiquement prévu pour gérer des informations issues de sources hétérogènes, ce qui dans notre cas, est un critère positif pour le balayage le plus exhaustif possible du domaine « (Daloz 2019).

5.4.2 Appropriation du modèle ontologique CIDOC CRM

La prise de connaissance du modèle est réalisée par une première lecture attentive des définitions de chaque classe et propriété sur le site web du modèle[131] avec une étude de la hiérarchie de classes et propriétés et saisie des liens entre propriétés et classes. À cela s'ajoute une lecture des différentes versions (notamment la 5.0.4 qui était la version officielle au début du travail et la 6.2.1, dernière version publiée en RDF à la date du début du travail). L'appropriation doit mener à une représentation mentale globale que l'on se fait du modèle. Pour cela, il faut pouvoir y accéder grâce à un point d'entrée.

[129] https://tectoniq.meshs.fr/
[130] https://www.doremus.org/
[131] http://www.cidoc-crm.org/versions-of-the-cidoc-crm

Hyvönen choisit de décrire le modèle en plaçant au centre la notion d'événement. Il compare celle-ci à de la « colle sémantique » mettant en relation différents aspects de la connaissance (les objets, les acteurs, le temps, les lieux, les appellations etc.) les uns avec les autres, d'une manière harmonisée dans une collection de patrimoine culturel (Hyvönen 2012, 44).

Nous choisissons de concevoir le modèle comme un arbre. Les classes de plus haut-niveau du modèle (pour la v.6.2.1) sont *E2 Temporal Entity*, *E77 Persistent Item*, *E52 Time-Span*, *E53 Place*, *E54 Dimension*. Toutes ces classes sont rattachées à une classe générique, *E1 CRM Entity*, ce qui rend possible la représentation des classes en arbre (figure 39) :

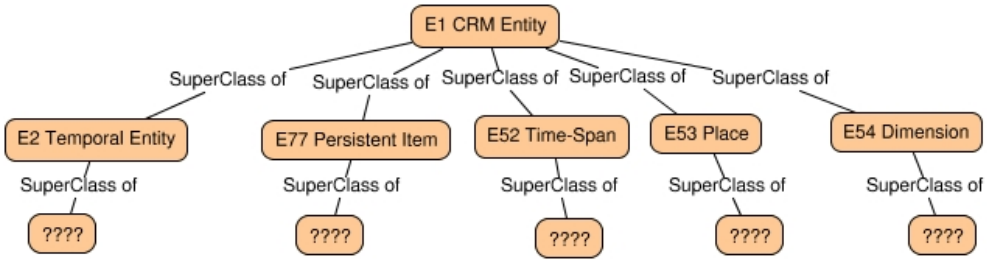


Figure 39 : Représentation en arbre des classes de haut-niveau du CIDOC CRM (v.6.2.1)

Il suffit ensuite de se représenter le modèle à partir des définitions données par chacune des classes de haut-niveau. Le modèle comprend : tous les phénomènes qui se produisent sur une étendue limitée dans le temps (*E2 Temporal Entity*), tous les éléments qui présentent des caractéristiques persistantes de nature structurelle substantiellement liées à leur identité et à leur intégrité, parfois appelés "endurants" en philosophie (*E77 Persistent Item*), des étendues temporelles abstraites, au sens de la physique galiléenne, ayant un début, une fin et une durée (*E52 Time-Span*), les étendues dans l'espace, en particulier à la surface de la terre, au sens pur de la physique : indépendamment des phénomènes temporels et de la matière (*E53 Place*) et les propriétés quantifiables qui peuvent être mesurées par des moyens calibrés et qui peuvent être approchées par des valeurs, c'est-à-dire par des points ou des régions dans un espace mathématique ou conceptuel, comme les nombres naturels ou réels, les valeurs RVB, etc. (*E54 Dimension*).

Cette représentation mentale n'est cependant pas complète puisqu'il manque toute la sémantique des liens et les instances qui donnent le sens au modèle. Il est alors nécessaire de lier ces premières études à des expériences pratiques. Ainsi, la participation en octobre 2019 à une école thématique intitulée DONIPAT[132] nous a permis de prendre part à différents ateliers sur le modèle ontologique et d'effectuer des exercices consistant par exemple à deviner à quelle classe appartenait telle ou telle instance ou bien à utiliser le logiciel Yed Graph Editor qui est un programme de création de diagrammes[133] et utilisé pour modéliser les données. Enfin, la lecture de travaux et méthodologies effectués sur un autre patrimoine industriel et utilisant le CIDOC CRM tel que le textile et notamment le mémoire de Master réalisé par

[132] Données Interopérables pour le Patrimoine
https://f-origin.hypotheses.org/wp-content/blogs.dir/1718/files/2019/10/DONIPAT_programme.pdf
[133] Nous n'utilisons cependant pas ce logiciel pour nos travaux.

Kaouther Ben Smida (Ben Smida 2016) dans le cadre du projet TECTONIQ (Kergosien et *al.* 2015b, 2015a, 2019) a été déterminante pour notre approche.

5.4.3 Confrontation avec d'autres travaux utilisant le modèle

Dans cette deuxième étape, nous reprenons la hiérarchie de classes du modèle en ajoutant une marque distinctive pour chaque classe sélectionnée dans le projet TECTONIQ (Ben Smida 2016, 49 à 53). Nous tentons ensuite pour chacune de ces classes de les peupler par des instances issues de notre connaissance sur le patrimoine minier (tableau 18). Certaines classes sont trop génériques pour être instanciées comme nous pouvons le voir sur ce même tableau 18. La colonne « Text. » correspond aux classes sélectionnées pour représenter le patrimoine textile dans le projet TECTONIQ. Une note « peu de renseignements là-dessus » précise la faible probabilité de trouver des instances pour la classe correspondante. 26 classes en plus seront sélectionnées pour décrire le patrimoine minier. Le lecteur retrouve ce travail en annexe 3.

Tableau 18 : Instanciation des classes par des connaissances issues du patrimoine minier

Id	Text.	Classe											Instances issues du patrimoine minier
E1	ok	CRM Entity											Toutes les classes et instances du CIDOC CRM
E2	ok	-	Temporal Entity										Cette classe est trop générique
E3	ok	-	-	Condition State									Fosse n° 4 des mines de Lens « en ruine » pendant la guerre 1914-1918
E4	ok	-	-	Period									Période de l'exploitation minière
E5	ok	-	-	-	Event								La Seconde Guerre mondiale, catastrophe de Courrières, la Grève des mineurs du Nord-Pas-de-Calais (1941)
E7	ok	-	-	-	-	Activity							Construction des corons
E8		-	-	-	-	-	Acquisition Event						La donation d'un objet au musée de Lewarde (mais peu de renseignements là-dessus)
E9		-	-	-	-	-	Move						Déplacement d'un objet (mais peu de renseignements là-dessus)
E10		-	-	-	-	-	Transfer of Custody						Changement de détenteur d'un objet (mais peu de renseignements là-dessus)
E11		-	-	-	-	-	Modification						Modification d'un objet matériel comme une restauration (peu de renseignements là-dessus) voir MBM

5.4.1 Classes et propriétés représentatives du patrimoine minier

Dans sa version 6.2.1, le CIDOC CRM possède 89 classes et 149 propriétés. Nous avons fait l'hypothèse que toutes les classes et propriétés n'étaient pas nécessaires pour la représentation du domaine du patrimoine minier. Le travail a donc consisté à revenir sur chacune afin de les sélectionner ou non.

5.4.1.1 Passage des descripteurs de thésaurus aux classes du modèle

Cette étape consiste à choisir les classes du CIDOC CRM qui représentent les concepts portés par les descripteurs de thésaurus sans se préoccuper des propriétés (tableau 19). Dans ce tableau, les activités domestiques sont représentées par la classe *E7 Activity* tandis que les appareils et matériels sont reliés à des *E18 Physical Thing*. Ce travail exploratoire est réalisé alors que le thésaurus n'est pas encore stabilisé. Les descripteurs *cuisine* ainsi qu'*entretien* n'apparaîtront par exemple pas dans ThesoMines v1.0 ; *cuisine* se révélant être ambigu et pouvant être indexé grâce au descripteur *repas* sous *coutume et tradition* et l'activité d'entretien remplacée par le descripteur nettoyage car plus fréquemment employé dans le corpus et le descripteur métier-entretien dans le cadre de l'activité minière a été ajouté. Nous montrons donc ici l'importance d'avoir stabilisé le thésaurus avant de passer à cette étape.

Tableau 19 : Classes CIDOC CRM attribuées aux descripteurs du thésaurus

DESCRIPTEURS		CLASSES CIDOC CRM CORRESPONDANTES
Activité domestique		E7 Activity
.	cuisine	E7 Activity
.	entretien	E7 Activity
.	jardinage	E7 Activity
Appareil et matériel		E18 Physical Thing
.	matériel d'exhaure	E18 Physical Thing
.	matériel de chauffage	E18 Physical Thing
.	matériel de communication	E18 Physical Thing
.	matériel de constitution, préparation et agglomération des charbons	E18 Physical Thing
.	matériel de forage	E18 Physical Thing
.	matériel de foudroyage	E18 Physical Thing
.	matériel de manutention	E18 Physical Thing
.	matériel de mesure et d'évaluation	E18 Physical Thing
.	matériel de nettoyage	E18 Physical Thing
.	matériel de prospection et sondage	E18 Physical Thing
.	matériel de sauvetage	E18 Physical Thing
.	matériel de soutènement	E18 Physical Thing
.	matériel d'abattage	E18 Physical Thing
.	matériel médical	E18 Physical Thing

5.4.1.2 Description des entités nommées avec le CIDOC CRM

Dans cette étape, nous faisons émerger des catégories de concepts qui sont : les entités spatiales, les entités temporelles et les acteurs. Celles-ci correspondent en fait aux entités nommées (Cf. 8.1). Nous sélectionnons ensuite les classes et propriétés correspondantes en suivant l'approche du projet TECTONIQ, que nous confrontons avec notre connaissance du domaine du patrimoine minier. Ce travail permet par exemple de décrire qu'un lieu (*E53 Place*) peut s'identifier par (*P87 is identified by*) une adresse (*E45 Adress*), des coordonnées spatiales (*E47 Spatial Coordinates*), un nom de lieu (*E48 Place Name*) et qu'il peut s'insérer (*P89 falls within*) dans un autre lieu (*E53 Place*). Quant aux entités temporelles (*E2 Temporal Entity*), elles sont identifiées (*P1 is identified by*) par une appellation (*E41 Appellation*) et une date (*E50 Date*). Plus précisément, cela peut être une période (*E4 Period*) ayant une durée (*E52 Time-Span*) ou plus généralement un événement (*E5 Event*) ayant une durée (*E52 Time-Span*). Enfin, les acteurs (*E39 Actor*) sont soit des personnes (*E21 Person*), soit des groupes (*E74 Group*) qui sont identifiés (*P131 is identified by*) par une appellation (*E82 Actor Appellation*). Le groupe (*E74 Group*) peut avoir été ou être composé (*P107 has current or former member of*) d'un agent ou d'un groupe (*E39 Actor*)(tableau 20).

Tableau 20 : définition des entités spatiales et temporelles et les acteurs

Les lieux		
E53 Place	P87 is identified by (identifies)	E45 Address
		E47 Spatial Coordinates
		E48 Place Name
	P89 falls within	E53 Place

Les entités temporelles		**Propriétés**		**Exemple donné par CIDOC CRM**
E2 Temporal Entity		P1 is identified by (identifies)	E41 Appellation	Le siècle dernier
E2 Temporal Entity		P1 is identified by (identifies)	E50 Date	1956 ou 12-12-1956
-	E4 Period	P4 has time-span (is time-span of)	E52 Time-Span	L'âge de bronze (E4)
	E5 Event	P4 has time-span (is time-span of)	E52 Time-Span	Le tremblement de terre à Lisbonne en 1955 (E5)

Les acteurs			
E39 Actor			
-	E21 Person	P131 is identified by (identifies)	E82 Actor Appellation
	E74 Group	P131 is identified by (identifies)	E82 Actor Appellation
	E74 Group	P107 has current or former member (is current or former member of)	E39 Actor

5.4.2 Traduction des classes et propriétés

Afin d'accéder au sens porté par les termes en anglais employés pour les classes et propriétés dans notre langue, il nous a semblé nécessaire d'entamer un travail de traduction de celles sélectionnées. Nous nous sommes néanmoins rendue compte que la traduction était disponible pour la plupart des classes et propriétés dans la version en RDF (l'évolution du modèle intégrant petit à petit les traductions jusqu'à établir des ressources multilingues spécifiques en allemand, grec, français, portugais, russe et chinois[134]). Travaillant tout d'abord sur la version 5.0.4, puis sur la 6.2.1, nous avons également noté l'absence de certaines classes dans la version 5.0.4. (tableau 21).

Tableau 21 : traduction des classes de l'anglais vers le français et note sur la présence ou non dans la version 5.0.4

CLASSES	traduction française	présence dans version 5.0.4
E22_Man-Made_Object	objet fabriqué	
E27_Site	site	
E3_Condition_State	état matériel	
E39_Actor	agent	
E41_Appellation	appellation	
E42_Identifier	identificateur d'objet	
E52_Time-Span	durée	
E53_Place	lieu	
E55_Type	type	
E57_Material	matériau	
E92_Spacetime_Volume	volume espace-temps	non

Le travail avec des traductions n'est néanmoins pas entièrement possible car certaines classes et propriétés ne sont pas traduites (tableau 22), pour cela le travail sera effectué en anglais, dans la langue originale du modèle.

[134] http://cidoc-crm.org/versions/cidoc_crm_v6.2.1_with_translations.html#E1

Tableau 22 : Classes du CIDOC CRM v.6.2.1 qui n'ont pas de traduction en français http://cidoc-crm.org/versions/cidoc_crm_v6.2.1_with_translations.html

E59 Primitive Value	Missing translation in 'French' language
E60 Number	Missing translation in 'French' language
E61 Time Primitive	Missing translation in 'French' language
E62 String	Missing translation in 'French' language

5.4.3 Des descripteurs du thésaurus à l'ontologie

Afin de ne pas représenter une vision unique du patrimoine minier, le passage des descripteurs du thésaurus aux classes du modèle ontologique se réalise à plusieurs en collaboration avec les membres du projet ANR MémoMines, Stéphane Chaudiron, Bernard Jacquemin et Éric Kergosien. Un certain nombre de réunions ont donc été réalisées de novembre 2019 à avril 2021. Cette étape prend tout d'abord en considération les premiers niveaux (*topconcept*) du thésaurus en partant du principe que le sens porté par la forme de ceux-ci correspond à ceux porté par les classes du modèle. Par exemple, le terme de premier niveau « activité » peut être comparé par traduction avec la classe *E7 Activity* et logiquement, tous les termes spécifiques devraient pouvoir se rattacher à cette classe. Or, ceci ne fonctionne pas lorsque la relation hiérarchique est de type partitif. Par exemple, le descripteur « règlement » sous « activité » n'est pas une activité (ici le terme « activité » fait référence au travail général dans les mines), mais il « fait partie » du concept de l'activité professionnelle. Ce ne sera donc pas automatiquement un *E7 Activity*. Un travail de précision des relations sémantiques utilisées pour construire le thésaurus doit ainsi être effectué pour pouvoir automatiser le processus.

La méthode consiste donc à rattacher chaque descripteur à une classe et d'étudier le sens des relations associatives du thésaurus pour identifier les propriétés. Ainsi, le descripteur « aérage », de la même manière que le descripteur « creusement » dans leur aspect intentionnel et en tant qu'action réalisée par un agent (cela peut être la compagnie minière responsable de l'étude –descripteur « patronat »– de l'aérage du fond ou bien les anciens mineurs chargés du creusement d'une galerie et d'un puits –descripteur « métier-creusement »– servant à l'aérage ou encore les ingénieurs ayant travaillé sur un plan) sont considérés comme un *E7 Activity*. Le modèle ontologique possédant des classes et des sous-classes, il s'agit également de trouver le bon niveau de granularité en réfléchissant à l'utilisation potentielle du résultat : est-ce que l'utilisateur voudra requêter des types spécifiques de *E7 Activity* telles que *E65 Creation* et savoir qu'un auteur a créé un objet d'information de type élément visuel ou bien que ce même auteur a simplement réalisé une activité ? Pour notre part, nous avons souhaité représenter un niveau de granularité assez fin pour être au plus près des propos recueillis auprès des anciens mineurs.

Le tableau 23 illustre la méthodologie utilisée pour réduire le CIDOC CRM au patrimoine minier à l'aide d'une réflexion sur les descripteurs et leur potentielle sémantique à instancier ensuite grâce aux corpus. Dans celui-ci les termes *domain* et *range* sont utilisés pour désigner le sujet et l'objet de la propriété.

Tableau 23 : Regroupement des descripteurs par classe					
Classe sélectionnée	**Descripteurs**	**Classe-*Domain***	**Propriété**	**Classe -*Range***	**Commentaire**
E22 Man-Made Object	chapelle, église, mosquée, presbytère : 3				
		E22 Man-Made Object	*P2 has type (is type of)*	*E55 Type*	Lien vers descripteur ThesoMines correspondant
		E55 Type	*P1 is identified by (identifies)*	*E41 Appellation*	Forme identifiée (mot) dans les corpus
		E22 Man-Made Object	*P10 falls within (contains)*	*E92 Spacetime Volume*	Un élément à l'intérieur de… (poupée russe)
			P44 has condition (is condition of)	*E3 Condition State*	Un état donné à un moment donné
		E3 Condition State	*P4 has time-span (is time-span of)*	*E52 Time-Span*	La période de l'état (voir entité temporelle)
			P45 consists of (is incorporated in)	*E57 Material*	Composé de matières
			P46 is composed of (forms part of)	*E18 Physical Thing*	Est composé d'objets physiques (partie de...comme les légos)
			P48 has preferred identifier (is preferred identifier of)	*E42 Identifier*	Identifiant URI
			P49 has former or current keeper (is former or current keeper of)	*E39 Actor*	Celui qui possède l'objet…qui gère…
			P51 has former or current owner (is former or current owner of)	*E39 Actor*	Celui qui est propriétaire de l'objet…
			P53 has former or current location (is former or current location of)	*E53 Place*	À noter que les P54 et 55 existent pour préciser si ce sont des localisations actuelles…

Des préconisations concernant le modèle du CIDOC CRM peuvent ensuite être données : par exemple, l'utilisation de la classe *E92 Spacetime Volume* sera précisée par un commentaire (*un élément à l'intérieur de… (poupée russe)* permettant d'accéder mentalement plus rapidement au sens de cette classe. De même, des commentaires sont notés concernant des choix qui nécessitent une justification comme par exemple la double sélection de certaines propriétés telles que la *P53 has former or current location* ou encore l'utilisation facultative d'une propriété comme la propriété *P19 was intended use of* par rapport à une autre, la *P16 used specific object*. Dans le premier cas, ceci permet de préciser le type de *E53 Place* à noter (soit la localisation spatiale à la surface de la terre, soit la profondeur) dans le deuxième, d'ajouter une précision sur l'utilisation d'un objet (intentionnelle d'une part, objet spécifique d'autre part).

5.4.4 Représentation des modèles obtenus : cartes conceptuelles

La réalisation de schémas et l'utilisation de l'outil de création de cartes conceptuelles CmapCloud nous a ensuite paru nécessaire. Les schémas réalisés reprennent pour chaque type de classe sélectionné dans le CIDOC CRM (comme *E7 Activity*, *E22 Man-Made Object*), les propriétés associées à chaque représentation différente. Par exemple, pour le descripteur du thésaurus « commémoration », nous avons

choisi la classe *E7 Activity* et le schéma de propriétés représenté sur la figure 40 ci-dessous. Pour chaque représentation commune, nous avons donné un identifiant numérique (ici, 16, visible dans l'encadré jaune). Le schéma se lit : « *an activity (E7) took place at (P7) a place (E53)* » ou « *a place (E53) witnessed (P7.i) an activity (E7)* » et ainsi de suite. 41 schémas ont ainsi été réalisés.

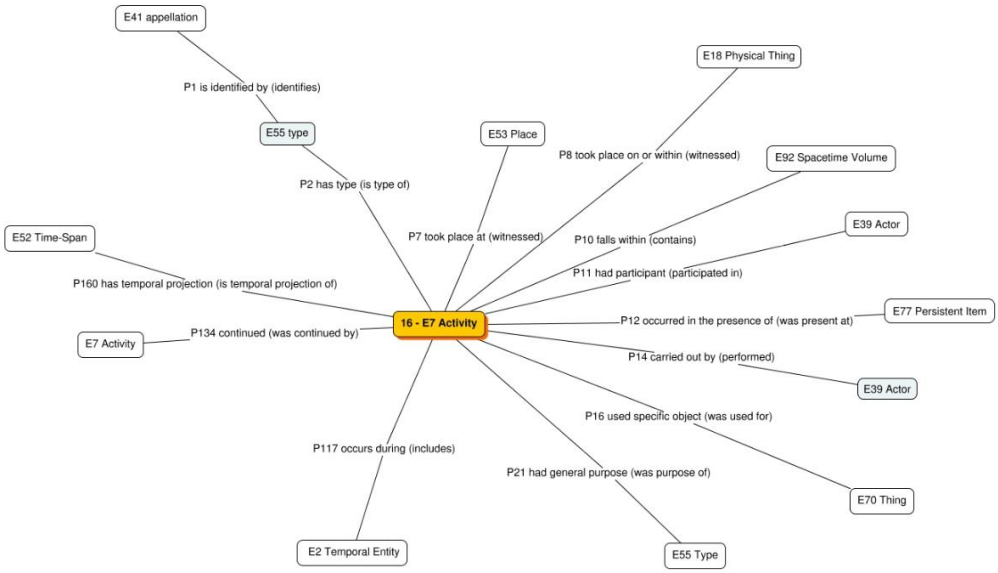


Figure 40 : Représentation conceptuelle des classes et propriétés sélectionnées pour décrire le concept de commémoration

Toutes les classes du CIDOC CRM possèdent des « propriétés propres » ainsi que des propriétés qu'ils héritent de leurs classes génériques. Nous appelons propriétés propres les propriétés qui apparaissent sous *properties* (figure 41) dans la description des concepts du CIDOC CRM. La version française de 1999 (Groupe de Normalisation documentaire ICOM/CIDOC et *al.* 1999) nomme ceci des attributs. Nous appelons les propriétés héritées les propriétés qui se situent sous *inherited properties*.

In First Order Logic:
E5(x) ⊃ E4(x)
Properties:
P11 had participant (participated in) : E39 Actor,
P12 occurred in the presence of (was present at) : E77 Persistent Item
Inherited properties:
List of properties
Inherited references:
List of references

Figure 41 : Capture d'écran des propriétés de la classe *E5 Event*

Ainsi, pour le schéma 16 (figure 40) la liste des propriétés correspondante est la suivante :

Trois propriétés propres :

P14 carried out by (performed) : *E39 Actor*,
P16 used specific object (was used for) : *E70 Thing*
P134 continued (was continued by) : *E7 Activity*

Et sept propriétés héritées :

E5 Event. *P12 occurred in the presence of (was present at)* : *E77 Persistent Item*
E5 Event. *P11 had participant* (participated in) : *E39 Actor*
E92 Spacetime Volume. *P10 falls within* (contains) : *E92 Spacetime Volume*
E4 Period. *P8 took place on or within* (witnessed) : *E18 Physical Thing*
E4 Period. *P7 took place at* (witnessed) : *E53 Place*
E1 CRM Entity. *P2 has type* (is type of) : *E55 Type* - *E1 CRM Entity*. *P1 is identified by* (identifies) : *E41 Appellation*
E92 Spacetime Volume. *P160 has temporal projection (is temporal projection of)* : *E52 Time-Span*

Comme nous pouvons le constater d'après la figure 42, les classes *E22 Man-Made Object*, *E7 Activity* et *E25 Man-Made Feature* permettent de décrire sémantiquement la plus grande variété d'objets. Ceci montre que notre sélection représente en fait des sous-classes spécifiques du CIDOC CRM. Ainsi, le schéma 3 serait une sous-classe de *E22 Man-Made Object* pour laquelle il faudrait trouver un terme générique commun aux descripteurs chapelle, église, mosquée, presbytère et appliquer un nouvel identifiant comme par exemple « E223 Bâti religieux ». Nous ne réalisons cependant pas ceci dans nos travaux et considérons ces descripteurs au même niveau sémantique que *E22 Man-Made Object*, ceci pouvant évidemment être contesté.

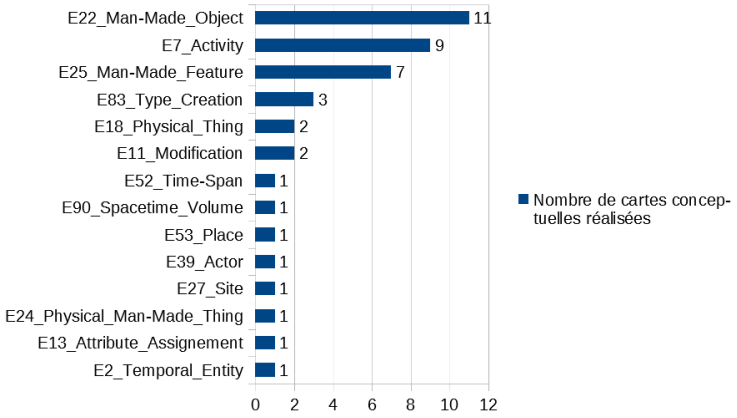


Figure 42 : Nombre de cartes conceptuelles réalisée par classe sélectionnée

Par souci de place et lisibilité, le schéma de la figure 40 (et tous les schémas effectués) ne représentent pas les classes génériques (ici *E5 Event*, *E92 Spacetime Volume*, *E4 Period* ou *E1 CRM Entity*) qui permettent de transmettre les propriétés héritées.

Ce premier travail nous a permis de pointer des défauts dans la structuration du thésaurus menant à six réunions avec Stéphane Chaudiron et Bernard Jacquemin de mai à septembre 2020 sur la restructuration et la validation du thésaurus grâce au deuxième sous-corpus audiovisuel (Cf. 5.1.4.1).

5.4.5 Enregistrement des données

La hiérarchie des descripteurs du thésaurus a été reprise puis pour chacun des descripteurs, nous avons effectué un lien vers des sources externes, un alignement avec d'autres thésaurus, nous avons ajouté la carte conceptuelle correspondante, un commentaire d'application pour l'instanciation de l'ontologie et une illustration à partir du corpus image. Le tableau 24 répertorie toutes les données qui peuvent être collectées pour les descripteurs de thésaurus afin d'orienter les choix des classes et propriétés du modèle ontologique. Nous ajoutons une mention sur la disponibilité des données dans chacun des SOC réalisés : TerminoMines, ThesoMines, OntoMines. La mention indisponible signifie que l'information n'est pas

enregistrée dans la structure des trois SOC mais qu'elle est tout de même disponible sur notre disque dur. Il s'agit ensuite, pour chaque type de données, d'utiliser un format permettant le passage au web sémantique. Ainsi, les alignements avec les descripteurs d'autres thésaurus ne peuvent actuellement se réaliser qu'avec des thésaurus dont le format respecte les standards du web sémantique. C'est par exemple le cas pour le thésaurus de l'UNESCO et celui des Thesaurus pour la description et l'indexation des archives locales anciennes, modernes et contemporaines et RAMEAU (disponibles au format RDF/XLM) mais pas pour les autres.

Tableau 24 : Données disponibles (ou potentiellement déposables) sur TerminoMines, ThesoMines et OntoMines

Disponibilité dans les SOC	**Donnée**	**Explication**
ThesoMines et lien à OntoMines par la propriété *P2 has type*	Identifiant	Identifiant du descripteur issu de ThesoMines (6 chiffres)
Pourrait être décrit dans OntoMines avec P5 consists of, *P2 has type*, P106 is composed of	Type de relation	Précision sur le type de relation avec le terme générique (partie de, type de, composant de…)
TerminoMines	Provenance du terme	Provenance du terme (vocabulaire contrôlé, terminologie, corpus, plan de classement de Lewarde, du CERCHAR etc.)
TerminoMines	Définition(s)	Définition(s) tirée(s) de la terminologie minière
ThesoMines	Note d'application	Note d'application du descripteur (pour le thésaurus)
Pourrait être décrit dans OntoMines avec P3 has note	Note générale	Commentaire ou notes personnelles à propos du descripteur
Indisponible	Lien URL vers ce schéma sur CMAPtool	Lien vers la carte conceptuelle correspondante en ligne
OntoMines instanciée	Instances dans le corpus	Instance du concept dans un corpus (extrait de corpus presse, audiovisuel, blog, autres…)
Ressources indexées par TerminoMines	Contexte	Contexte du terme
- pour le moment, nulle part mais peut apparaître dans OntoMines	Lien vers source externe	Lien URL vers le concept utilisé dans un document trouvé sur le web.
Indisponible mais les articles de presse nettoyés pourraient être versés sur le plateforme OKAPI pour être indexés par OntoMines	Nombre de hits dans le corpus presse (VDN)	À partir du logiciel Ransack, requêtes effectuées à partir du descripteur dans le deuxième corpus presse. Sont alors notés le nombre de hits dans le nombre d'item avec « hits » correspondant au nombre de fois où le terme apparaît et « item » au nombre de fichier.
Pourra être disponible dans ThesoMines	Alignement vers RAMEAU	Vedette-matière équivalente dans RAMEAU (recherches effectuées à partir des notices d'autorité)
Indisponible	AUCLAIRE	Alignement avec le thésaurus d'Auclaire (Auclaire 1990)
Indisponible	CERCHAR	Alignement avec le plan de classement (CERCHAR 1976)
Indisponible	TESAURO	Alignement avec le thésaurus Tesauro de la minera y Materias Afines, (Servicio Nacional de Geología y Minería et Sociedad nacional de minería 1997)
Indisponible	AUGER&ROUSSEAU	Alignement avec la terminologie d'Auger et Rousseau (Auger, Rousseau, et Québec. Office de la langue française 1973)
Indisponible	TURPIN	Alignement avec les thématiques du dictionnaire de Turpin (B. Turpin et Goudaillier 2004)
Indisponible	Blegny-Mine	Alignement avec le thésaurus du centre de documentation de Blegny-Mine (Centre liégeois d'archives et de documentation de l'industrie charbonnière (CLADIC), s. d.)
Pourra être disponible dans ThesoMines	Thesaurus pour la description et l'indexation des archives locales anciennes, modernes et contemporaines	Alignement avec le thésaurus pour la description et l'indexation des archives locales anciennes, modernes et contemporaines
Pourra être disponible dans ThesoMines	UNESCO	Alignement avec le thésaurus de l'UNESCO
Indisponible	MOTBIS	Alignement avec le thésaurus Thésagri

5.5 Du modèle ontologique réduit à l'ontologie de domaine instanciée

L'instanciation est effectuée grâce à la plateforme OKAPI conçue par l'INA, plateforme également utilisée par l'INALCO pour les archives audiovisuelles de la recherche. Le but principal d'OKAPI est d'organiser le travail d'annotation tout en contraignant les choix de l'annotateur. La plateforme permet de noter des connaissances multilingues, elle est évolutive et permet de décrire des médias ainsi que leur contenu en utilisant les ontologies. Les différents documents (vidéos, audios, images, objets numérisés en 3D ou textes au format PDF) peuvent être segmentés puis être indexés et décrits.

Pour notre travail, une version OWL du modèle du CIDOC CRM a été déposée sur la plateforme afin tester l'instanciation de notre modèle à partir des documents audiovisuels créés dans le cadre du projet ANR MémoMines (documents également déposés sur la plateforme). Afin d'obtenir une ontologie de domaine, l'instanciation de plusieurs documents différents est nécessaire. Nous avons vu que les documents que nous possédons sont des visites de sites miniers, présentations d'objets, et des témoignages d'ancien mineur. L'instanciation est effectuée à partir des connaissances historiques, non ambigües, vérifiables. Le repérage des instances se réalise à partir des concepts du thésaurus ainsi que des entités nommées extraites des différents corpus. Le thésaurus ThesoMines v1.0 au format *SKOS* est en effet versé sur la plateforme et indexe la transcription des documents audiovisuels à partir des formes des descripteurs et des non-descripteurs (ceci ne prend donc pas en compte les formes fléchies des termes). Ceci permet néanmoins un premier repérage des instances de concept. Ci-dessous, le repérage de la forme *abattage* dans les tours de parole transcrits (figure 43 et figure 44).

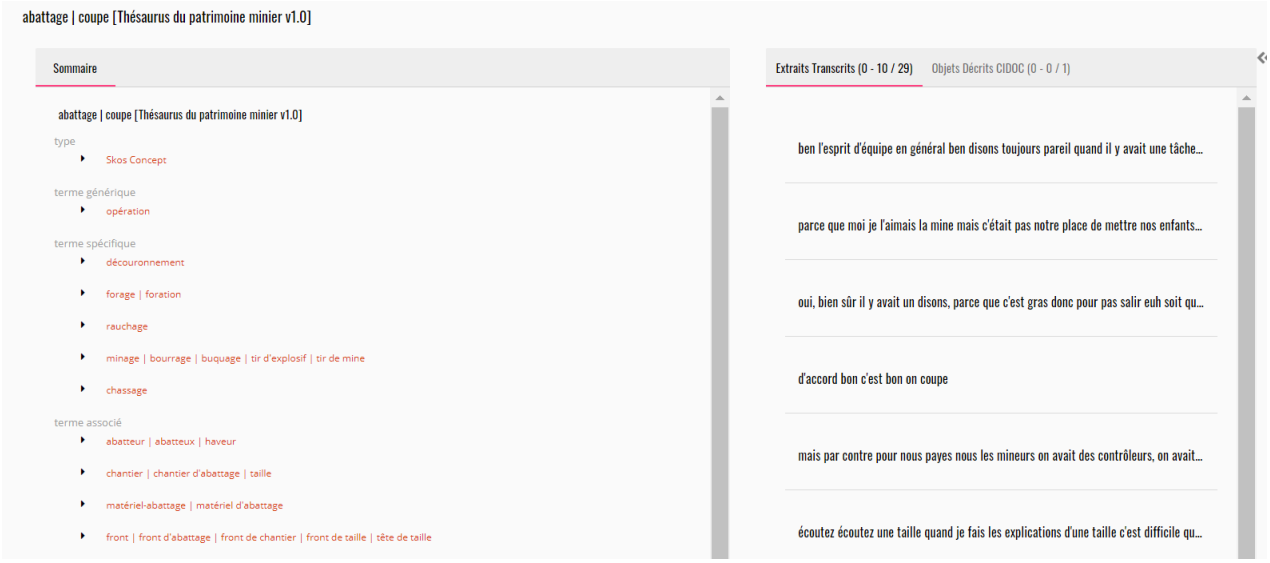


Figure 43 : Indexation de 29 tours de parole du corpus audiovisuel avec le terme *abattage* (EP *coupe*)

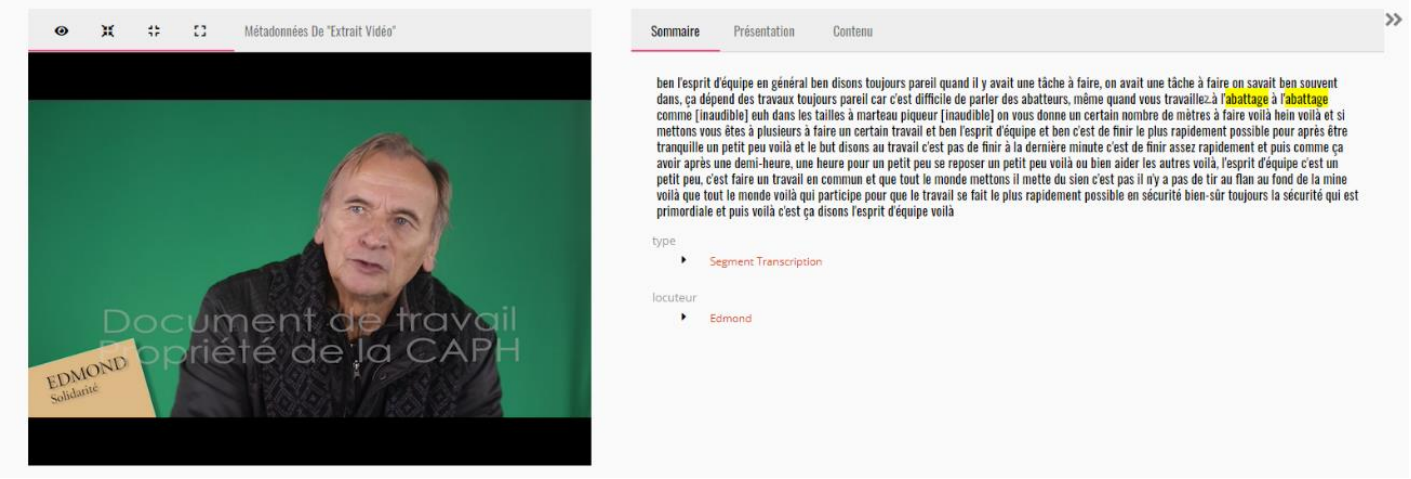


Figure 44 : Forme « abattage » dans un tour de parole transcrit d'Edmond

La méthodologie d'instanciation est composée de cinq étapes principales. Elle consiste tout d'abord à segmenter le média en tours de parole et à leur attribuer une thématique en fonction du contenu, puis à analyser les connaissances du contenu de chaque séquence en séparant celles pouvant être indexées par le thésaurus, des connaissances spécifiques à la personne interrogée dont l'affectif, l'aspect familial ou des connaissances ambigües etc. Une troisième étape consiste à relever et à lister les instances dans la transcription se rattachant aux classes sélectionnées dans notre modèle réduit (activités (*E7 Activity*), objets fabriqués (*E22 Man-Made Object*), lieux (*E53 Place*), agents (*E39 Actor*), modification (*E11 Modification*), entité temporelle (*E2 Temporal Entity*), affectation d'attribut (*E13 Attribute Assignement*), chose matérielle (*E18 Physical Thing*), choses matérielles fabriquées (E25 Physical Man-Made Thing), caractéristique fabriquée (*E25 Man-Made Feature*) et site (E27 Site) etc.). La quatrième étape consiste à relever le lien sémantique entre les différentes instances dans les tours de parole et à vérifier que notre sélection de propriété permet de réaliser ce lien. La cinquième repose sur le repérage et la précision de nouvelles classes (par exemple, *E9 Move* pour *immigration* ou pour *descente* initialement rattachées à *E7 Activity*) en s'adaptant au contexte. Une dernière étape consiste à relier les instances à un ou plusieurs descripteur(s) de thésaurus grâce à la propriété *P2 has type*.

CHAPITRE 6.Présentation des trois SOC

Ce chapitre présente les trois SOC réalisés, chacun ayant été nommés dans un souci d'identification et de cohérence : TerminoMines, ThesoMines, OntoMines. Les résultats montrent factuellement la richesse des trois SOC, qui seront ensuite évalués et discutés selon quatre critères, la couverture informationnelle, la granularité, l'expressivité des formalismes de représentations des connaissances et l'interopérabilité, dans la partie 3.

6.1 Une terminologie de la mine : TerminoMines

La ressource TerminoMines (tableau 25) dans sa dernière version (18 mai 2021) est constituée de 1378 termes-concepts pour 1 805 entrées indexées à l'aide de 737 thématiques au plus spécifique et 59 au plus générique dont la liste est en annexe 6. Au niveau sémantique, trois relations sont utilisées : la synonymie, l'association (un terme est renvoyé à un autre sans précision sur sa relation sémantique), les variantes (sans distinction sur le type, orthographiques et/ou dialectales). Le corpus de ressources lexicales est constitué des 20 ressources identifiées par un code de type « memo-mines_RL_nombre ». Lorsque cela a été possible, l'équivalent français a été noté, ceci correspondant à 289 termes. 157 synonymes dans le même bassin et 51 synonymes provenant des autres bassins miniers sont disponibles, 200 termes possèdent un renvoi vers d'autres termes, 42 termes possèdent un antonyme et 69 possèdent des variantes. Le fichier est au format ods[135] et est à aligner avec le standard TBX[136]. La figure 45 représente la volumétrie des ressources lexicales.

Tableau 25 : Extrait de TerminoMines

entrée	source bibliographi	thématique	définition	équivalent français	synonyi	renvois / voir	variante
ABANDACHE	memo-mines_RL_11	boisage&soutènement_techniques	Boisage provisoire destiné à maintenir la paroi et le	ABANDAGE	(vide)	(vide)	(vide)
ABANDAGE	memo-mines_RL_16	boisage&soutènement_techniques	Boisages pour éviter selon les cas les chutes de pi	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_19	boisage&soutènement_techniques	Boisage pour éviter les chutes de pierres à l'entrée	(vide)	(vide)	(vide)	(vide)
ABANDER	memo-mines_RL_11	actions_boisage	Boiser provisoirement pour attaquer la veine de cha	(vide)	(vide)	(vide)	ABINDER
ABATACHE	memo-mines_RL_42	actions_abattage	Action d'abattre le charbon	(vide)	(vide)	(vide)	(vide)
ABATTACHE	memo-mines_RL_11	actions_abattage	1 - action de détacher le charbon de la veine | 2 - lie	ABATTAGE	(vide)	(vide)	(vide)
ABATTAGE	memo-mines_RL_12	actions_abattage	Action de détacher le charbon de la veine.	(vide)	(vide)	(vide)	ABATAGE
	memo-mines_RL_14	actions_abattage	Travail consistant à extraire le charbon de la veine.	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_16	actions_abattage	Faire tomber des blocs, à l'aide d'un outil, tant de p	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_20	actions_abattage	Opération d'arracher le charbon au massif à l'aide d	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_21	actions_abattage	Action de détacher d'un gisement (d'une veine) le ch	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_25	actions_abattage	Opération consistant à détacher un bloc de charbon	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_27	actions_abattage	Opération consistant à détacher un bloc de charbon	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_29	actions_abattage	Action de détacher le charbon de la veine.	(vide)	(vide)	(vide)	ABATAGE
	memo-mines_RL_30	actions_abattage	Action de détacher le charbon de la veine	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_31	actions_abattage	Action d'arracher le charbon.	(vide)	(vide)	(vide)	(vide)
ABATTEUR	memo-mines_RL_19	métier_ouvrier_abattage	Mineur qui abat le charbon	(vide)	(vide)	(vide)	(vide)
	memo-mines_RL_35	métier_ouvrier_abattage	Ouvrier à la veine chargé d'abattre le charbon.	(vide)	(vide)	(vide)	(vide)
ABATTEUX	memo-mines_RL_11	métier_ouvrier_abattage	Ouvrier chargé d'abattre le charbon.	ABATTEUR	(vide)	CARBONNIER | H	(vide)
	memo-mines_RL_42	métier_ouvrier_abattage	Ou piqueur ou haveur. Ouvrier qui abat le charbon	(vide)	PIQUEUR	(vide)	(vide)
ABENDAGE	memo-mines_RL_18	résultat d'une action_boisage	Boisage qui se fait en bordure d'une devanture de v	(vide)	(vide)	(vide)	(vide)
ABENDER	memo-mines_RL_14	actions_caler	Poser la première " rallonge " à la " devanture " (ra	(vide)	(vide)	(vide)	RABINDER
ABINDACHE	memo-mines_RL_11	résultat d'une action_boisage	Boisage provisoire destiné à maintenir la paroi et le	BOISAGE PROVISOIR	(vide)	ABANDACHE |	(vide)
ABINDER	memo-mines_RL_11	actions_boisage	Boiser provisoirement pour attaquer la veine de cha	BOISER PROVISOIRE	(vide)	ABANDER |	(vide)
	memo-mines_RL_39	actions_boisage	Boiser provisoirement	(vide)	(vide)	(vide)	(vide)
ABISTOQUACHE	memo-mines_RL_11	productivité&appointements_boisage	Action d'abistoquer, de faire un travail hâtivement	ABISTOQUAGE	(vide)	(vide)	(vide)
ABISTOQUER	memo-mines_RL_11	productivité&appointements_boisage	Faire un boisage rapide et provisoire	TOQUER	(vide)	(vide)	(vide)
	memo-mines_RL_42	productivité&appointements_boisage	Effectuer un travail rapide, sommaire, se dit souven	(vide)	(vide)	(vide)	(vide)
ABLO	memo-mines_RL_11	1 ; boisage&soutènement_pièces ;	1 – pièce de bois servant d'appui pour entailler un a	(vide)	(vide)	(vide)	HABLOT | ABLOC
		2 ; transport_homme&matériel_système	1 – pièce de bois servant d'appui pour entailler un a	(vide)	(vide)	(vide)	HABLOT | ABLOC

[135] *Open Document Spreadsheet* est un format de document pour tableurs.

[136] Pour être partagées, les données doivent pouvoir être reconnues par des standards. Parmi les sous-catégories de la norme 01.020 Principes et coordination de la terminologie rédigées par l'Organisation internationale de normalisation (ISO), celle sur les « systèmes de gestion de la terminologie, de la connaissance et du contenu, TermBase eXchange (TBX) » ou ISO 30042 publiée en 2008 nous intéresse particulièrement car elle présente un grand nombre de catégories variées et normées qui permettent d'établir des relations entre les termes de manière normalisée.

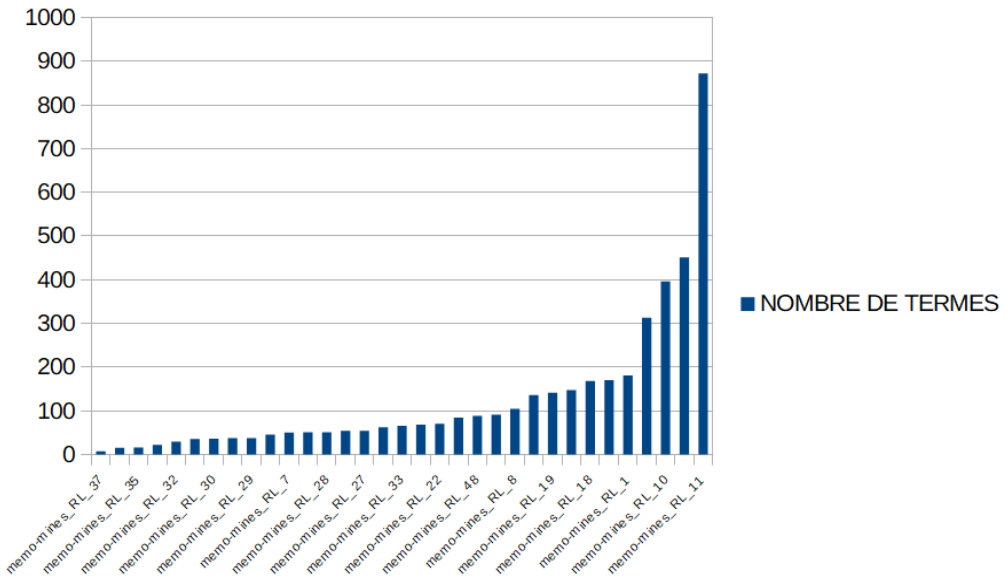


Figure 45 : Nombre de termes par ressources lexicales

En parallèle, un travail a consisté à extraire manuellement les termes des corpus audiovisuels, presse et documentaires pour enrichir la terminologie. 169 termes ont par exemple été retirés du corpus presse en tant que potentiels candidats et une première indexation libre a été établie (tableau 26). L'indexation fait la différence entre les termes qui peuvent permettre de catégoriser, les instances, les objets de mémoire et les termes miniers. L'influence du CIDOC CRM se reflète dans les indexations « appellation de personne » pour *E82 Actor Appellation*, « instance : acteur : groupe » pour *E39 Actor* de type *E74 Group*. Concernant les termes miniers, certains articles étant consacrés à ceux-ci comme celui du dimanche 7 juillet 2013, nous avons relevé les termes absents dans TerminoMines (tableau 27) ainsi que les informations telles que la langue (français ou dialecte) et d'éventuels équivalents, définitions et étymologies.

Tableau 26 : Extraction manuelle de termes dans les articles de *La Voix du Nord* et indexation. Termes non présents dans la terminologie. 144 termes

Indexation	Terme candidat	Source
catégorisation : terme minier	terme minier	VDN di7juil2013
catégorisation : acteur : spécialiste de la mine	spécialiste de la mine	VDN lu25mars2013 (2sur2)
catégorisation : appellation de personne/métier	mineur-collectionneur	VDN di15sept2013
catégorisation : appellation de personne/métier	chef de service des sondages de prospection	VDN di15sept2013
catégorisation : appellation de personne/métier	femme du mineur	VDN lu25mars2013 (1sur2)
catégorisation : appellation de personne/métier	ouvrier	VDN ve21avril2000
catégorisation : appellation de personne/métier	communauté des mineurs	VDN lu25mars2013 (1sur2)
catégorisation : appellation de personne/métier	enfant de mineur	VDN lu25mars2013 (2sur2)
catégorisation : appellation de personne/métier	ancien mineur	VDN lu25mars2013 (2sur2)
catégorisation : appellation de personne/métier : accident	rescapé	VDN lu25mars2013 (1sur2)
catégorisation : art	art	VDN lu25mars2013 (1sur2)
catégorisation : cadre de vie et sociabilité	cadre de vie et sociabilité	VDN lu25mars2013 (1sur2) ; VDN lu25mars2013 (2sur2)

catégorisation : culture minière	culture minière	VDN lu25mars2013 (2sur2)
catégorisation : équipement	appareils de production	VDN di8nov2015
catégorisation : équipement	outil industriel ; machine industrielle	VDN di8nov2015
catégorisation : équipement	équipement	VDN sa18déc2010
catégorisation : équipement	matériel roulant	VDN lu25mars2013 (2sur2)
catégorisation : équipement	chevalement emblématique	VDN ve1erjuil2011
instance : acteur : groupe	Houillères du Nord—Pas-de-Calais ; Houillères	VDN lu25mars2013 (1sur2) ; VDN lu25mars2013 (2sur2)
instance : acteur : groupe	Mission Bassin minier	VDN lu25mars2013 (2sur2)
instance : acteur : groupe : association	association Bassin minier uni	VDN lu25mars2013 (1sur2)
instance : acteur : personne	Alain Rambeau	VDN di7juil2013
instance : acteur : compagnie	Compagnie d'Anzin	VDN ve21avril2000
instance : catastrophe : date	catastrophe de Courrières en 1906	VDN lu25mars2013 (1sur2)
instance : date	21/12/90	VDN sa11déc2010
instance : document : film	Germinal	VDN ve1erjuil2011
instance : document : journal	journal des Houillères	VDN ma9janv2001
instance : document : magazine	Magazines du mineur	VDN lu25mars2013 (1sur2)
instance : document : maquette	maquette de la fosse Barrière à Anzin	VDN ve21avril2000
instance : document : publication	Mémoires de Gaillette	VDN ve21avril2000
instance : équipement	dernière berline remontée de la fosse 9-9bis le 21 décembre 1990	VDN lu25mars2013 (2sur2)
instance : évenement	nationalisation des houillères	VDN lu25mars2013 (1sur2)
instance : galerie	galerie du décor de Germinal	VDN lu25mars2013 (2sur2)
instance: objet de mémoire : gaillette	dernière gaillette	VDN sa11déc2010
objet de mémoire : géologie/sciences naturelles	formation du bassin houiller	VDN di15sept2013
objet de mémoire : géologie/sciences naturelles	faune	VDN di15sept2013
objet de mémoire : géologie/sciences naturelles	flore	VDN di15sept2013
objet de mémoire : pratique	colombophilie	VDN lu25mars2013 (1sur2)
objet de mémoire : pratique	musique	VDN lu25mars2013 (1sur2)
objet de mémoire : pratique	jardinage	VDN lu25mars2013 (1sur2)
objet de mémoire : technique/action	creusement des galeries	VDN di15sept2013
objet de mémoire : technique/action	technique d'extraction	VDN lu25mars2013 (1sur2)
objet de mémoire : technique/action	extraction charbonnière	VDN ve1erjuil2011
objet de mémoire : technique/action	déblayage des roches	VDN di7juil2013
objet de mémoire : technique/action	teneur en grisou	VDN di7juil2013
objet de mémoire : tourisme	visite du bassin minier ; visite d'un site minier	VDN lu25mars2013 (1sur2) ; VDN lu25mars2013 (2sur2)
objet de mémoire : valeur	solidarité	VDN di7juil2013
terme minier : partie de la mine	bowette	VDN di7juil2013 ; VDN ve21avril2000
terme minier : repas	briquet	VDN di7juil2013 ; VDN ma9janv2001

Tableau 27 : Extraction terminologique à partir d'un article de presse. Relevé de la langue, des équivalents, des définitions, de l'étymologie et indexation thématique – 25 termes

SOURCE	TERME	LANGUE	EQUIVALEN	DEFINITION	ETYMOLOGIE
VDN di7juil2013	BOWETTE	DIALECTE			du picard « bove » qui veut dire c
VDN di7juil2013	MUSETTE	DIALECTE			
VDN di7juil2013	SALLE DES PENDUS	DIALECTE	LAVABOS		
VDN di7juil2013	LOQUES ED' FOSSE	DIALECTE			
VDN di7juil2013	LAMPISTERIE	DIALECTE			
VDN di7juil2013	BRIQUET	DIALECTE			
VDN di7juil2013	COULONS	DIALECTE			
VDN di7juil2013	L'GARDIN	DIALECTE			
VDN di7juil2013	TROUS DE FORAGE	FRANÇAIS			
VDN di7juil2013	TIRS DES EXPLOSIFS	FRANÇAIS			
VDN di7juil2013	DÉBLAYAGE DES ROCHES	FRANÇAIS			
VDN di7juil2013	CADRES DE SOUTÈNEMENT	FRANÇAIS			
VDN di7juil2013	LAMPE BENZINE	FRANÇAIS			
VDN di7juil2013	TENEUR EN GRISOU	FRANÇAIS			
VDN di7juil2013	BOUTEFEU	DIALECTE			de l'expression du Moyen-Age «
VDN di7juil2013	PORION	DIALECTE		agent de maîtris	de l'italien « caporione », meneur

6.2 Un thésaurus du patrimoine minier : ThesoMines

ThesoMines v1.0 est disponible en accès libre sur le serveur HumaNum sur Opentheso[137] (figure 46). Les données (format *SKOS/RDF*) ainsi qu'une documentation sont téléchargeables sous demande.

Il est composé de 559 concepts descripteurs dont 543 termes génériques et 144 termes spécifiques au dernier niveau, 764 synonymes non-descripteurs, 136 notes d'indexation et 129 relations d'associations. 24 % des termes ont une note explicative, ce qui signifie que les 76% ont été jugé être transparents (au vu de leur situation dans le thésaurus). Les termes descripteurs sont à utiliser pour indexer, les termes associés pour rebondir entre sujets proches, les termes spécifiques pour identifier des instances dans les corpus et pour indexer, les termes génériques pour organiser le vocabulaire minier, ou pour indexer, les notes d'application pour une indexation guidée et les non-descripteurs pour identifier des instances dans les corpus.

En ce qui concerne la structuration, le thésaurus du patrimoine minier présente actuellement 9 points d'entrée (figure 46 à gauche sous concept) qui correspondent à un découpage mixte mettant en valeur les points saillants des analyses. Tantôt thématique, tantôt par facette, le thésaurus a pour particularité d'avoir été construit sur une base de ressources multiples et reflète plusieurs évolutions dans la méthodologie. Les entrées *histoire*, *économie*, *culture et éducation*, *environnement naturel* relèvent d'un classement thématique dont l'émergence conceptuelle a été influencée par des thésaurus génériques et thématiques comme celui de l'UNESCO ou Motbis. Les descripteurs spécifiques à l'intérieur de ces concepts proviennent d'analyses de contenu documentaire. En revanche, les entrées *activité*, *opération*, *gisement*, *bâti*, *matériel* ont été construites principalement sur la base du vocabulaire technique minier et certaines peuvent correspondre à différentes facettes définies par le *Classification Research Group* (CRG). En tant qu'actions portées par des agents humains l'entrée *opération* prend sa dénomination de cette même facette. L'entrée *activité* correspond à un environnement sémantique reliant la facette *agent* en tant que personne et les *actions/modifications* ayant un impact sur celle-ci. Nous avons ainsi le descripteur *maladie professionnelle* comme modification de l'état d'un agent, mais également l'agent

[137] https://opentheso.huma-num.fr/opentheso/api/theso/MemoMines

abatteur qui réalise l'opération d'*abattage* (descripteur présent dans opération). L'entrée méthode pourrait correspondre quant-à-elle à la facette *kinds* en tant que type de disposition naturelle du souterrain nécessitant des équipements et méthodes d'extraction spécifiques. Enfin *bâti* et *matériel* pourraient correspondre aux facettes *things/entities* en tant qu'objets construits. Les points d'entrée correspondent donc alternativement à un accès tantôt scientifique, tantôt technique du patrimoine minier (culturel et industriel). Ils sont issus de l'analyse des différentes structures classificatoires et des définitions du patrimoine culturel UNESCO et industriel du TICCIH, mais ne sont en aucun cas la seule façon de lire le domaine. Une analyse des concepts par facette est approfondie dans le cadre de l'élaboration du modèle ontologique du domaine.

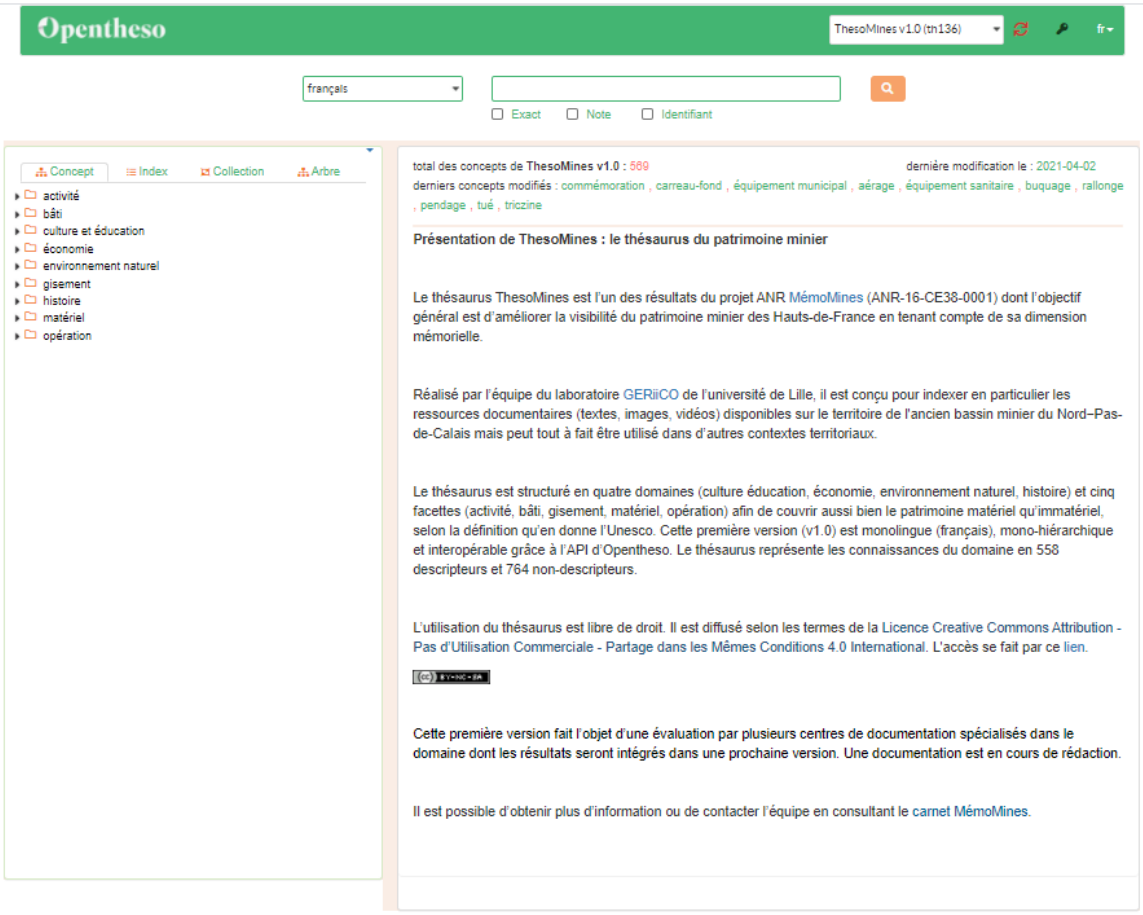


Figure 46 : Page d'accueil de ThesoMines

Les différents niveaux (cinq au plus spécifique mais une majorité de trois) permettent des points de vue plus ou moins spécifiques sur le domaine. Sous *coutume et tradition* pourront être retrouvés des documents musicaux sur la ducasse du nord ou bien un tour de parole d'un ancien mineur sur la consommation du tabac chez les mineurs. Les termes les plus spécifiques permettent d'identifier les instances dans les corpus pour l'enrichissement d'OntoMine.

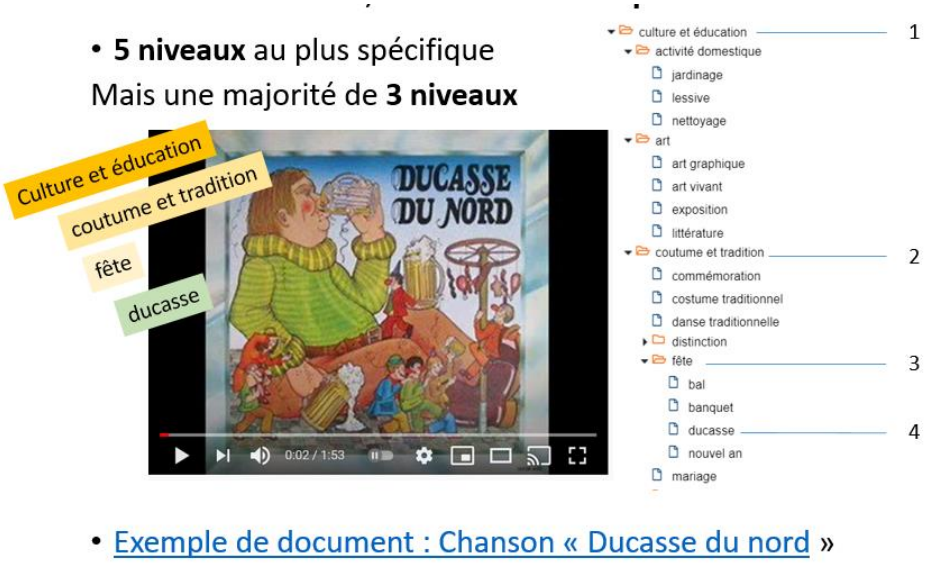


Figure 47 : Les niveaux de spécificité de ThesoMines

Le thésaurus est monolingue français (même si quelques variantes sont en dialecte lorsqu'aucun équivalent français n'existe) et monohiérarchique, c'est-à-dire que les termes n'appartiennent qu'à une seule branche.

ThesoMines ne contient pas les instances suivantes car elles sont destinées à enrichir OntoMines :

- Les noms de lieux et de sites (ex: fosse n°6 des mines de Marles)
- Les noms de personnes ou groupes (ex: compagnie d'Ostricourt etc.)
- Les dates, périodes (ex: 1989, époque Germinal…)
- Les instances des concepts (ex : église de Loos en Gohelle, catastrophe de Courrières etc.)
- Les types d'objet d'exposition (ex : maquette, plan, sculpture)
- Les types documentaires (ex : roman, poésie, presse…)

ThesoMines a été testé sur un corpus de 70 photos qui proviennent de visite de sites miniers. Pour cela, 82 descripteurs ont été utilisés. Comme il s'agissait de matériel d'exposition, le descripteur *exposition* est systématiquement utilisé. Avec le versement de ThesoMines sur OKAPI, les tours de parole ont pu être indexés avec plus de 1000 tours indexés.

6.3 Une ontologie de la mine : OntoMines

OntoMines correspond d'une part à la réduction de la version 6.2.1 du modèle du CIDOC CRM par projection des connaissances du domaine du patrimoine minier sur sa structure et d'autre part à son instanciation avec un entretien d'un ancien mineur de 45 minutes. Elle ne correspond donc pas à une ontologie de domaine mais à une ontologie de corpus. Elle a cependant le mérite d'être réalisée grâce à une analyse fine dont les résultats sont qualitatifs et de proposer un cadre d'analyse strict et rigoureux.

6.3.1 Classes et propriétés du CIDOC CRM sélectionnées pour décrire le patrimoine minier

Le modèle réduit se compose de 52 classes (surlignées en bleu sur la figure 48) et de 55 propriétés. Ce résultat ne prend pas en compte les propriétés inverses[138]. Pour le travail d'instanciation, cela permet d'alléger le modèle lors d'un travail d'annotation. Le lecteur retrouve la sélection complète en annexe 4. Sur la figure 48 qui représente la hiérarchie des classes du CIDOC CRM, les classes génériques des classes sélectionnées sont encadrées en bleu et font également partie du modèle OntoMines.

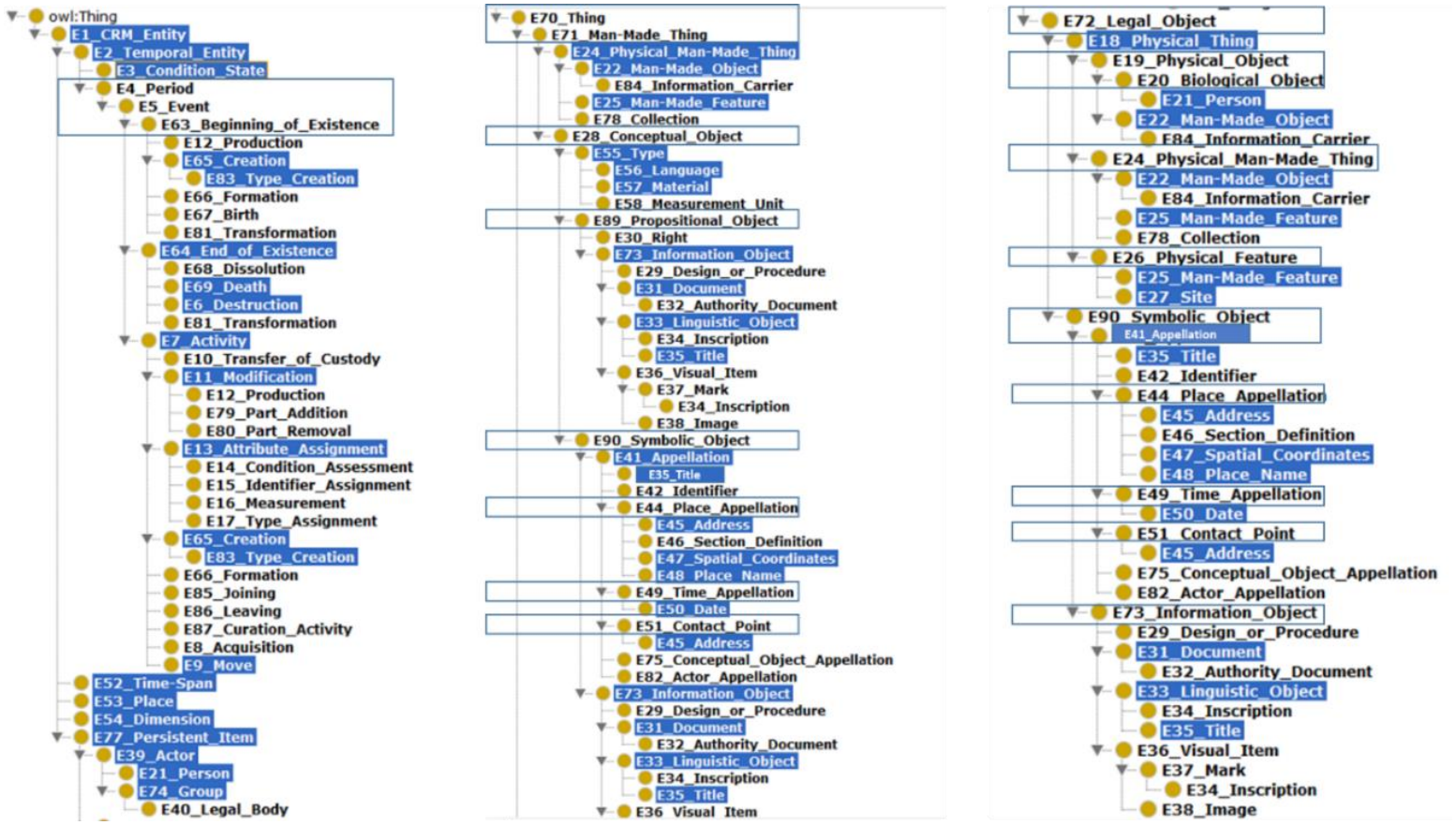


Figure 48 : Classes du CIDOC CRM sélectionnées pour représenter et décrire le patrimoine minier.

6.3.2 Représentation et formalisation des données

L'entretien avec l'ancien mineur Jean a été segmenté en tours de parole significatifs. Les tours de parole qui ne contiennent pas de référence au patrimoine minier ne sont pas instanciés. La figure 49 illustre l'instanciation du modèle pour le premier tour de parole de Jean que nous avons intitulé thématiquement « présentation de Jean ». Les instances possèdent le format suivant :

1_(2)_3

Avec,

1 : forme du descripteur ThesoMines dont le concept a été identifié puis retenu dans le tour de parole,

[138] Chaque propriété du modèle possède une propriété inverse marquée par un « i », par exemple la propriété inverse de *P32 used general technique* est *P32i was technique of*.

2 : traduction de la classe correspondante du CIDOC CRM,

3 : E désignant l'entité pour distinguer l'instance des descripteurs ThesoMines.

Chaque instance de classe est ensuite reliée à une liste de propriétés issues de la réduction du CIDOC CRM pour décrire le sens contenu dans le tour de parole, elles-mêmes reliées à d'autres instances de classe. Ainsi, l'instance *immigration polonaise_(déplacement)_E* de la figure 49 est développée sémantiquement en figure 50 (la petite flèche pointant vers le bas désigne le développement). L'instance *immigration polonaise* de la classe déplacement (*E9 Move*) est reliée au concept *immigration polonaise* de ThesoMines, puis, grâce à la propriété *P11 had participant* (a eu pour participant en français), à l'instance de la classe personne (*E21 Person*) *père de Jean*, puis à la propriété *P1 is identified by* (est identifié par) les instances *E41 polonais* et *E41 origine polonaise* qui correspondent aux mots utilisés dans l'entretien permettant d'identifier le concept d'*immigration polonaise*. Ces instances sont notées *appellation* car elles sont des instances de la classe *E41 Appellation*. Des données concernant la provenance du déplacement (*E53 Pologne*) et la date (*E50 1923*) sont également relevées et notées. La transcription de l'entretien est disponible en annexe 1. Les instances relevées ont été mises en gras dans le texte. Sur la plateforme OKAPI, les classes et propriétés peuvent être affichées dans la ou les langues fournie(s) par la version du modèle. Pour plus de transparence nous avons choisi de présenter les résultats en français. Sur la figure 49, la propriété *P161 has spatial projection* est néanmoins en anglais car elle n'est pas traduite dans la version du CIDOC CRM utilisée (Cf. 5.4.2).

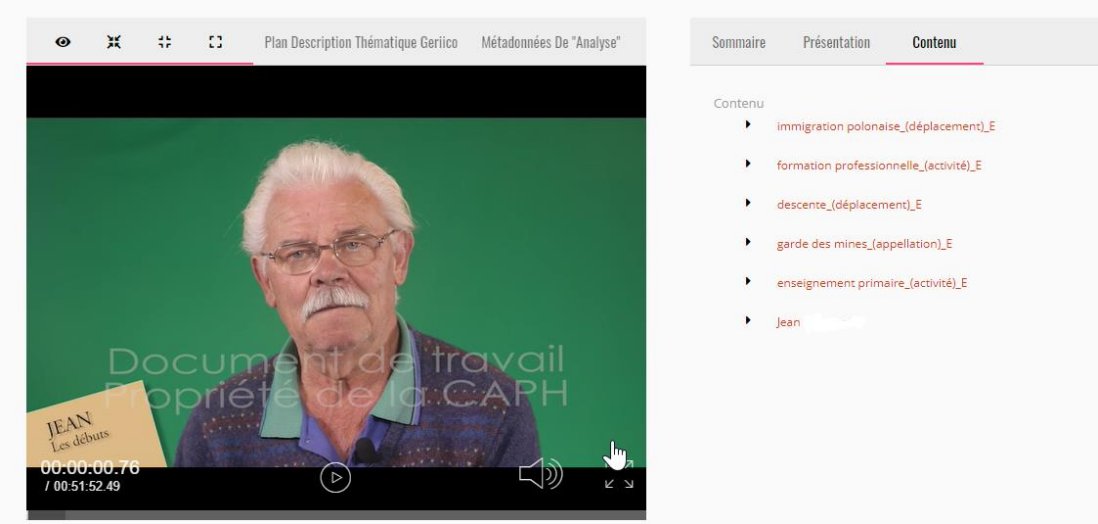


Figure 49 : Instances de classes relevées dans un tour de parole de Jean (ancien mineur)

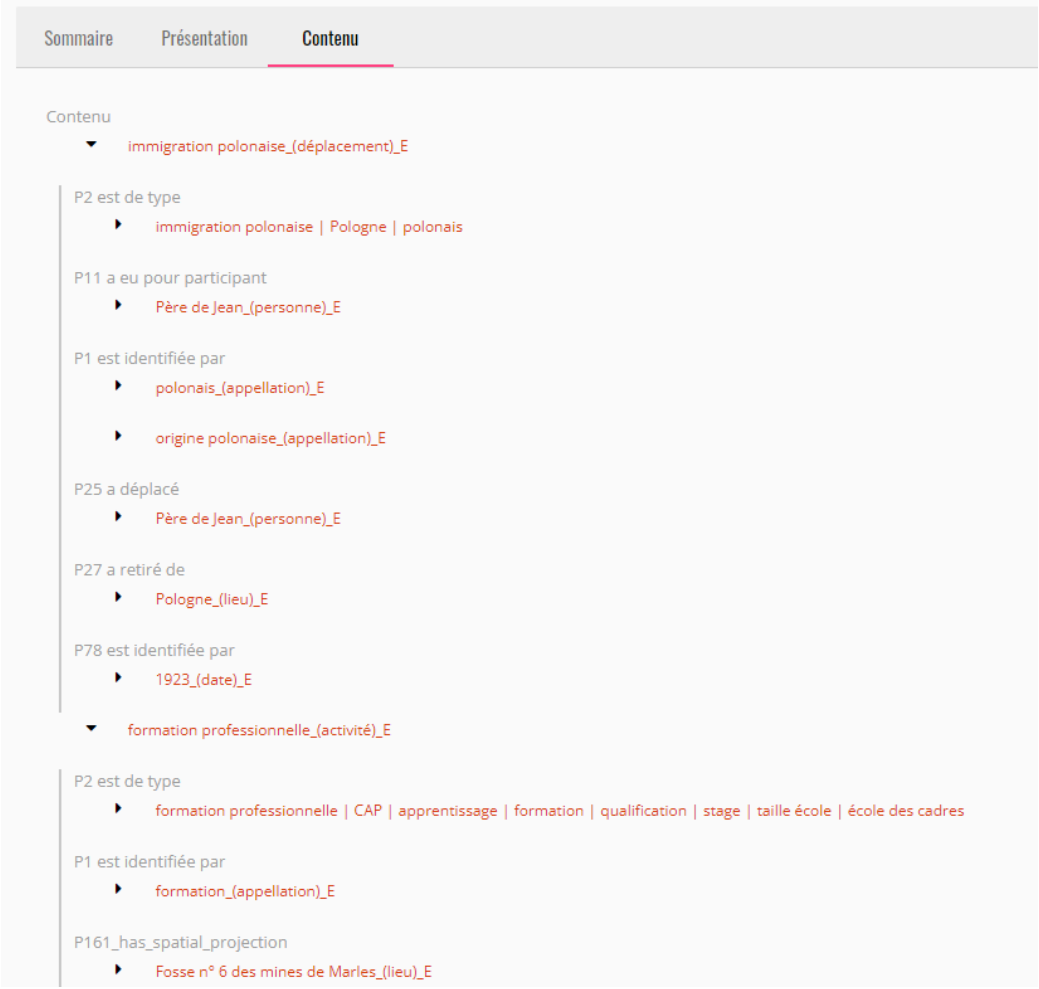


Figure 50 : Illustration d'une instanciation

La figure 51 correspond à une capture d'écran représentant l'instanciation du tour de parole *présentation de Jean*. Les symboles caméra représentent les segments de vidéo auxquels sont rattachées les instanciations, les instances de classes sont en orange, les concepts de thésaurus sont en vert et les instances non partagées dans les différents segments sont en bleu.

La figure 52 illustre l'instance *tout en haut de ch'terril* de la classe *E73 Information Object*, qui est une chanson de Renaud mentionnée par Jean, ajoutée dans les connaissances partagées[139] sur OKAPI avec l'ajout d'un lien au descripteur *musique* de ThesoMines. L'instance est décrite dans un modèle défini en amont par l'équipe d'OKAPI intitulé modèle « domaine MémoMines » permettant de ne pas mélanger les instances créées dans d'autres travaux sur la plateforme OKAPI avec celles créées pour le projet MémoMines.

[139] Les connaissances partagées sur OKAPI permettent de pouvoir appeler une même instance (même URI) lors de l'instanciation de plusieurs médias différents.

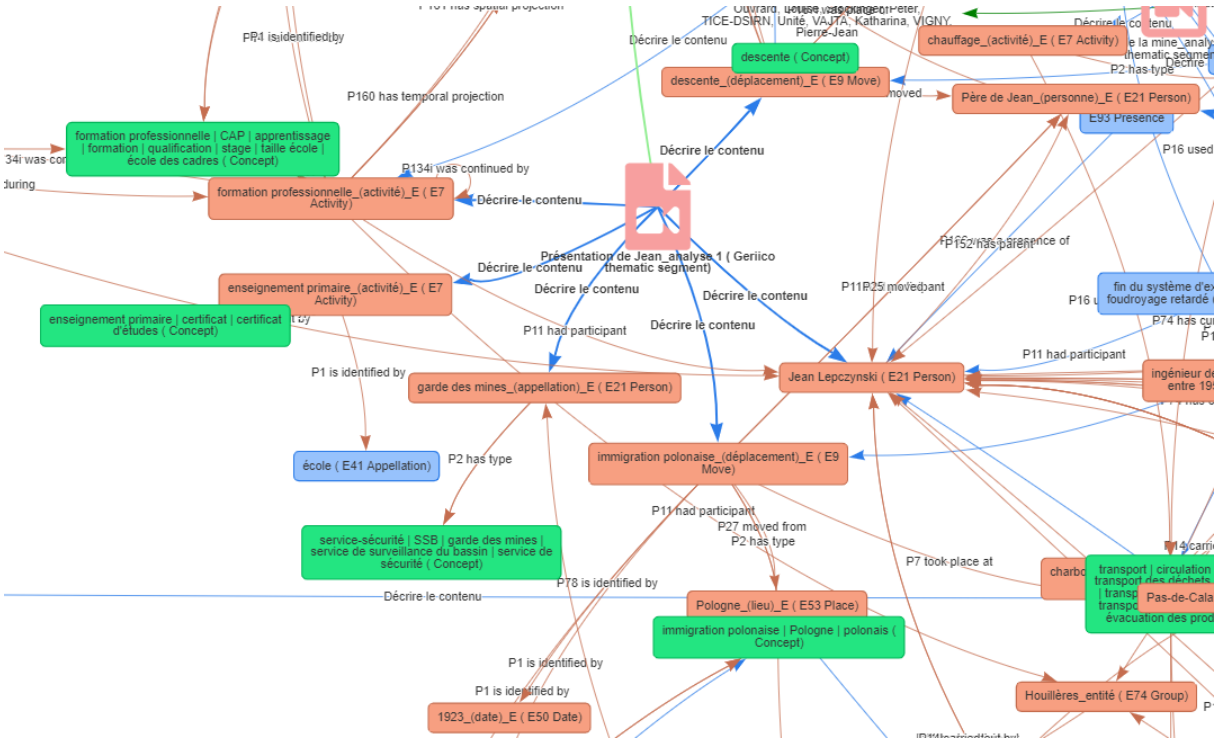


Figure 51 : Extrait du graphe représentant l'instanciation du segment « Présentation de Jean »

Figure 52 : Instance ajoutée dans les connaissances partagées sur OKAPI et ajout d'un lien au descripteur « musique » de ThesoMines

La figure 53 représente le graphe produit grâce à l'instanciation de l'entretien avec Jean. La représentation n'est pas lisible. Ceci illustre ce que Pietriga et Lee mentionnent dans leur article sur la manipulation et la visualisation des données (Pietriga et Lee 2009), il y a un réel souci dans la présentation des résultats. En revanche, il est intéressant de voir la disposition générale de l'instanciation et de remarquer que l'instance *E21 Person* de Jean se place au centre du graphe (encadré en noir). L'algorithme utilisé est ForceAtlas2, un algorithme de modélisation de graphe, disponible dans la librairie vis.js[140] développé par Jacomy et *al.* (2014). Dans cet algorithme, plus un nœud (unité de base d'un réseau) fait réseau (soit car il a un nombre important de liens entrants et sortants, soit car ses nœuds voisins ont un nombre important de liens entrants et sortants), plus il est au centre du graphe. Chaque

[140] https://visjs.github.io/vis-network/docs/network/physics.html

type de nœuds du graphe (instances, segments de document, documents, etc.) a le même poids dans le calcul des distances et positionnement. Le graphe montre que certaines instances ne sont pas reliées à d'autres dans l'entretien (en jaune). Celle en bas à droite n'est pas du tout reliée au reste du graphe contrairement à celles en haut à gauche qui sont reliées grâce à deux instances.

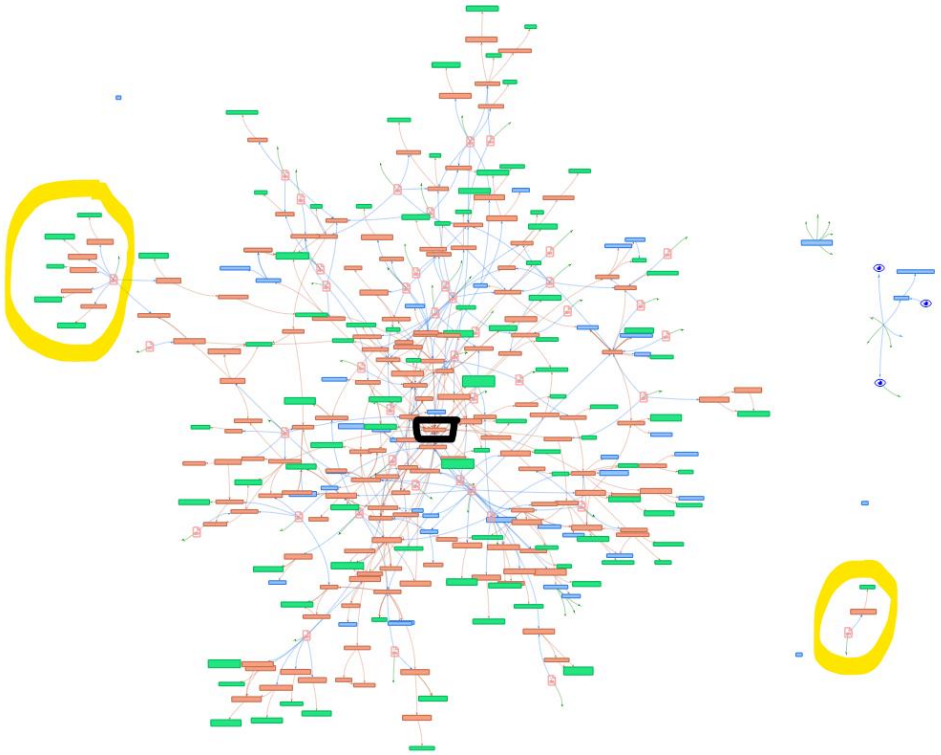


Figure 53 : Graphe représentant l'instanciation de l'entretien avec l'ancien mineur Jean, en noir au centre l'instance E21 Person, Jean

La figure 54 est un zoom de la figure 53 qui montre que l'instance au centre est celle représentant l'interviewé Jean. Cette instance possède en effet le plus de liens entrants (elle ne possède aucun lien sortant) car Jean parle de sa vie et est sujet ou l'objet de nombreuses actions et est d'une manière générale relié à un certain nombre d'événements. S'il n'est peut-être pas nécessaire de devoir représenter ce graphe pour savoir que la référence à Jean est la plus importante dans l'entretien, il n'est pas évident de représenter par la transcription ou grâce à un thésaurus le fait que certains sujets ne sont reliés à aucun autre, et inversement. Le graphe montre son potentiel d'expressivité (Cf. chapitre 9) au niveau de l'articulation des concepts dans le discours.

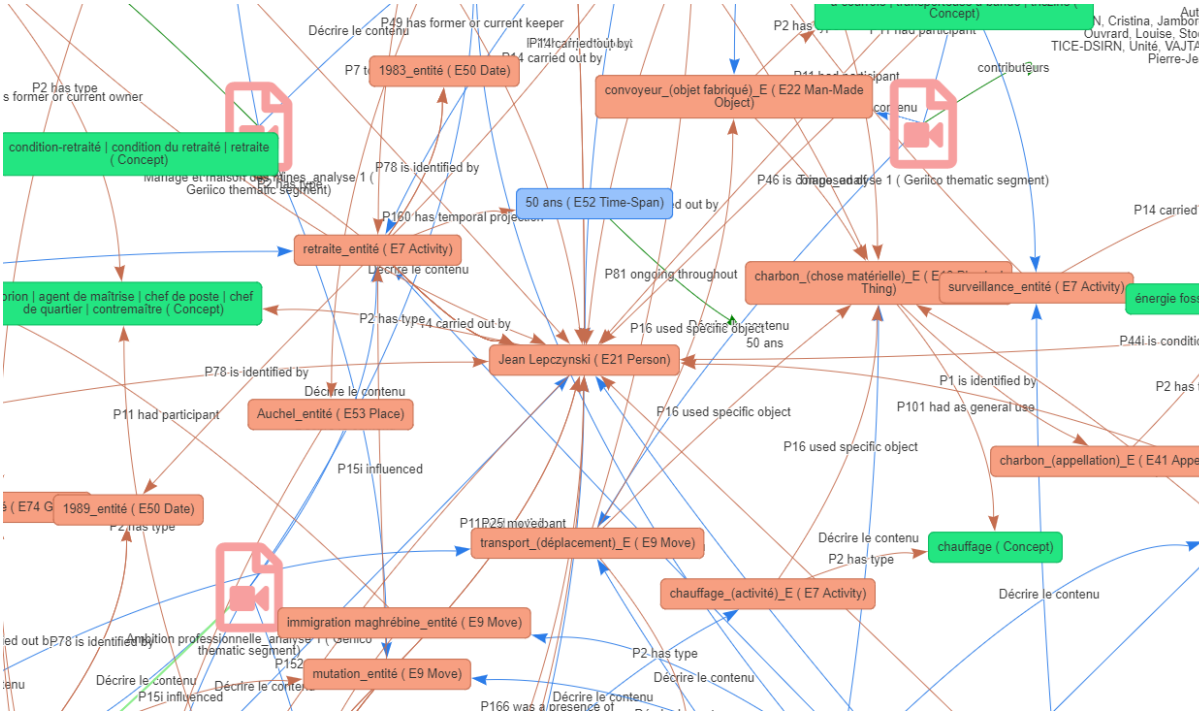


Figure 54 : Zoom sur l'instanciation, Jean est au centre du graphe

Enfin, l'ontologie a pu être téléchargée au format RDF/XML. Le fichier comprend le modèle ontologique du CIDOC CRM, la strate d'analyse de l'entretien avec Jean, les segments (tours de parole), les entités nommées, les déclarations des propriétés et classes utilisées en plus du CIDOC CRM dans la description correspondant à l'ontologie centrale de la plateforme OKAPI. Pour ces déclarations, seul le type et les labels sont repris sans les *domain* et *range*. Pour le logiciel d'édition d'ontologie que nous utilisons ensuite, Protégé[141], tout ceci est traité comme étant dans la même ontologie CIDOC CRM mais avec des espaces de noms[142] différents (Cf. 10.1.3).

6.3.3 Extrait de retranscription instancié

Afin d'illustrer l'instanciation d'entretien oral avec un ancien mineur, nous reprenons ci-dessous un extrait disponible en contexte dans l'annexe 1 du témoignage de Jean. Les classes d'OntoMines sont notées en exposant des instances en gras tandis que les descripteurs ThesoMines sont notés en indice. Les liens sémantiques entre les instances sont représentés par des crochets et les segments surlignés en gris sont ceux dont l'instanciation est effectivement représentée sur la figure 55. Le schéma instancié est représenté grâce au logiciel CMapTool[143] pour plus de lisibilité.

> « C'est grâce au **film**$^{\text{E31 Document}}_{\text{art graphique}}$, nous on peut dire un grand merci à **Claude Berri**$^{\text{E21 Person}}$ parce qu'on, si bien quand [**on a arrêté**$^{\text{E7 Activity}}_{\text{arrêt-exploitation}}$ en **mars 89**$^{\text{E61 TimePrimitive}}$] [j'étais déjà pour dire, à la **retraite**$^{\text{E7 Activity}}_{\text{condition-retraite}}$], mais je dis, qu'est-ce qu'il va devenir et après il y a eu ce **Claude Berri**$^{\text{E21 Person}}$ qui cherchait l'environnement

[141] « Protégé is a free, open-source ontology editor and framework for building intelligent systems. » https://protege.stanford.edu/
[142] Espace de noms (*namespace* en anglais) : « Partie d'un modèle de structure à l'intérieur de laquelle les noms utilisés pour désigner des fonctions, des classes ou des variables sont définis selon des règles établies par convention. » (GDT)
« Un cas particulier est celui de l'URI utilisé pour identifier un ensemble de termes, un vocabulaire, un schéma. Dans ce cas, l'URI est appelé un espace de noms, ou espace de nommage ou *namespace* » (Gandon, Faron-Zucker, et Corby 2012, 32)
[143] Le schéma est disponible à l'adresse : https://cmapscloud.ihmc.us:443/rid=1XCZ751CK-127HDWK-594VGK

pour un **film**[E31 Document] [art graphique] **Germinal**[E35 Title] parce que [**Germinal**[E35 Title] c'est quand même un **livre**[E31 Document] [littérature] qui a été écrit par **Emile Zola**[E21 Person] en **1800 et quelques**[E61 TimePrimitive]] et quand Claude Berri il a vu les vieux corons parce que les corons ils étaient encore à l'ancien style et tout parce qu'il aurait pu partir en Pologne, ça aurait revenu moins cher mais comme ici c'étaient presque des mineurs du coin alors lui il voulait des gens vrais comme nous qu'on connaissait le travail qu'on pouvait donner même des conseils alors de là c'est parti alors quand [**on a tourné le film**[E31 Document] [art graphique]] croyez-moi on était enchanté hein [trouver des **Depardieu**[E21 Person] des **Renaud**[E21 Person] des **Miou Miou**[E21 Person]] enfin il y en a au moins 7, 8 qui sont déjà décédés mais la mémoire elle reste parce que maintenant on peut dire, un type comme [**Renaud**[E21 Person]] c'était le gars euh un cœur d'or il mangeait avec nous il chantait avec nous, c'est d'ailleurs de là que c'est parti [les **chansons**[E31 Document] [musique] **en patois**[E56 Language] [langue] **qu'il chante**[E65 Creation] [musique] "**tout en haut de ch'terril**[E31 Document] ", "**viens boire une goutte ed jus**"[E31 Document]], ch'est à table qu'on chantait tous. »

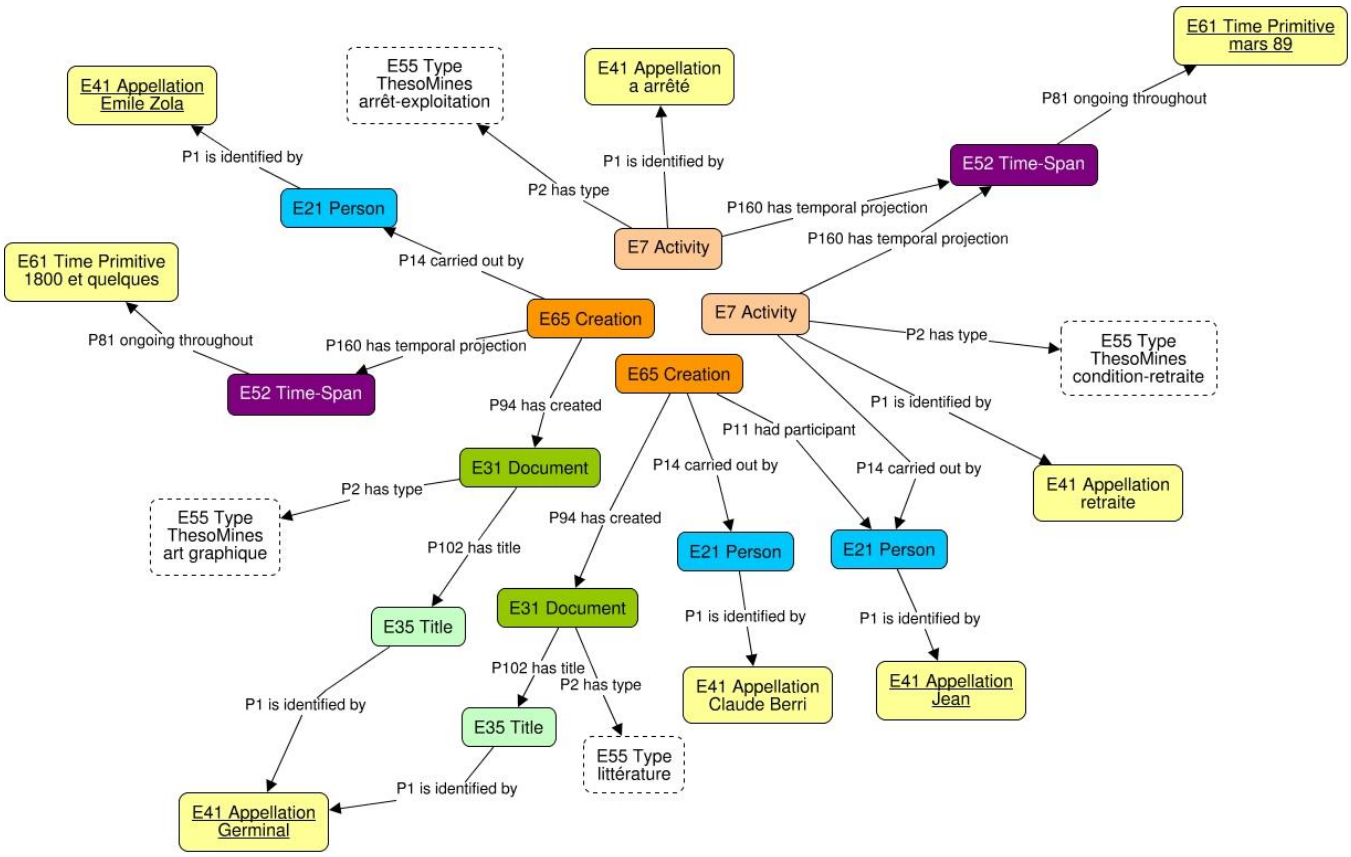


Figure 55 : Extrait de représentation de l'instanciation des *E65 Creation* et *E7 Acivity* à partir d'un extrait d'entretien

Sur le graphe de la figure 55, les instances soulignées (*mars 89*, la référence à *Jean, Germinal, Émile Zola*) montrent l'articulation du discours autour de quatre activités (*E7 Activity* et son fils *E65 Creation*) dans ce tour de parole. Les liens peuvent en effet s'effectuer directement grâce aux instances, ou grâce aux descripteurs ThesoMines qui représentent un niveau conceptuel.

CONCLUSION DE LA PARTIE 2

La partie 2 avait pour objectif de définir le patrimoine minier comme domaine de connaissances et de présenter la méthodologie de conception des trois SOC. Dans le chapitre 4, nous avons tout d'abord vu que ce patrimoine pouvait être qualifié de culturel (immatériel et matériel), naturel et industriel par les définitions de l'UNESCO et du TICCIH. Nous avons ensuite vu en quoi ce patrimoine était en danger malgré son inscription à l'UNESCO en 2012 en tant que « paysage culturel » et nous avons choisi de délimiter le domaine grâce à deux périodes spécifiques : celle de l'activité minière (du début de l'exploitation minière en 1720 à 1990, date symbolique de l'arrêt) et celle de l'après-mine (jusqu'à aujourd'hui). Nous avons ensuite expliqué pourquoi le patrimoine langagier était en danger, notamment avec la disparition progressive des locuteurs parlant un vocabulaire spécifique au territoire (chti, rouchi) et à une activité arrêtée définitivement (mines de charbon). Nous avons montré les enjeux de reconstituer fidèlement la réalité, l'ambiance du travail souterrain, souterrain qui a depuis la fin des exploitations été remblayé. Enfin, nous avons montré en quoi la documentation sur le domaine pouvait être riche mais hétérogène et dispersée sur le territoire avec un accès parfois difficile, d'où le besoin de créer de la visibilité sur ces ressources, de les identifier et de les organiser de façon à ce que leurs systèmes d'information soient interopérables.

Dans le chapitre 5, nous avons donc présenté notre méthodologie de conception des trois SOC avec la constitution d'un corpus documentaire hétérogène et la formalisation progressive des ressources lexicales à l'ontologie en passant par la conception d'un thésaurus. La constitution des différents sous-corpus a nécessité la rencontre avec les acteurs détenteurs et organisateurs de la documentation sur le domaine. Nous avons vu en quoi leurs pratiques et leurs SOC différaient d'un acteur à l'autre et l'enjeu par rapport à l'accès aux documents. L'identification des ressources lexicales spécifiques au domaine a montré la richesse du vocabulaire employé et la nécessité d'organiser celui-ci pour y accéder plus facilement et pouvoir se le représenter. Nous avons vu la différence entre les ressources lexicales (traitant des mots) et les terminologies (traitant des termes) et l'utilité de ces dernières pour la constitution d'un thésaurus. Les différentes étapes de celle-ci ont été décrites et exemplifiées à l'aide des termes, des définitions, des concepts et des notions issus du travail à partir des différents corpus. Nous avons montré notre méthode d'enrichissement du thésaurus ainsi que le logiciel utilisé pour sa conception, Opentheso. Nous sommes enfin revenue sur les étapes de passage du thésaurus au modèle ontologique du CIDOC CRM réduit et sur son instanciation à l'aide de la plateforme OKAPI de l'INALCO, utile pour faire le lien entre les instances et leur contexte d'apparition, ici un entretien d'un ancien mineur interviewé sur sa vie et son travail dans les mines.

Enfin, le chapitre 6 présentait factuellement les trois SOC réalisés : TerminoMines, ThesoMines et OntoMines au regard de résultats qualitatifs et illustratifs.

PARTIE 3. Enjeux conceptuels des SOC du patrimoine minier

Cette dernière partie aborde les enjeux conceptuels et méthodologiques de la conception des trois SOC au regard de la problématique de la valorisation d'un patrimoine. La partie 2 montrait que le patrimoine minier pouvait être considéré comme un domaine de connaissance. Dans le champ de la documentation, un domaine se définit comme « un ensemble de "concepts" considérés comme une unité dans l'activité humaine »[144]. De plus, nous avons vu en partie 1 que les concepts étaient des représentations mentales. Or, au sens philosophique, la valorisation est l'« action de donner de la valeur à un objet ou à une représentation mentale »[145] (Djambian 2010, 109). Par raisonnement logique, ceci signifie qu'en donnant de la valeur aux représentations mentales (donc aux concepts) du domaine du patrimoine minier ou à ses objets, on participe à la valorisation de celui-ci. La réalisation des trois SOC permet de formaliser ce patrimoine dans ses dimensions linguistiques, extralinguistiques et métalinguistiques (trois dimensions nécessaire à la représentation du sens) et de ce fait, de le faire exister. Les systèmes tels que le thésaurus et l'ontologie ont vocation à indexer et donc à représenter des corpus documentaires ainsi que des objets de la réalité tandis que la terminologie permet de désigner les concepts en langage naturel. Dès lors que les liens entre la réalité linguistique (le patrimoine documentaire et sa terminologie), la réalité extralinguistique (la mémoire du vécu et son ressenti) et la réalité métalinguistique (l'accès aux concepts) est réalisée, le patrimoine est perceptible, il a donc une valeur. Avant de et pour favoriser l'attractivité d'un territoire (il s'agit ici de la valorisation économique) suite à l'arrêt d'une activité, il est primordial de réfléchir au sens à redonner aux objets qui n'en ont plus.

Dans ce chapitre, les exemples illustrant les analyses sont tirés du témoignage vidéo d'un ancien mineur ayant exercé dans le Pas-de-Calais puis venu travailler à Arenberg en 1970. Celui-ci a travaillé 36 ans au fond dans les veines de charbon et a pris sa retraite en 1986. Le document, d'une durée de 51 minutes, fait partie du premier sous-corpus audiovisuel de 2013 (Cf. 5.1.4.1). Il est constitué de 127 tours de parole transcrits via le logiciel Transcriber[146] (64 tours de parole pour l'interviewé et 63 pour l'interviewer).

[144] Traduit par l'auteure à partir de la définition de domaine donnée dans le GDT par Wersig, Gernot, 1976 : « A set of "concepts" looked at as a unit in human activity. »
[145] Définition du Petit Larousse (édition de 1991)
[146] Logiciel libre développé par Claude Barras de l'Université Paris-Sud https://transcriber.fr.uptodown.com/windows

CHAPITRE 7.Couverture informationnelle

Ce chapitre étudie la couverture informationnelle des trois SOC grâce à leur recouvrement et à la saturation. Ici, la couverture informationnelle représente la composition d'un SOC. Nous l'avons vu en 1.1, l'information est « un ensemble de données dont la structure et le contexte permettent de véhiculer un sens, et donc susceptible d'informer la personne à qui les données sont transmises » (Arsenault et *al.* 2009, 105). Étudier la couverture informationnelle d'un SOC revient donc à se poser des questions quant aux données qu'il contient, à leur manière d'être représentées, d'être structurées et à leur facilité à être traitées. Il s'agit de se demander comment le contexte de ces données est représenté dès lors que celles-ci sont extraites de leur contexte d'apparition. Est-ce que la représentation du contexte correspond au contexte initial ? Qu'en est-il du sens véhiculé ? La notion de couverture implique une unité de comparaison qui dans notre cas est le domaine à représenter.

Ensuite, ici, la saturation est l'état d'un SOC tel que son enrichissement n'apporte plus de nouveaux concepts. En d'autres termes, un système ayant pour objectif de représenter un domaine pour sa valorisation arrive à saturation lorsque tout nouveau concept identifié peut être inséré sans en modifier sa structure. Un « bon » SOC devra avoir une structure assez souple. La couverture informationnelle des SOC s'analyse donc au regard du domaine à couvrir, c'est-à-dire le patrimoine houiller des Hauts-de-France, domaine que nous avons délimité sémantiquement, spatialement et temporellement. Sémantiquement tout d'abord, grâce à TerminoMines et aux définitions de l'UNESCO et du TICCIH, spatialement ensuite, grâce à la zone de cohérence paysagère définie par la Mission Bassin minier et temporellement à partir de l'événement marquant du 3 février 1720, date de la découverte du charbon jusqu'à aujourd'hui. Nous illustrons dans un premier temps l'évaluation de la couverture grâce à l'étude des données de ThesoMines au regard de celles de TerminoMines. Nous présentons ensuite le processus d'enrichissement des SOC qui permet d'évaluer leur structure et de les perfectionner à partir de l'indexation de témoignages d'anciens mineurs.

7.1 Recouvrement et différenciation conceptuels

Le recouvrement des SOC ne peut pas être évalué si nous ne distinguons pas la nature de leurs données. La forme, contrairement au sens, n'est pas une représentation mentale et elle ne peut pas être ambigüe. Elle est ce qui pourra être reconnu (par la machine et par l'homme) comme une suite de caractères identiques. Le sens, nous l'avons vu en (1.2.1), est une représentation mentale. Pour rendre cette représentation mentale visible et utilisable, le sens peut être représenté graphiquement, le signe graphique pouvant être reproduit. Représenté graphiquement, le sens est donc un outil. Ainsi, les phrases (en langage naturel) sont une façon de formaliser et de préciser notre pensée avec l'utilisation d'une proposition ou de plusieurs en relation de coordination ou de subordination. Les définitions en langage naturel, qui peuvent s'apparenter à un type de phrase, sont des objets d'étude de la lexicographie et sont une manière formelle (car construites à partir de théories) de représenter le sens d'une unité linguistique, métalinguistique ou extralinguistique que nous appelons entrée, c'est-à-dire le point d'accès à

l'information recherchée. Enfin, les termes-concepts correspondent aux formes utilisées pour regrouper plusieurs entrées dont le sens est équivalent comme nous l'avons vu en (5.2.4).

La figure 56 représente les SOC et le recouvrement de leur contenu (identifié pour plus de lisibilité par des lettres). Chaque SOC possède une forme différente qui représente une unité. La taille des formes ne représente par le poids de l'une par rapport à l'autre mais pourrait être calculée. F représente le contenu de TerminoMines, G représente le contenu de ThesoMines, H représente le contenu du modèle d'OntoMines, E représente les classes et propriétés mobilisées dans l'instanciation du témoignage de Jean, I représente le contenu d'OntoMines instanciée, D représente les liens entre les instances d'OntoMines et les descripteurs de ThesoMines (moins B), A représente le nombre de formes de TerminoMines insérées dans ThesoMines (moins B), C représente le nombre d'instances d'OntoMines issu de TerminoMines (moins B), B représente les instances d'OntoMines dont les formes et sens sont contenus dans TerminoMines et dont le terme-concept est contenu dans ThesoMines.

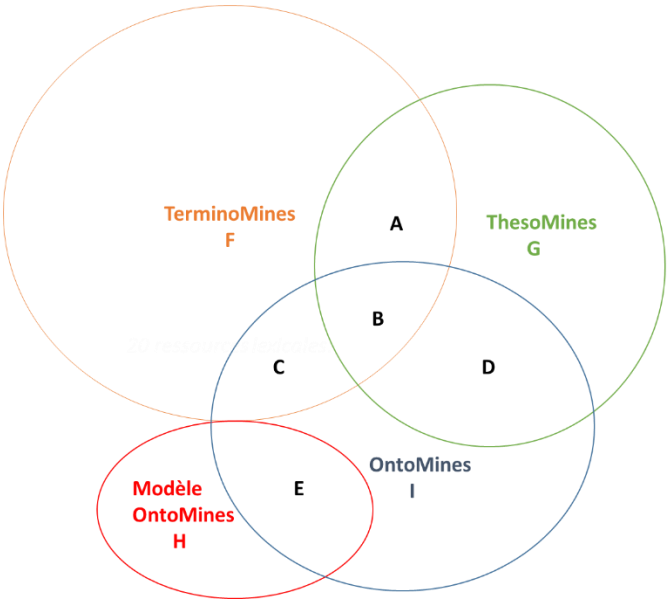


Figure 56 : Représentation du recouvrement des SOC

La figure 57 représente en détail le recouvrement de TerminoMines et ThesoMines. TerminoMines, résultat de la concaténation de 20 lexiques, contient 3076 entrées et autant de définitions, 1371 formes (nombre d'entrées dédoublonnées) et 1806 termes-concepts (formes analysées selon leur sens). Thesomines, contient 1322 formes dont 558 termes-concepts descripteurs et 764 termes non-descripteurs. 321 termes de TerminoMines (sur 1371 formes) sont insérés dans ThesoMines dont 91 comme termes préférés, donc 230 comme équivalents (EP). Cela signifie que 23% des termes de TerminoMines ont été inclus dans ThesoMines : 7% en termes préférés et 16% en EP. Ces chiffres peuvent être justifiés par la forte présence de verbes, d'expressions, de variantes orthographiques et dialectales très spécifiques dans TerminoMines ainsi que de nombreuses références au domaine de la géologie minière, domaine que nous avons choisi de ne pas représenter dans ThesoMines car il existe déjà des thésaurus pour ce domaine. Ce schéma ne représente cependant pas le nombre de termes de TerminoMines qui peuvent être indexés par les descripteurs de ThesoMines. Ce nombre peut être obtenu

une fois que chaque indexation thématique a été reliée aux descripteurs de ThesoMines (ce travail, en cours, est disponible en annexe 6).

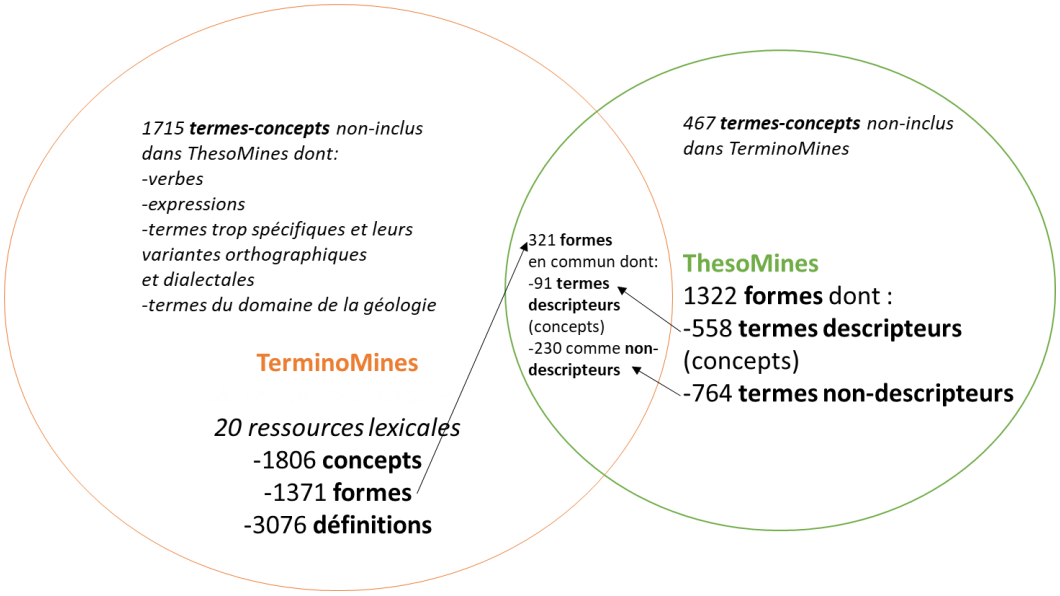


Figure 57 : Recouvrement de TerminoMines et ThesoMines

La figure 58 représente le pourcentage des termes issus de TerminoMines insérés comme descripteurs dans chacune des branches de ThesoMines. D'après ce diagramme, il apparaît clairement que les ressources lexicales ne couvrent pas la totalité du domaine et que le thésaurus complète cette couverture.

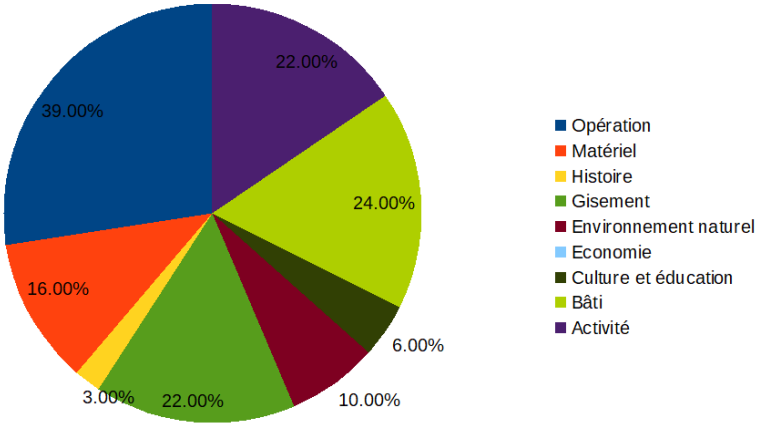


Figure 58 : Pourcentage des descripteurs issus de TerminoMines dans chacune des branches de ThesoMines

Enfin, nous l'avons vu en 6.3, le modèle d'OntoMines (réduction du CIDOC CRM) contient 52 classes et 55 propriétés. L'instanciation du témoignage de Jean a mobilisé 27 classes (donc 52% du modèle d'OntoMines) et 34 propriétés (dont 9 n'ont pas été choisies pour le modèle d'OntoMines et permettent de l'enrichir) pour décrire 331 instances. Les instances ont une forme en contexte (instances de *E41 Appellation* formalisées par un terme de TerminoMines si disponible), elles peuvent être rattachées à un descripteur de ThesoMines (les descripteurs sont des instances de la classe *E55 Type*) et sont parfois reliées en contexte par des liens (en langage naturel dans les corpus textes) décrits par des propriétés qui lient leurs classes entre elles.

Le tableau 28 répertorie le type de données que chacun des SOC contient et la nature de leur recouvrement. La mention « oui » signifie que le recouvrement des SOC contient ce type de données, « non » l'inverse. Les classes sont des types de termes-concepts et les propriétés des types de relations.

		Tableau 28 : Nature des données des SOC			
		forme	**terme-concept** **classe**	**définition** **de concept** **de terme**	**relation** **entre concept** **entre classe (propriété)** **entre terme**
Termino-ThesoMines	A	oui	oui non	oui oui oui	oui oui non oui
Termino-Theso-OntoMines	B	oui	oui oui	oui oui oui	oui oui oui oui
Termino-OntoMines	C	oui	oui oui	oui oui oui	oui oui oui oui
Theso-OntoMines	D	oui	oui oui	oui oui oui	oui oui oui oui
Modèle Onto-OntoMines	E	oui	oui oui	oui oui non	oui oui oui oui
TerminoMines	F	oui	oui non	oui non oui	oui non non non
ThesoMines	G	oui	oui non	oui oui oui	oui oui non non
Modèle OntoMines	H	oui	oui oui	oui oui non	oui oui oui non
OntoMines	I	oui	oui oui	oui oui non	oui oui oui oui

Ce tableau montre l'hétérogénéité des données des trois SOC et l'importance de les distinguer lors du calcul de recouvrement. Il montre également que la couverture d'OntoMines sur TerminoMines ou sur ThesoMines est la combinaison qui permet d'avoir le plus de données hétérogènes, par conséquent d'obtenir plusieurs points de vue différents sur le domaine (et différents points d'accès).

7.2 Évaluation de la saturation

7.2.1 Saturation absolue des SOC

La figure 59 représente la configuration idéale des SOC lorsqu'ils sont arrivés à saturation, c'est-à-dire qu'aucun nouveau concept n'est identifié dans la documentation du domaine. OntoMines (I) est représentée par l'ensemble du modèle OntoMines (H), TerminoMines (F) et ThesoMines (G). OntoMines doit permettre la représentation du domaine du patrimoine houiller des Hauts-de-France.

Elle doit avoir pu instancier, nommer et décrire au moins un de chaque type d'objet du domaine et ayant pu nommer et décrire au moins un lien de chaque type entre chaque instance d'OntoMines. Ces instances doivent pouvoir être représentées par les termes de TerminoMines et indexées par les concepts de ThesoMines. Chaque SOC est représentatif d'un point de vue différent sur le domaine. TerminoMines représente tous les termes techniques issus de la documentation faisant mention des réalités du domaine. Par l'indexation des termes, ThesoMines indexe la documentation. Le modèle OntoMines doit pouvoir instancier, nommer et décrire toutes les données de TerminoMines et ThesoMines. Nous l'avons vu, la figure 56 représente OntoMines instanciée par un seul document, elle est loin d'être arrivée à saturation.

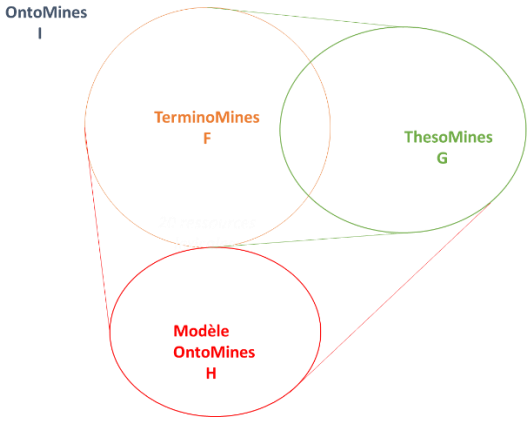


Figure 59 : OntoMines saturé

7.2.2 Point de vue sur le domaine

Si l'UNESCO et le TICCIH fournissent des définitions pour le patrimoine culturel, naturel et industriel, ces instances restent souples quant à leur interprétation. Pour construire ThesoMines, nous avons choisi d'étudier le langage utilisé par les mineurs, vecteur du patrimoine immatériel. Notre volonté étant en effet de sauvegarder la mémoire de ceux-ci, il nous a semblé intéressant de concevoir un outil permettant de le mettre en lumière dans la documentation. Pour cela, l'utilisation de ressources majoritairement réalisées par les anciens mineurs ou à partir de témoignages d'anciens mineurs comme ressources pivot, nous a paru primordial. Nous l'avons vu, le résultat se regroupe sous neuf termes de tête[147] qui représentent la couverture sémantique de ThesoMines. Nous revenons ci-dessous sur la signification de chacune des branches au regard des définitions de l'UNESCO et du TICCIH :

Bâti représente les monuments ou édifices, les ensembles de constructions du patrimoine culturel ou les bâtiments, les structures et infrastructures de transport, les lieux utilisés pour des activités sociales en rapport avec l'industrie du patrimoine industriel. D'un côté, l'architecture industrielle considérée comme l'ensemble des bâtiments industriels fait partie des caractéristiques essentielles du bassin minier du Nord–Pas-de-Calais. L'exploitation minière a requis la construction de bâtiments pour l'exercice de différentes fonctions liées à l'activité minière, la construction d'infrastructures nécessaires à la manutention des produits, à l'extraction etc. D'un autre côté, le patrimoine culturel minier prend en considération les activités culturelles et sociales pratiquées par la population, exercées dans des

[147] Le terme de tête est le terme préférentiel qui représente le concept de plus haut niveau dans la hiérarchie à laquelle il appartient.

bâtiments spécifiques. Enfin, le patrimoine minier prend en compte les logements des employés comme preuve d'un microcosme autour de l'activité même. Des 353 éléments inscrits sur la liste du patrimoine mondial de l'UNESCO en 2012, 14 types de bâtis sont considérés comme tels par la Mission Bassin minier sur 58 communes différentes (ce sont des équipements, des édifices, des fosses, des monuments, des salles des fêtes (classés monuments historiques ou non) etc. Dans ThesoMines, nous distinguons trois types de bâtis, s'opposant entre eux par leurs fonctions ou leurs dispositions :

- le bâti civil : par opposition fonctionnelle avec le bâti du carreau de surface et le bâti du carreau souterrain. Ces deux dernières catégories concernent l'activité de l'exploitation minière (le travail).
- le bâti du carreau de surface : infrastructures (équipements ou installations) nécessaires à l'activité d'exploitation minière situées en surface et non obligatoirement disposées dans l'enceinte de l'activité même.
- le bâti du carreau souterrain : selon les époques, ouvrages et infrastructures en cours de construction, déjà construits ou ayant disparus se situant dans le souterrain.

Activité représente les valeurs du patrimoine industriel qui est « le témoignage des activités qui ont eu et qui ont encore des conséquences historiques profondes » (TICCIH 2003, 1). Ceci regroupe les risques et nuisances liés à l'exercice de l'activité minière tels que les accidents de travail ou la mort, la réglementation mise en place pour éviter ces risques, les types de métiers et tout ce qui relève de la délimitation dans le temps des tâches ou de poste précis ainsi que du repos.

Matériel ne représente pas que les machines (TICCIH 2003, 1), mais tous les types d'outils, d'équipements nécessaires à l'activité d'exploitation du charbon, de l'organisation à la construction de dispositifs de sécurité à l'accès au gisement, à l'exploitation, au transport du charbon et à son traitement permettant sa vente qui ont une valeur historique.

Culture et éducation indexe les réalisations artistiques (œuvres), les activités liées à la vie domestique des femmes de mineurs, les habitudes festives et alimentaires ou de consommation spécifiques, la formation du début de la scolarisation à la formation professionnelle pour devenir mineur et l'évolution de son métier, l'aspect linguistique, les événements religieux, la pratique de sport et de loisirs spécifiques, la vie des associations, les événements réalisés par elles et la documentation associée. (UNESCO 2003a, 2-3)

Économie représente l'impact économique de l'activité aussi bien dans la période de prospérité que dans son déclin, les actions de redynamisation du territoire comme le tourisme et des activités comme la création de nouveaux emplois. Ceci s'insère dans les problématiques de maintenance et de conservation énoncées dans la charte du TICCIH : « continuer à adapter et à utiliser des bâtiments industriels évite des pertes d'énergie et contribue à pérenniser le développement économique. Le patrimoine industriel peut jouer un rôle important dans la régénération de régions sinistrées ou en déclin. La continuité que le

réemploi implique peut fournir un équilibre psychologique aux communautés confrontées à la perte soudaine d'emplois durables » (TICCIH 2003, 4).

Environnement naturel représente le patrimoine naturel (UNESCO 1972, 2) et indexe la composition et la délimitation des ensembles formant l'ancien bassin minier du Nord–Pas-de-Calais. Cela prend en compte : le classement de celui-ci au patrimoine de l'UNESCO en tant que *paysage culturel évolutif*, les types d'énergie utilisés pour l'exercice de l'activité, la consommation de cette énergie, les questions écologiques avec l'équilibre et la réhabilitation de l'espace et les risques naturels ainsi que leurs enjeux.

Gisement représente des « vestiges de la culture industrielle qui sont de valeur […], sociale, […] ou scientifique » (TICCIH 2003, 1) et indexe les différentes méthodes d'exploitation employées dans les mines (étudiées au niveau scientifique) et le type d'inclinaison du gisement, variant selon les lieux et modifiant la difficulté du travail des mineurs (phénomène étudié au niveau social).

Opération indexe les techniques et processus employés pour pouvoir exploiter le charbon en sécurité et correspond aux enjeux de conservation et de maintenance énoncés dans la charte du TICCIH : « Les savoir-faire impliqués dans de nombreux processus industriels, anciens ou obsolètes, sont des sources d'importance capitale dont la perte peut être irremplaçable. Elles doivent être soigneusement enregistrées et transmises aux jeunes générations. » Les opérations sont réalisées à l'aide des concepts *matériel* et des types de métiers sous *activité*.

7.3 Processus d'enrichissement : démarche évaluation et perfectionnement

Nous présentons dans cette section le processus d'enrichissement des SOC qui permet d'évaluer leur structure et de les perfectionner à partir du corpus de témoignages transcrits des anciens mineurs. Une phase d'évaluation de ThesoMines a ainsi été réalisée grâce à l'indexation automatique d'un corpus audiovisuel d'une part et à la projection de TerminoMines sur le deuxième sous-corpus audiovisuel défini en 5.1.4.1 d'autre part.

7.3.1 Test par indexation automatique d'un corpus

Afin de tester la couverture de ThesoMines, nous avons indexé le corpus audiovisuel déposé sur la plateforme OKAPI, qui est constitué des entretiens effectués avec six anciens mineurs ayant travaillé à Arenberg (réalisés en 2013), ainsi qu'une visite du site minier d'Arenberg par un autre ancien mineur (2018). 2384 tours de parole ont été transcrits (chaque nouvelle prise de parole par l'interviewé est considéré comme un tour de parole, les prises de parole par l'interviewer ne sont pas prises en compte). Plus de 1000 tours de parole ont été indexés avec les descripteurs de ThesoMines sur OKAPI, ce qui représente environ 41% du corpus. La figure 60 illustre l'indexation avec le descripteur *accrochage* qui apparaît dans cinq tours de parole de deux anciens mineurs. Dix occurrences du terme font émerger plusieurs contextes utiles à l'enrichissement des trois SOC. L'indexation des tours de parole via la plateforme OKAPI permet d'avoir un lien entre le descripteur de thésaurus, la prononciation des termes à partir du média et un contexte court d'énonciation correspondant au tour de parole.

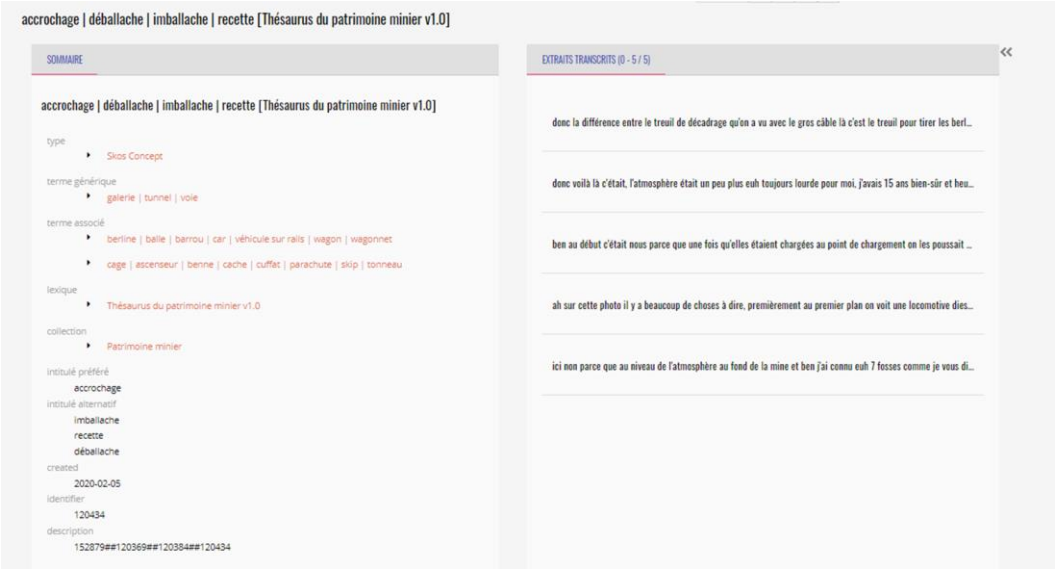


Figure 60 : Cinq tours de parole indexés par le descripteur *accrochage* et ses équivalents sur OKAPI

L'indexation étant réalisée en plein texte avec les formes des descripteurs et non-descripteurs (c'est-à-dire à partir des chaînes de caractère de ceux-ci), elle n'est donc pas entièrement représentative du contenu conceptuel des médias et fait parfois apparaître des soucis d'ambiguïté, comme le montre la figure 61 pour le descripteur café inséré dans ThesoMines comme établissement mais apparaissant comme boisson dans le corpus. Une note d'application est d'ailleurs donnée dans ThesoMines, à café : '*Pour indexer le café comme boisson, utiliser le descripteur "boisson"*'. Si ThesoMines seul permet d'accéder au contenu de médias transcrits via la forme, son indexation n'est sémantiquement pas assez précise. Pour que celle-ci y soit d'avantage, il faudrait par exemple ajouter aux non-descripteurs toutes les formes fléchies des termes.

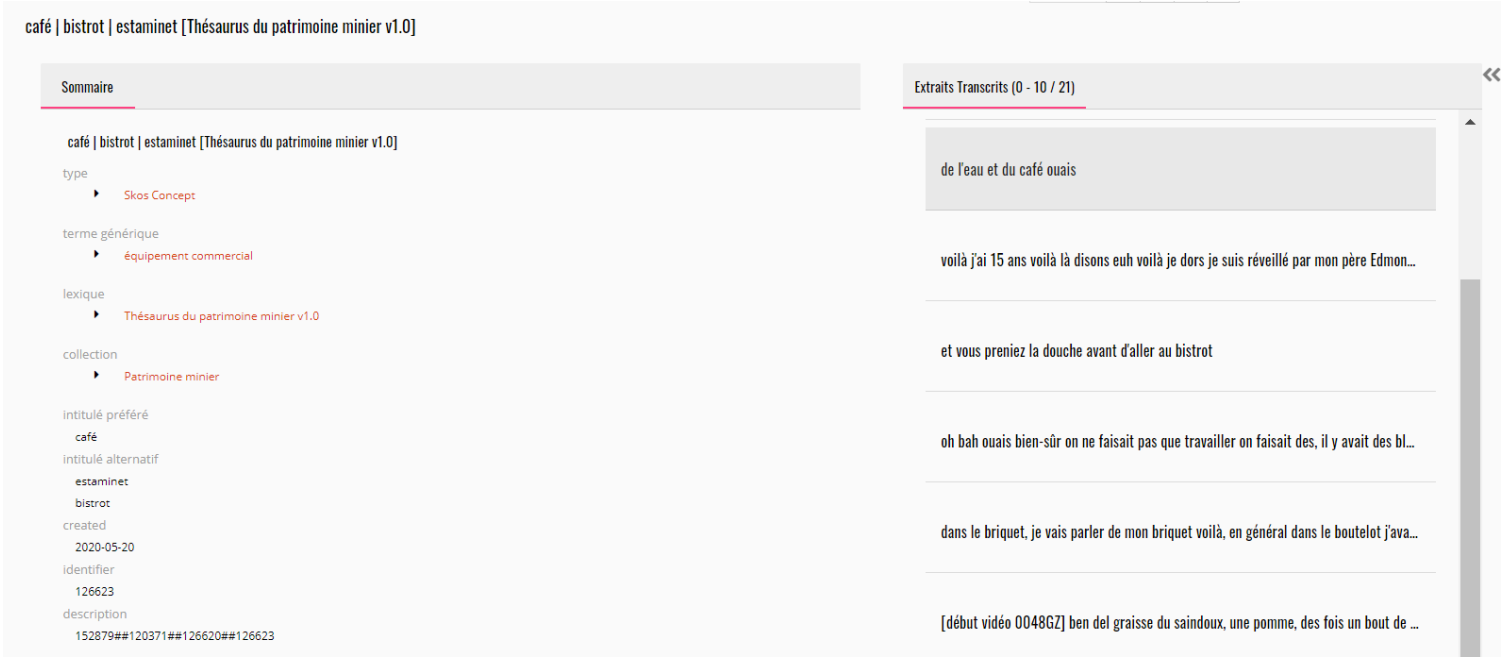


Figure 61 : Ambiguïté du descripteur *café*

7.3.2 Test de la souplesse de la structure

Afin de tester la souplesse de la structure de ThesoMines, une phase d'évaluation est ensuite réalisée grâce à la projection de TerminoMines sur le deuxième sous-corpus audiovisuel défini en 5.1.4.1. De nouveaux concepts sont également relevés.

Lors de cette phase, 109 nouveaux termes sont introduits dans la structure (en descripteur ou en EP). L'introduction des termes nécessite parfois l'enrichissement de la structure. Par exemple, le terme *médaille* présent dans le témoignage de Raymond, désigne la distinction et l'objet remis aux mineurs ayant travaillé un certain nombre d'année. Celui-ci, représentant le concept général des distinctions honorifiques et la numismatique, justifie l'introduction du terme *distinction* sous *coutume et tradition* qui était jusqu'alors absent.

Des concepts non inclus dans TerminoMines mais présents dans tous les entretiens sont également relevés. Par exemple, nous décidons que le concept de *famille*, bien qu'important, n'enrichit pas ThesoMines car il est considéré comme ne relevant pas directement du domaine. Le fait de ne pas inclure certains concepts importants dans ThesoMines nécessite l'analyse du modèle OntoMines et de se demander si cette réalité peut être décrite par OntoMines. Dans le cas du concept *famille*, des personnes peuvent être décrites comme des membres d'une même famille avec la propriété *P152 has parent (is parent of)*.

D'autres concepts comme les émotions, les sentiments, les valeurs et comportements ne sont pas non plus inclus dans ThesoMines, bien que la peur, la joie, l'amitié, la solidarité soient régulièrement mentionnés dans les entretiens. L'aspect émotionnel est une partie intégrante du patrimoine et ceci peut être une limite dans la couverture des SOC car de son côté, OntoMines ne permet pas d'exprimer ces réalités. Il faudrait donc certainement intégrer une nouvelle branche *psychologie*, sous *culture et éducation* dans laquelle nous placerions *émotion*, *sentiment*, *comportement* qui pourraient représenter l'environnement émotif et psychologique du mineur (figure placée au centre des SOC). Le rôle du thésaurus apparaît ici clairement : son enrichissement peut pallier des manques au niveau du modèle du CIDOC CRM.

Une autre réalité qui n'est pas complètement représentée, nous l'avons vu, est celle du domaine de la *géologie*. Même si celle-ci est déjà représentée dans d'autres ressources classificatoires et que les mineurs interrogés n'en parlent pas de la même manière que les scientifiques, ce domaine fait néanmoins partie intégrante du patrimoine minier. Des termes associés à ce domaine sont toutefois insérés. Les branches *méthode-exploitation* et *gisement* ont d'ailleurs fait l'objet de nombreuses discussions sur leur insertion dans ThesoMines. Au moment de la phase d'évaluation, la structure hiérarchique concernant les méthodes (d'exploitation) était :

méthode
- taille
  - tranche
  - sous niveau
  - chambre pilier
  - chambre magasin
- morphologie
  - dressant
  - plateure
  - semi-dressant

Or l'étude des témoignages montrait que le terme de *taille* (seul) était toujours utilisée en synonyme de *chantier* ou *chantier d'abattage* et jamais comme type de méthode comme cela était structuré dans le thésaurus d'Auclaire (Auclaire 1990), structure sur laquelle nous nous étions initialement appuyées. Nous avions de plus déjà inséré le terme *chantier* sous *bâti*. Ceci engendrait donc de l'ambiguïté. Concernant le terme *tranche*, nous l'avons finalement exclu du thésaurus car polysémique : c'est en effet l'entaille réalisée pour délimiter les blocs de charbon à abattre, mais aussi un morceau (de lard) ou encore un outil. Enfin, nous nous sommes posées des questions sur le concept de *méthode* au regard de celui de *morphologie* et de leur lien sémantique. La morphologie désignait plutôt l'*inclinaison* de la veine de charbon, désignée par les termes régionaux suivants par les anciens mineurs : *dressant* (couche verticale), *plateure* (couche plate) et *semi-dressant* (couche inclinée). Entre le terme *morphologie* et *méthode*, le lien sémantique n'est pas hiérarchique, mais associatif. La morphologie de la veine a une influence sur la méthode d'exploitation, tout comme la profondeur de la veine de charbon par exemple. Pour représenter ce concept de *veine*, terme qui peut être polysémique, nous avons donc introduit le terme *gisement* qui, s'il est peu utilisé par les anciens mineurs, représente cette idée de « position, disposition naturelle des couches (souterraines) de minéraux » (TLFi).

Le concept de *méthode*, peut signifier plusieurs choses. Dans le plan de classement du CERCHAR (CERCHAR 1976) les méthodes d'exploitation prennent en compte l'exploitation des *couches* de charbon. Les *couches* de charbon peuvent être prises en une, voire plusieurs *tranches* (nous avons vu que nous n'incluons pas cette spécificité polysémique dans le thésaurus). Ici *couche* est équivalent à *gisement*. Le CERCHAR classe également sous méthodes, le creusement des galeries, le remblayage, le foudroyage, le fonçage des puits. Or, nous avons placé ces trois concepts sous *opération* avec les descripteurs respectifs *creusement*, *remblayage* et *foudroyage*. Ces termes peuvent en effet représenter des *méthodes* ou des *opérations*, nous choisissons de les spécifier comme des suites d'actions[148] donc des *opérations* qui impliquent la notion d'agent (le mineur) et peuvent être instanciées par la classe *E7 Activity* d'OntoMines. Enfin, nous éliminons le terme « sous niveau » qui n'apparaît jamais dans le corpus et ajoutons le terme *longue-taille* qui désigne la méthode d'exploitation par taille et qui caractérise la longueur de la taille (les anciens mineurs mentionnent parfois cette longueur), le terme est également une vedette-matière de RAMEAU[149]. Enfin, le terme *charbon* est placé en EP d'*énergie*

[148] Opération : « Suite d'actions plus ou moins complexes supposant la combinaison de moyens et de ressources pour obtenir un résultat déterminé. « (GDT, 2010)
[149] https://catalogue.bnf.fr/ark:/12148/cb14499388h

fossile et est associé au terme *gisement*. Cette évaluation permet d'obtenir la structure actuelle suivante où *gisement* est terme de tête dans ThesoMines :

gisement EP couche, veine TA énergie fossile
- méthode-exploitation EP méthode d'exploitation
  - chambre-pilier EP méthode d'exploitation par chambres et piliers
  - chambre-magasin EP méthode d'exploitation par chambres-magasins TA voie-base, voie-tête
  - longue-taille EP méthode d'exploitation par taille, taille chassante, taille rabattante
- inclinaison
  - dressant
  - plateure EP tranche horizontale
  - semi-dressant EP tranche inclinée, tranche montante, tranche descendante

D'autres questions se posent alors quant au terme *gisement* placé au même niveau que *culture et éducation*, *histoire* ou *économie*. En effet, *gisement* est un terme de granularité plus spécifique que ces derniers. Étant donné le peu de termes désignant le domaine de la *géologie* insérés dans ThesoMines, nous avons préféré laisser ce terme en tête. Enfin, le lien sémantique entre *gisement* et *méthode-exploitation* n'est toujours pas hiérarchique mais associatif. Ceci ne dérange pas dans la mesure où OntoMines permet de spécifier plus en détail ces relations sémantiques. Ceci pourrait poser problème dans le cas d'une indexation automatique.

Tandis que ce travail d'évaluation du thésaurus permet d'enrichir TerminoMines avec des termes pertinents trop spécifiques pour être insérés dans ThesoMines, il enrichit également OntoMines et permet de valider le modèle. L'instanciation du témoignage de Jean a en effet permis de mobiliser 27 classes (donc 52% du modèle d'OntoMines correspondant à 46% du modèle du CIDOC CRM v.6.2) et 34 propriétés pour décrire 331 instances. Parmi ces 34 propriétés, 9 propriétés du CIDOC ont été jugées nécessaires d'être ajoutées au modèle OntoMines pour représenter la réalité du domaine. 45% des propriétés d'OntoMines ont donc été utilisées et 22% du CIDOC (34 propriétés sur 149).

Lors de cette phase d'enrichissement et d'évaluation, chaque nouveau descripteur inséré dans ThesoMines nécessite d'être évalué en fonction des schémas déjà réalisés (en annexe 4, ces descripteurs se situent sous « Descripteurs ajoutés »). Dans certains cas, de nouveaux schémas sont réalisés. 8 nouvelles classes ont été ajoutées et décrites en détail (dont *E3 Condition state*, *E9 Move*, *E28 Conceptual object*, *E54 Dimension*, *E64 End of existence*, *E69 Death*, *E72 Legal object*, *E73 Information object*). Les schémas de 5 classes sur 15 sont modifiés.

L'enrichissement continu des SOC jusqu'à saturation grâce à l'analyse des documents agit sur la structure de ceux-ci. La saturation des systèmes dépend du niveau de granularité des SOC.

CHAPITRE 8. Finesse de description ou granularité

Ce chapitre étudie le niveau le plus fin de description des trois SOC. Ces unités minimales de sens sont appelées instances (ou individus) et représentent la granularité de chaque système. Les instances peuvent être regroupées par concept, et dans certains cas ces concepts sont des classes : d'un côté nous avons les instances de concept, ce sont les « concepts feuilles », les plus spécialisés du SOC, d'un autre les instances de classe. Par exemple, *berline* pourrait être une instance de la classe *matériel* et instance du concept *transport*. La différence entre les deux est que *matériel* fait référence à berline en tant qu'objet tandis que *transport* l'insère en tant que fonction. Si l'instance est une unité linguistique désignant un élément précis de l'univers du discours rattachée à un référentiel, celle-ci est appelée entité nommée. C'est le cas par exemple des noms propres, des références calendaires, des mesures etc. Nous verrons que ces instances sont traitées différemment dans les trois SOC. Tandis que TerminoMines et ThesoMines ne contiennent pas ou très peu d'entités nommées (que nous divisons en entités temporelles, entités spatiales et acteurs), OntoMines permet, grâce au modèle du CIDOC CRM, de les décrire d'une manière assez précise. Nous verrons également comment sont traités les objets matériels et les activités au regard des entités nommées dans le témoignage de l'ancien mineur Jean. La segmentation de l'entretien en tour de parole permet de réfléchir à l'impact de ce choix sur la granularité d'instanciation et sur la représentation sémantique des connaissances.

Le terme « granularité » vient de *granulaire* avec le suffixe *–ité*, qui dérive lui-même du latin *granulum* « petit grain » avec le suffixe *–aire*. Dans granule, il y a l'idée d'une valeur infime, d'une petite quantité (mais pas forcément la plus petite quantité) de quelque chose de plus grand. La granularité pose question lors de l'interopérabilité entre systèmes : comment faire fonctionner ensemble deux systèmes qui traitent du même objet mais à différents niveaux ? Il est primordial de traiter celle-ci pour éviter les erreurs et pouvoir effectuer des alignements corrects.

La notion de granularité est donc reliée à celle de système. Le système correspond à un tout, un ensemble et on attribue une valeur (qui peut varier) aux unités de ces systèmes. L'ensemble de ces unités forment le tout. Ainsi, le système numérique définit l'unité (représentée par le chiffre 1) et permet le découpage des unités en infiniment petit et par opération telle que la somme, en infiniment grand. Le système temporel traditionnel découpe le temps en secondes, en jours, en semaines, en mois etc. Le système informationnel découpe l'information en données. Le système informatique découpe l'information en donnée numérique. Le système linguistique découpe la langue en phrases, mots, syllabes, phonèmes. Le système typographique découpe la typographie en lettres, caractères. Le système lexical découpe le lexique en lexème, en morphèmes grammaticaux. Le système sémantique découpe le sens en concept et en instance. Le système documentaire découpe le document en contenus. Le système d'organisation des connaissances découpe la connaissance en facettes, en disciplines, en domaines, en thématiques, en sujets, en termes, en concepts, en instances et son organisation en liens, relations, propriétés et sous-liens, sous-relations, sous-propriétés.

Dans notre cas, étudier la granularité, c'est étudier les unités des SOC. La granularité la plus fine d'un système est atteinte lorsque l'on ne peut plus découper l'unité. Canivet (2011) relie la granularité aux besoins des utilisateurs : « jouer sur la granularité de l'information consiste à manipuler cette dernière en fonction des besoins de l'utilisateur ». La granularité de l'information ou granularité informationnelle dépend donc des usages envisagés. La granularité d'information peut par exemple être étudiée sur les sites web. En effet, de nos jours, la notion de granularité est liée à celle de référencement et en *Search Engine Optimisation* (SEO)[150], elle permet de découper le contenu d'un site et d'une page pour le hiérarchiser.

Les documents ont une granularité documentaire variable. L'indexation des documents doit correspondre aux requêtes des utilisateurs, c'est-à-dire que le niveau de granularité des requêtes doit correspondre au niveau de granularité des mots-clés, descripteurs, thématiques utilisées.

Le découpage de l'information peut être effectué à plusieurs niveaux :

- Niveau du document (ici on pourrait vouloir indexer le contenu de la transcription, transcription délimitée par le début et la fin des tours de parole)
- Niveau du tour de parole, (cela peut-être une indexation thématique comme nous l'avons fait suite à la lecture de chacun des tours de parole, elle serait différente d'une personne à l'autre) – le tour de parole peut permettre de faire émerger un thème et un rhème (la connaissance qu'on apporte en plus).
- Niveau d'une phrase, (de la même façon que le tour de parole, la lecture de la phrase permet de relever une thématique (thème) et parfois un rhème, parfois seulement dans le cadre d'un corpus oral car les phrases ne sont pas toujours finies)
- Niveau d'un syntagme, (linguistique) – (la lecture du syntagme peut permettre de relever un thème et un rhème) -niveau d'un lexème, (linguistique) (morphème lexical d'un lemme)
- Niveau d'un morphème (linguistique – unité minimale, niveau de granularité le plus fin)

La plateforme OKAPI permet par exemple d'indexer au niveau du contenu de l'image vidéo (il peut être envisagé d'indexer des expressions de visage, des objets présentés, un paysage en arrière-plan etc.), il est également possible d'indexer une image en 3D.

Ici, l'analyse est focalisée sur le plus petit niveau de sens porté par les lexèmes et syntagmes dans chacun des trois SOC.

[150] Traduction : Optimisation pour les moteurs de recherche

8.1 Les entités nommées

8.1.1 Les acteurs

Les acteurs, nous l'avons vu, sont des entités qui exercent une action importante dans le domaine du patrimoine minier. Ces acteurs sont soit des personnes, soit des groupes de personnes qui forment une unité dans leurs actions. Dans TerminoMines, les unités linguistiques (substantifs) se rapportant à des personnes désignent plusieurs réalités du domaine essentiellement concentrées sur l'activité minière passée. Elles ne sont en revanche pas précises comme des entités nommées car leur rôle est de qualifier la personne ou l'activité et non de l'identifier. Certaines ont une part de subjectivité et sont :

- liées au travail et représentent les concepts de :
    - qualité du travail effectué en général : *arland, hachepoteux, manoqueux, brile* comme 'mauvais ouvrier' ou *blatier* 'ouvrier négligent', *bradeu* 'ouvrier qui travaille trop vite au gré de ses coéquipiers', *bineux* 'tricheur, ouvrier qui cherche à passer devant les autres', *roufion* ou *lougnard* 'rapporteur, espion'
    - qualité du travail dans une tâche particulière : *saqueux* 'mineur abattant le charbon très rapidement'
    - temps de travail : *rat d'fosse* 'ouvrier restant au travail plus longtemps (avant 1950)' ou *rabioteux* 'ouvrier qui a dû effectuer un travail au-delà du temps du poste'
    - type de travail : se retrouvent ici tous les termes désignant les différentes tâches que les mineurs effectuent comme *ravacheux* 'galibots chargé de pousser les wagonnets dans les cages', *machinisse* 'mécanicien d'entretien chargé de la bonne marche de la cage' ou *mesureux* 'ouvrier chargé de mesurer le charbon destiné aux mineurs'. Cette catégorie est la plus importante (58 termes-concept pour 98 formes différentes).
    - chômage : *joqueux* 'ouvrier au chômage'
    - groupe ou un ensemble de mineurs : *ploton* 'groupe de mineurs se suivant dans un travail, notamment groupe de galibots se suivant dans une galerie en poussant des wagonnets' ou *coupe* 'ensemble des ouvriers descendus à la même heure'
- représentent le concept d'âge du mineur :
    - âgé : *raccomodeu* 'vieux mineur'
    - jeune : *galibot* 'jeune apprenti des mines de moins de 18 ans'
- représentent le concept de conflit social :
    - *jaune* 'ouvrier non gréviste' ou *blanques orelles* 'ouvrier qui n'est pas syndiqué'

- *délégué* ‘mineur élu syndicalement, responsable de la sécurité et des conditions de travail’

- représentent le concept de condition ou d’état du mineur :
  - *maca* ‘homme fort, diminutif de macaroni en italien, surnom dont les mineurs français gratifiaient les Italiens qui étaient selon eux les plus acharnés à la production’
  - *quoiché* ‘blessé’
  - *récapé* ou *rescapé* ‘mineur qui est réchappé d’un accident’
- représentent le concept des loisirs :
  - *coulonneux* ‘personne pratiquant la colombophilie’
- représentent le concept du lieu de travail :
  - *blanc joseph* ‘habitant du nord venu travailler dans les mines du pas-de-calais’, *boyau rouge* ‘appellation ironique employée par le mineur du nord pour désigner le mineur du pas-de-calais’

La seule référence à une entité nommée dans TerminoMines est *Fenzy* qui est un ‘ingénieur de la station de mesure de Liévin qui mit au point un appareil respiratoire à réglage automatique du type à oxygène et régénération […]’. Dans les corpus, cette entité ne désigne pas forcément la personne mais l’appareil respiratoire créé par celle-ci (figure 62). Dans ce cas, le terme *appareil* ou *appareil de sauvetage* placé avant ou après le nom propre permet de faire la différence avec l’entité nommée de personne.

Figure 62 : Appareil de sauvetage Fenzy – Photographies prises par l’auteure à la fosse 2 d’Oignies

ThesoMines ne contient pas de termes subjectifs sur les mineurs et indexe la catégorie la plus conséquente de TerminoMines qui concerne le type de travail et les tâches effectuées que nous avons regroupés sous le terme générique *métier*. Les concepts-feuilles de cette branche se situent au troisième ou au quatrième niveau de spécificité en fonction de s’ils sont regroupés par type de tâche ou non. Au troisième niveau se situent certains termes qui désignent plusieurs types de tâches comme *mécanicien*,

nous l'avons vu en 5.3.1, pouvant se placer sous *métier-transport* et *métier-entretien*. Le choix d'un thésaurus mono-hiérarchique a été motivé par la réalisation de l'ontologie sur sa base.

Dans ThesoMines, d'autres références peuvent désigner des acteurs et sont pour la plupart sous *histoire sociale* (surlignés en jaune sur la figure 63). Ces termes ne proviennent pas de la concaténation des 20 ressources lexicales initiales formant TerminoMines mais de son enrichissement. Sous cette branche, certains descripteurs renvoient à des EP qui désignent des personnes ou des groupes de personnes, c'est le cas d'*immigration belge*, *italienne* etc. qui ont pour EP *belge*, *italien* faisant référence à l'origine d'un acteur (mais pas seulement, cela peut être également d'une coutume, d'un objet etc.). Les acteurs associatifs peuvent être identifiés grâce aux descripteurs *vie associative* et son EP *association* et les acteurs muséaux par le descripteur *musée*. Les instances (entités nommées) telles qu'*association des amis de la mine* ou *musée de la mine de Harnes* ne sont pas contenues dans ThesoMines, elles peuvent en revanche être décrites par OntoMines.

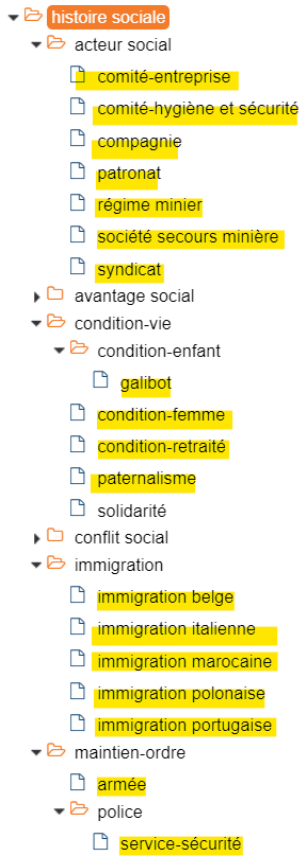


Figure 63 : Descripteurs (surlignés en jaune) faisant référence à des acteurs dans ThesoMines

OntoMines permet en effet d'enregistrer les acteurs grâce à la classe générique *E39 Actor*, en tant que personne *E21 Person* ou bien en tant que groupe avec *E74 Group*. Le tableau 29 contient les propriétés sélectionnées pour décrire les instances de ces classes.

Tableau 29 : choix des propriétés associées aux classes *E21 Person* et *E74 Group* pour décrire les acteurs du patrimoine minier
E21 Person
E21 Person : *P1 is identified by* : *E41 Appellation*
E21 Person : *P2 has type* : *E55 Type*
E21 Person : *P44 has condition* : *E3 Condition State*

Tableau 29 : choix des propriétés associées aux classes *E21 Person* et *E74 Group* pour décrire les acteurs du patrimoine minier
E21 Person : *P74 has current or former residence* : *E53 Place*
E21 Person : *P152 had parent* : *E21 Person*
E74 Group
E74 Group : *P1 is identified by* : *E41 Appellation*
E74 Group : *P2 has type* : *E55 Type*
E74 Group : *P74 has current or former residence* : *E53 Place*
E74 Group : *P107 has current or former member* : *E39 Actor*
E74 Group : *P107.1 kind of member* : *E55 Type*

La propriété *P1* permet de lier un acteur à sa dénomination, par exemple *Jean* sera une instance de *E41 Appellation*. La propriété *P2* permet de lier cette personne à un descripteur de thésaurus qui est enregistré comme instance de *E55 Type*. La propriété *P44* permet de décrire l'état d'une personne (par exemple son état de santé). La propriété *P74* permet de mentionner le lieu (*E53 Place*) où une personne/un groupe a sa résidence. La propriété *P152* est une relation de parenté entre une personne et une autre. La propriété *P107* permet de relier un acteur à un groupe en décrivant que cet acteur fait partie ou a fait partie de ce groupe. Enfin la propriété *P107.1* permet de spécifier le type d'adhésion ou le rôle qu'a le membre dans le groupe.

De plus, les acteurs du patrimoine minier peuvent effectuer des activités (*E7 Activity*), des modifications (*E11 Modification*), ils affectent des attributs (*E13 Attribute Assignement*), ils créent (*E65 Creation*), ils créent des types (*E83 Type Creation*) et sont participants d'activités (*E7 Activity*), de déplacements (*E9 Move*), de modifications (*E11 Modification*), d'affectation d'attribut (*E13 Attribute Assignement*), de création de type (*E83 Type Creation*), d'événements (*E5 Event*) et sont ou ont été détenteurs ou propriétaires d'objets fabriqués (*E22 Man-Made Object*).

OntoMines permet donc de décrire un énoncé fictif de ce type : « Gilbert, fils de Jérôme, autrefois abatteur, vit actuellement en Belgique. Ses longues années de travail au fond de la mine ne l'ont pas épargné, il souffre depuis des années de silicose. Aujourd'hui Jean est membre de l'association *Nom de l'association* ». La figure 64 représente cet énoncé, les propriétés sont nommées par leur identifiant et sont représentées par des flèches. Les références à ThesoMines sont surlignées en jaune. Le lien entre la classe générique *E39 Actor* et ses classes spécifiques *E74* et *E21* est représenté par des pointillés.

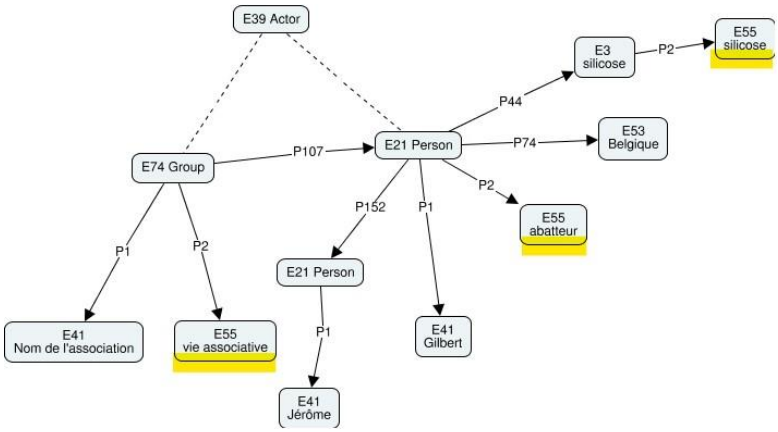


Figure 64 : Représentation d'un énoncé avec surlignage en jaune des références à ThesoMines

Si l'énoncé inventé correspond aux instances appelées par le modèle du CIDOC CRM, l'instanciation des témoignages n'est pas si évidente. En effet, le discours des anciens mineurs n'est pas aussi précis et si la référence à une entité nommée est faite dans le discours, ceci ne signifie pas qu'elle est forcément reliée sémantiquement à une autre, ou bien que cette autre entité nommée est pertinente pour le domaine. Nous donnons ci-dessous un exemple issu du travail d'instanciation du témoignage de Jean.

24 instances « candidates » ont pu être relevées pour les classes *E21* et *E74* (figure 65) afin d'enrichir OntoMines. Sur ces 24 instances, 16 sont des entités nommées, les 8 autres n'en sont pas car le contexte ne permet pas d'identifier précisément la personne ou le groupe de personne même si ces unités linguistiques font effectivement référence à des personnes (par exemple *musiciens*, *directeur*, *surveillants*, *maghrébins*). Une des questions qui se pose ici est : est-ce que le simple fait d'être mentionné par un ancien mineur fait de ces personnes ou groupes de personnes, des acteurs du patrimoine minier ? Est-ce que le *Paris Saint-Germain Football Club* est un acteur du patrimoine minier ? Le fait que les autres anciens mineurs ne mentionnent pas ce club et sa faible apparition dans les corpus (seulement une occurrence dans les deux corpus presse) pourrait nous permettre d'exclure celui-ci. En revanche, la référence à Lens en tant que ville, faisant partie du bassin minier (d'après les bases de données de la Mission Bassin minier) est à prendre en compte dans l'analyse.

Le choix doit également s'appuyer sur le contexte :

> Jean : « […] j'avais aussi une bonne possibilité en plus de la mine on avait possibilité de faire une carrière au football et moi j'avais des qualités de footballeur […], j'ai même joué au stade Bollaert contre Lens, au parc des Princes contre Paris et j'avais des propositions quand j'avais 16 ans d'aller à Toulouse ».

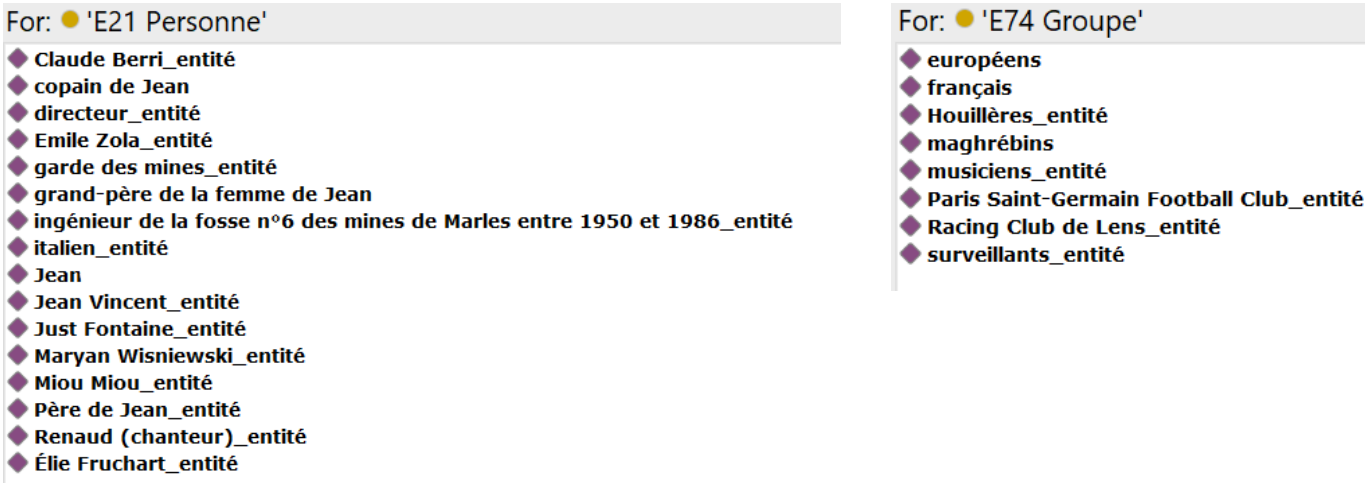


Figure 65 : Instances de *E21 Person* et *E74 Group* issues du témoignage de Jean

Grâce aux multiples références à ce concept dans les différents corpus (48 occurrences dans le premier corpus presse, 22 dans le deuxième, 11 dans le corpus audiovisuel, 19 dans l'inventaire des images d'archives du CHM), nous pouvons affirmer que le football est un loisir présent chez les mineurs, ce qui permet de justifier l'insertion d'entités nommées liées au football. Si, la mention « jouer contre » suivi

d'une entité nommée de type ville permet d'inférer que la ville représente en réalité une équipe de football, nous pouvons identifier une équipe du club de football de Lens et une équipe du club de football de Paris. Dénommer Lens par *Racing Club de Lens* et Paris par *Paris Saint-Germain Football Club* ne serait cependant pas possible même en croisant toutes les données issues du témoignage. En effet, concernant Lens, d'autres sources indiquent que de 1906 à 1969, le club de football de Lens avait le nom de *Racing club lensois*. Or Jean parle de cette activité qui a eu lieu pendant sa carrière (1950-1986, dates calculées grâce à la date de sa retraite qu'il mentionne dans son témoignage 1986 et le nombre d'années passées à la mine, au nombre de 36, donc 1986-36=1950). Seule la mention « j'avais des propositions quand j'avais 16 ans » dans la suite de son énoncé permet de situer à peu près cette activité dans les années 1952 sachant que Jean est né en 1936 (1936+16=1952) (donnée issue des métadonnées du document). Par hypothèse, le club de Lens dont parle Jean serait donc le *Racing club lensois*. Enfin, il ne s'agirait pas du club même mais de l'équipe du Racing club lensois (composée de joueurs spécifiques) ayant joué ce jour-là contre Jean. Cet exemple nous montre la difficulté à identifier de réelles entités nommées en ne prenant en compte que la forme. Le besoin de croiser un maximum de données reste la meilleure façon de rendre les instances précises. Ainsi le club de Lens a pu être identifié dans un certain nombre de documents sous les formes : *racing club de lens*, *RCL*, *Racing*, *club de Lens*, *Racing-club*, *club de football lensois* dans les corpus audiovisuels et presse, mais également dans des publications scientifiques. Dans le corpus presse, l'histoire de mineurs empruntant le chemin professionnel du football n'est pas unique comme le montre l'article de la *Voix du Nord* du samedi 4 mars 2017 issu de notre corpus presse sur Raymond Kopaszewski qui déclare au journal : « *Ma chance fut que mon entraîneur était également l'ingénieur de la fosse où je travaillais* », ou l'article de la *Voix du Nord* du jeudi 15 septembre 2005 qui mentionne la présence importante de polonais dans les équipes de football. Dans les sources scientifiques, Chovaux traite par exemple du concept de football en lien avec l'intégration des populations immigrées dans les mines du Nord (Chovaux 2001, 2006). L'identification de liens entre les mêmes concepts et entités nommées de même nature dans plusieurs ressources donne de la pertinence à l'instanciation au regard du domaine. La figure 66 montre les liens entre les instances issues du témoignage. Les concepts d'immigration et de football dans les mines peuvent s'étudier grâce au modèle OntoMines en repérant les liens entre classes et propriétés :

- E7 Activity (*football*) : P11 had participant : E21 Person (*joueur de football*)
- E21 Person (*joueur de football*) : P152 has parent : E21 Person (*parent du joueur de football*)
- E7 Activity (*football*) : P11 had participant : E74 Groupe (*équipe du club de football de ?*)
- E9 Move (*immigration*) : P25 moved : E21 Person (*parent du joueur de football*)
- E9 Move (*immigration*) : P27 moved from : E53 Place (*pays de provenance*)

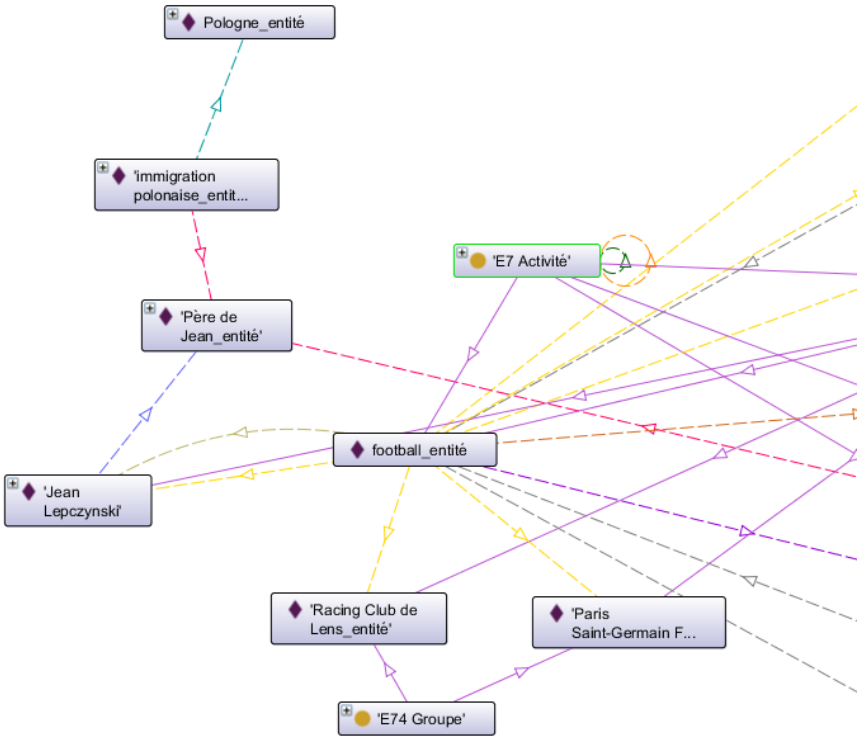


Figure 66 : Illustration de l'environnement des instances (losange violet) en lien avec l'instance nommée *Racing Club de Lens*. Schéma réalisé avec OntoGraph sur Protégé

Le terme « club » ne s'apparente pas qu'au football et représente des groupes pouvant être enregistrés comme instances de la classe *E74 Group*. Une requête a été effectuée dans le deuxième corpus presse et nous a permis de retirer une trentaine d'occurrences. L'identification a été réalisée à partir du logiciel *Agent Ransack* sur le premier lot de fichiers de 2004 du corpus de *La Voix du Nord* issu d'Europresse. Les entités nommées peuvent se décomposer selon le type de membre (selon l'âge, selon l'expertise, selon l'activité), l'objet du club (domaine ou activité), le nom du club et/ou sa localisation comme nous le montrons dans le tableau 30, ceci nous permettant d'enrichir OntoMines et notamment la classe *E53 Place* avec les noms de localisations et de sites qui entrent dans le périmètre défini. L'objet des clubs représente une partie des domaines et activités développés depuis l'après-mine et qui sont moins présents dans ThesoMines : la pétanque, le modélisme sont des activités qui sont de potentielles candidates.

Tableau 30 : Acteurs de type *club* dans une partie du corpus Europresse

Acteur	**Prép.**	**Membre**	**Prép.**	**Objet**	**Nom**	**Prép.**	**Localisation**
club		ados					
club	des	aînés				du	Val-de-Souchez
club	des	aînés				du	Numéro 2
club		amateurs	d'	histoire			locale
club	des	collectionneurs				de	Berck
club				colombophile			
club	de			football			
club				Généalogique		de l'	Artois
club				généalogique			Liévin
club				histoire			locale

Tableau 30 : Acteurs de type *club* dans une partie du corpus Europresse							
Acteur	**Prép.**	**Membre**	**Prép.**	**Objet**	**Nom**	**Prép.**	**Localisation**
club	d'			histoire			
club				historique		de	Roeulx
club	des			loisirs		de la	fosse 9
club				modélisme			
club	de			pétanque		de la	cité 5
club		philatéliste				de la	Chapelle-d'Armentières
club	de			prévention de la Vie Active			
club	de			prévention spécialisée	« avenir des cités »	de	Harnes
club				Racing		de	Lens
club					Bassin Minier UNESCO		
club						du	Pays de Condé
club					« Nord Madame »	à	Lewarde
club					LéoLagrange	à	Hergnies
club						de	Maulde
club					Patrimoine		
club					Noyelles UNESCO		
club					Unesco		
club					Roches		
club					AMICI		
club					Cigales		

8.1.2 Les entités spatiales

Dans TerminoMines, les entités spatiales se réfèrent aux lieux, espaces et éléments situés par rapport à d'autres grâce à des mentions comme « au fond », « au jour », « entre X et Y » etc. Par exemple, la *cage* est un 'ascenseur pour le transport, *entre la surface et le fond*, du personnel, du matériel et du charbon', et un *cayat* est un 'plan incliné *reliant deux étages*, et par où circulaient les wagonnets manœuvrés par un treuil - d'où le nom donné dans certaines mines : treul'. Ces entités peuvent être placées sur un plan ou un schéma représentant une fosse ainsi que ses composants mais ne sont pas précises comme des entités nommées, elles permettent cependant d'identifier des objets qui se placent ou se déplacent d'une manière générale d'une certaine façon dans un espace limité.

Ces entités sont organisées dans ThesoMines en fonction des critères *fond*, *surface* et *civil* (ce dernier représentant l'espace extérieur à une fosse) sous la branche *bâti*. ThesoMines ne contient en revanche pas d'entités nommées pouvant être précisément identifiées grâce à un référentiel géographique telles que des coordonnées *Global Positioning System* (GPS) (ville, pays etc.) mis à part le descripteur « zone tampon » qui correspond à un ensemble de points, espace défini par la Mission Bassin minier et les principaux noms des pays de l'immigration insérés comme EP tel que Pologne, Italie, Belgique, Maroc et Portugal.

De la même manière que les acteurs, OntoMines permet de décrire et d'enregistrer les entités spatiales d'une façon précise. Par ceci, nous signifions tous les lieux et noms de lieux qui peuvent être situés sur une carte grâce à des coordonnées GPS tels que les villes, villages, régions. Nous avons cependant choisi d'instancier OntoMines en nous limitant aux concepts pouvant être indexés par ThesoMines et aux lieux en lien avec ces indexations, limités aux villes incluses dans la zone tampon et aux départements (Nord et Pas-de-Calais) qui contiennent le bassin minier.

Le tableau 31 relève les propriétés qui définissent les lieux dans le modèle d'OntoMines.

Tableau 31 : Choix des propriétés associées à la classe *E53 Place* pour décrire les lieux du patrimoine minier
E53 Place : *P87 is identified by* : *E44 Place Appellation*
E53 Place : P87 is identified by : E45 Address
E53 Place : P87 is identified by : E47 Spatial Coordinates
E53 Place : P87 is identified by : E48 Place Name
E53 Place : *P89 falls within : E53 Place*

La propriété *P87* permet de nommer le lieu, de définir une adresse et des coordonnées spatiales. Grâce à *P89*, un lieu peut être situé dans un autre lieu. Ainsi, OntoMines pourra décrire la phrase suivante : « Arenberg, ville située dans le département du Nord a pour coordonnées spatiales XXX et YYY » :

E53 Place (Arenberg) : *P87 is identified by : E44 Place Appellation (*Arenberg*)*

E53 Place (Arenberg) : *P87 is identified by : E47 Spatial Coordinates* (XXX YYY)

E53 Place (Arenberg) : falls within : *E53 Place* (Nord)

Les lieux ne sont en revanche jamais décrits dans le discours par leurs coordonnées GPS. Un travail d'identification du lieu en question est donc nécessaire avant d'instancier. De plus, nous l'avons vu avec *Lens* pour l'équipe de football, les noms de villes n'identifient parfois pas la ville en elle-même mais une autre entité.

Une base de données reliant coordonnées et nom de lieu doit être établie. L'organisme Bassin minier Patrimoine mondial met à notre disposition une base de données[151] contenant le nom des biens inscrits à l'UNESCO ainsi que les communes sur lesquelles ils sont placés. Le site de l'UNESCO[152] fournit quant à lui des coordonnées pour les 109 éléments remarquables qui forment les biens inscrits mais ceux-ci correspondent à des points uniques et non à des surfaces.

Le croisement des deux sources de données peut néanmoins se réaliser grâce au nom des biens. Par exemple, le bien « Paysage et ensemble miniers de Chabaud-Latour et paysage et ensemble miniers de Sabatier » a pour coordonnées « N50 26 23.00 E3 35 11.00 » d'après le site de l'UNESCO. Ceci correspond à un point sur une carte, situé à Fresnes-sur-Escaut. En réalité, la surface du bien est de 715 ha d'après la base de données, surface dans laquelle se situe par exemple une partie d'un cavalier[153], lui-même traversant les villes d'Escautpont et Raismes, répertoriées par la base de données. Le

[151] Rubrique « donnée » à « SIG – Bien inscrit au Patrimoine mondial de l'UNESCO » https://bassinminier-patrimoinemondial.org/ressources-bm/#1614608475617-d274f31c-56e5
[152] https://whc.unesco.org/fr/list/1360/multiple=1&unique_number=1776
[153] Cavalier : 'Installations qui permettaient autrefois le transport du charbon par voie ferrée autrement dit, chemin de fer.'

document « Index : éléments par commune »[154] reprend quant à lui l'ensemble des 353 éléments par commune (figure 67) mais ne cite pas les 109 ensembles. Le croisement des données entre le document des éléments par commune et la base de données peut néanmoins se réaliser grâce à un numéro d'élément, ici le cavalier à le numéro 8 M, où 8 correspond au numéro de l'ensemble correspondant (parmi les 109).

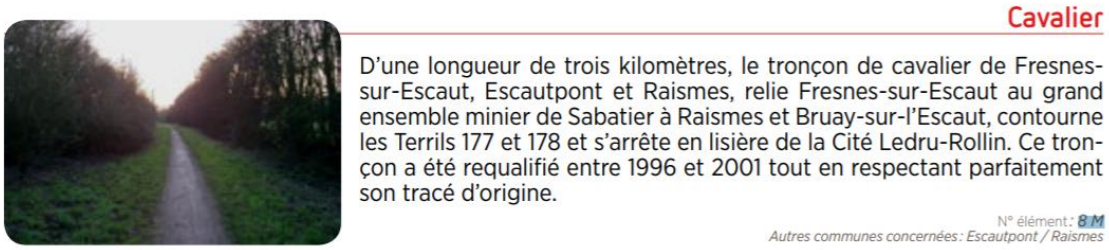


Figure 67 : Cavalier inscrit à l'UNESCO. Extrait de la brochure Index éléments par commune à https://bassinminier-patrimoinemondial.org/wp-content/uploads/2017/05/Porter-a-connaissance-Interactif-1.pdf

La page Wikipédia « Liste des biens du bassin minier du Nord-Pas-de-Calais inscrits sur la liste du patrimoine mondial »[155] répertorie les 109 ensembles en site numérotés avec leurs composants inscrits à l'intérieur. Certains composants sont renvoyés vers des pages les décrivant mais tous ne possèdent pas de coordonnées GPS. Enfin, nous avons vu que nous définissions le domaine grâce à la démarcation spatiale de la zone tampon. Or, ceci empêche d'instancier avec des entités spatiales pertinentes extérieures à celle-ci telles que la Pologne, l'Italie etc. L'instanciation se réalisant également sur la base des concepts de ThesoMines, ces entités sont tout de même prises en compte.

Pour illustrer ceci, nous instancions le témoignage de Jean. Celui-ci contient 15 instances candidates pour la classe *E53 Place* que nous relevons dans le tableau 32 avec leur environnement sémantique issu de la transcription. Les connaissances liées au domaine et leur granularité sont analysées par numéro d'extrait sous le tableau.

Tableau 32 : Instanciation des *E53 Place* dans la transcription du témoignage de Jean

Numéro d'extrait	*E53 Place*	Extrait de transcription
(1)	Pas-de-Calais, France, cirque de Gavarnie, Pyrénnées, brèche de Roland, Gavarnie, Alpes	parce qu'il y a beaucoup de **sites** qui sont **classés** maintenant à l'**UNESCO** aussi bien là-bas dans le **Pas-de-Calais**, dans toute la **France** euh là-bas hier on a vu une émission euh les racines et les ailes, le **cirque de Gavarnie** et moi j'ai été en haut des montagnes parce que à **l'école des cadres pendant les 4 ans** avec les **Houillères** on était obligé de faire des **stages** dans les montagnes une année on a été dans les **Pyrénées** donc la **brèche de Roland**, la **Gavarnie** et la 2ème année les **Alpes** alors on nous faisait grimper à presque 3800 mètres le point alors hier quand j'ai revu l'émission, que j'ai revu le refuge où on a dormi, le **refuge Felix-Fort**, les glaciers, il y en a presque plus parce que 40 ans en arrière là, il y avait des patinoires hein alors ça reste des souvenirs incroyables
(2)	Auchel, Pas-de-Calais, France, Boulogne	c'est l'année où **Juste Fontaine** il a marqué les 13 buts jamais égalés jusqu'à ce jour et de l'autre côté il y avait un autre **joueur Jean Vincent** c'était des **joueurs** qui ont joué ensemble à **Auchel** dans le **Pas-de-Calais** parce qu'il y en avait d'autre il y avait **Maryan Wisniewski**, il y en a eu il y en avait pas mal de **joueurs** qui avaient des qualités on avait un bon niveau on avait un bon niveau de **football** et surtout qu'on avait un **entraîneur** très sérieux qui a fait aussi une bonne carrière **Élie Fruchart Élie Fruchart** c'était un

[154] https://bassinminier-patrimoinemondial.org/ressources-bm/#1614608477182-ca60885a-2aca
[155] https://fr.wikipedia.org/wiki/Liste_des_biens_du_bassin_minier_du_Nord-Pas-de-Calais_inscrits_sur_la_liste_du_patrimoine_mondial

		entraineur qui a quand même été pris pour entraîner **l'équipe de France dans les années 58** mais regretter euh on peut plus retourner en arrière après j'avais une proposition d'aller à **Boulogne** quand **Boulogne** a instauré le professionnel mais ma femme elle était enceinte du premier alors c'est moi qui a eu peur parce que on avait quand même la **maison** on avait la **maternité des mines** on avait tous les **avantages des mines** donc c'est moi qui a abandonné le projet et puis voilà je suis resté jusqu'à la fin de ma carrière
(3)	fond, 6 de Calonne, Pologne	il avait 13 ans quand il est descendu au **fond** et moi j'ai commencé avec lui à la sortie de l'**école** parce qu'à la sortie de l'**école** on avait même pas besoin d'aller chercher du travail c'est le **garde des mines** qui venait nous chercher il était au courant de nos années et tout donc voilà, j'ai commencé là-bas au **6 de Calonne** avec mon père et on était formés sur le tas par nos vieux comme on dit mon père, les oncles et tout alors venus de **Pologne** en **1923** voilà alors moi j'ai commencé à la sortie d'**école** voilà
(4)	jour, Roubaix	ah bah c'est au **triage** que je parle des tr- oui parce que le **charbon** brut il **remontait** au **jour** et était **transporté** sur des **convoyeurs** et elles elles **triaient**, elles mettaient les gros blocs de caillou d'un côté et le **charbon** de l'autre elles **triaient**, voilà on les appelait les **cafus**, un surnom donné, terme minier hein, on disait **trieuse** mais ça après ça s'est amélioré parce qu'ils ont pu aller dans les usines, dans les filatures, il y en a beaucoup qui sont partis qui ont quitté la mine qui sont partis dans les filatures à **Roubaix**
(5)	Nord, Pas-de-Calais	il fallait encore faire un stage de 6 mois et comme les **Houillères** fermaient au **Pas-de-Calais** on nous a proposé deux choses, soit de **quitter la mine**, **aller à l'usine** ou aller **continuer dans les mines** il fallait **partir dans le Nord** alors moi j'étais déjà marié on avait des enfants et je dis on va risquer on va **dans le Nord** et d'un côté on a bien fait, parce que j'ai pu avoir ma **retraite** complète 3 ans avant la **fermeture** c'est-à-dire j'avais l'âge limite **50 ans c'était l'âge limite**

De ces six extraits, seules les instances *Pas-de-Calais*, *Auchel*, *6 de Calonne*, *Pologne* et *Nord* font partie du périmètre géographique du domaine. Les connaissances liées à ces instances sont relevées ci-dessous.

Extrait (1)

Le premier extrait concerne le lien entre le classement de sites à l'UNESCO dont le souvenir est stimulé par un programme de télévision regardé la veille par Jean qui présentait le *cirque de Gavarnie* (amphithéâtre inscrit au patrimoine mondial). Ce bien inscrit au patrimoine mondial anime chez Jean son souvenir de la *brèche de Roland*, située à proximité du *cirque de Gavarnie*, lieu où il est allé lors d'un stage lorsqu'il était à l'école des cadres. Dans cet extrait, la connaissance sur le domaine est multiple : le fait qu'un stage soit réalisé lors de la formation des mineurs, que l'école des cadres soit tenue par les Houillères et ait une durée de quatre ans et que le stage soit effectué dans les montagnes deux années de suite avec comme activité une ascension forte.

Pour l'extrait (1), ces choix permettent de n'enregistrer que *Pas-de-Calais*. Cette instance est liée en contexte au terme *sites* qui peut être indexé par *site et paysage* de ThesoMines, au terme *classés* pouvant être indexé par *classement* et au terme *UNESCO* indexé par *patrimoine mondial*. Le sens de cette phrase est : il existe des sites dans le département du Pas-de-Calais qui sont classés patrimoine mondial. Cette connaissance peut être enregistrée dans OntoMines comme le montre la figure 68. Au vu du graphe réalisé pour une seule phrase, la granularité de l'instanciation se pose. Quant aux connaissances spatiales liées au stage, nous verrons en 8.3 comment celles-ci peuvent être enregistrées en tant que *E7 Activity*.

Extrait (2)

Les instances candidates sont *Auchel* et *Pas-de-Calais*. *France* et *Boulogne* ne font pas partie du périmètre du domaine. L'ambiguïté de l'entité *Auchel* ne permet pas d'instancier OntoMines. En effet, est-ce jouer dans l'équipe de football d'Auchel ? Est-ce jouer dans la ville d'Auchel ? Si c'est la ville, l'instanciation permet d'enrichir le graphe de la figure 68.

Extrait (3)

Les instances candidates sont *6 de Calonne* et *Pologne*. *Fond* n'est pas assez précis car il n'y a pas de référence claire à la fosse dans laquelle est descendue le père de Jean à 13 ans, même si nous pouvons faire l'hypothèse grâce au contexte que Jean et son père ont travaillé à la même fosse, nommée *6 de Calonne* par Jean. *6 de Calonne* n'est pas inclus dans la base de donnée des biens inscrits à l'UNESCO (25% des biens de la totalité du patrimoine minier a été retenu). En revanche, *Calonne* fait référence à une ville, Calonne-Ricouart, ville minière répertoriée dans la liste ACOM (Association des Communes Minières de France) à laquelle la source (Corbion 2016, Département 62-2) fait référence « On y trouvait la Fosse n°6 (3 Puits), dépendant de la Cie des Mines de Marles (Marles-les-Mines 62540) fondée en 1862 et nationalisée en 1946 ». Le numéro 6 fait donc référence à la fosse n°6. La fosse est également appelée « Fosse n° 6 des mines de Marles » sur Wikipédia[156]. Enfin, *Pologne* désigne le pays de provenance du père de Jean, qui est venu travailler dans les mines en France et fait donc référence à l'immigration polonaise.

Extrait (4)

Il n'y a pas d'instance candidates pour cet extrait. *Jour* est employé d'une manière générale, en effet le charbon est remonté au jour et cela est vrai pour toutes les fosses. S'il n'est pas instancié dans OntoMines, il sera tout de même indexé par ThesoMines grâce au descripteur *carreau-surface*. *Roubaix* ne fait pas partie du périmètre du domaine, en revanche, cette instance mérite attention car elle fait référence aux filatures, terme spécifique au domaine du textile et fait ainsi référence au projet TECTONIQ. Le fait de ne pas instancier la ville de Roubaix pourrait empêcher de faire le pont entre ces deux domaines.

Extrait (5)

Les instances candidates pour cet extrait sont *Pas-de-Calais* et *Nord*. La connaissance transmise par cet extrait à propos de ces deux départements concerne la fermeture des mines dans le département du Pas-de-Calais et le fait que cette fermeture ait obligé certains mineurs à aller travailler dans les mines du département du Nord (figure 68). Cet exemple montre la construction du discours et son instanciation par les entités spatiales de type E53 Place.

[156] https://fr.wikipedia.org/wiki/Fosse_n%C2%B0_6_des_mines_de_Marles

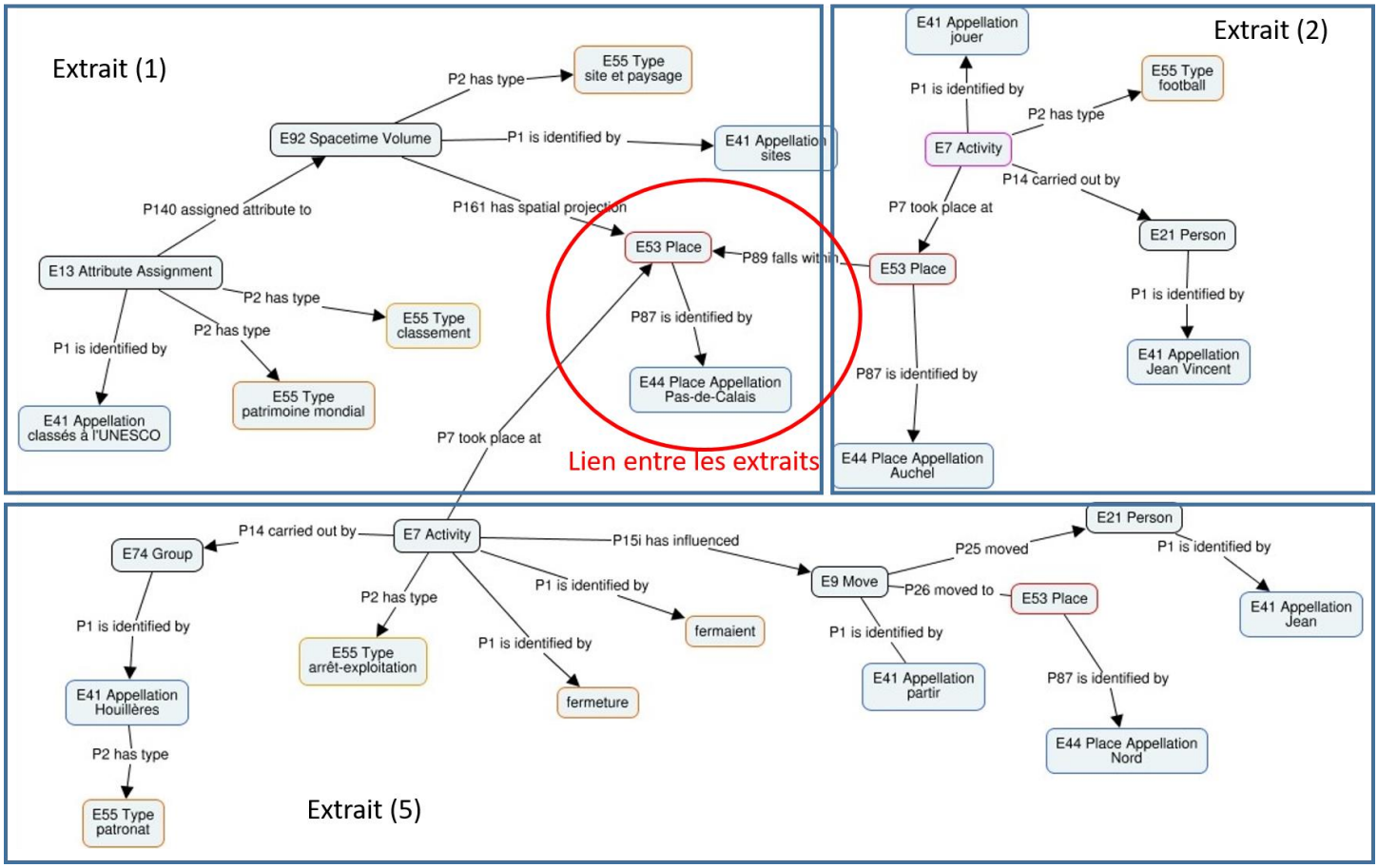


Figure 68 : Connaissances « sites classés à l'UNESCO dans le Pas-de-Calais » et « Jean Vincent a joué au football à Auchel, ville du Pas-de-Calais » et « Influence de la fermeture des mines dans le Pas-de-Calais sur le départ de Jean vers le Nord » décrites par OntoMines. En orange : liens avec ThesoMines, en bleu : termes utilisés par Jean, en rouge : E53 Place. Schéma disponible à https://cmapscloud.ihmc.us:443/rid=1XC9DD50R-1TK40PZ-1VXFC8

8.1.3 Les entités temporelles

Dans TerminoMines, les entités temporelles sont représentées par des termes qui désignent soit un découpage du temps de travail du mineur (ou ce temps), soit des périodes géologiques. Par exemple, dans la première catégorie, *coupe à carbon* est la 'période de temps pendant laquelle on travaille au charbon', *coupe du matin* correspond aux 'travaux d'abattage de la première partie de la journée', *coupe à terres* est la 'période de temps (l'après-midi) pendant laquelle on travaille aux remblais et aux services divers' et la *quinzaine* 'distingue aussi bien un demi mois de travail que la paye versée au terme de cette période'. La deuxième catégorie contient des termes comme *discordance* qui est la 'période géologique pendant laquelle la suite des formations géologiques est interrompue par un arrêt de la sédimentation et/ou l'érosion' ou encore *ères primaires, secondaires, tertiaires* et *quaternaires* qui sont :

> 'des grandes époques géologiques, la plus ancienne étant l'ère primaire et la plus récente l'ère quaternaire :
>
> -quaternaire : à partir de -1,8 millions d'années
>
> -tertiaire : -65 à -1,8 millions d'années
>
> -secondaire : -245 à -65 millions d'années

-primaire : -530 à -245 millions d'années'.

Par rapport aux premières, ces dernières entités sont des entités nommées précises comme elles contiennent une référence à des années.

Dans ThesoMines, ces premières entités sont insérées sous le descripteur *temps-travail* avec des termes comme *descente*, *remonte*, *pointage*, *briquet* qui désignent des activités qui peuvent se situer chronologiquement entre elles par rapport à une référence (ici le temps de travail, descripteur *poste* qui se décompose en *poste-matin*, *poste-après-midi*, *poste-nuit*) et peuvent être précises qu'en contexte. Les descripteurs correspondant à des fêtes religieuses telles que *Noël*, *Pâques*, *Sainte-Barbe* et *Sainte Catherine* sont précises mais en raison de leur caractère répétitif annuel, elles restent peu précises tant qu'une année n'est pas mentionnée. Les descripteurs correspondants aux conflits militaires *bataille-Denain*, *première guerre mondiale*, *deuxième guerre mondiale*, *guerre franco-allemande* sont des entités nommées précises délimitées dans le temps. Les conflits sociaux comme *grève* et différentes vagues d'immigration avec *immigration belge*, *immigration italienne* etc., et les périodes de l'histoire de l'exploitation comme *après-mine*, *arrêt-exploitation*, *bataille-charbon*, *découverte du charbon*, *nationalisation* peuvent également être situés dans le temps par des périodes.

OntoMines permet de décrire et d'enregistrer les entités temporelles. Dans ce cas, les entités temporelles sont forcément reliées à un événement ou à une activité précise. Les classes *E52 Time-Span* et *E50 Date* permettent d'exprimer des périodes ou des dates.

Tableau 33 : Choix des propriétés associées à la classe *E52 Time-Span* pour décrire les périodes du patrimoine minier
E52 Time-Span : *P78 is identified by* : E50 Date
E52 Time-Span : *P2 has type* : *E55 Type*
E52 Time-Span : *P81 ongoing throughout* : *E61 Time Primitive*

La propriété *P78* permet de relier une date à une période. La *P2* permet de relier cette période à un descripteur de thésaurus et la *P81* relie un laps de temps à des primitives temporelles. Ainsi, pour décrire qu'une activité dure six mois entre 1987 et 1990, on notera :

E52 Time-Span (6 mois) : *P81 ongoing throughout* : *E61 Time Primitive* (1987-1990).

Pour relier une activité à un laps de temps, à une période ou pour la relier à une date, il est nécessaire d'utiliser la propriété *P160*. Pour décrire un stage (EP de formation professionnelle) ayant une durée de six mois (1) ou réalisé en 1967 (2), il faut instancier ainsi :

(1) *E7 Activity* (formation professionnelle) : *P160 has temporal projection* : *E52 Time-Span* (6 mois)

(2) *E7 Activity* (formation professionnelle) : *P160 has temporal projection* : *E52 Time-Span* : *P78 is identified by* : *E50 Date* (1967).

Le témoignage de Jean nous permet de relever cinq instances de la classe *E50 Date* reliées à des événements tels que la venue du père de Jean en France (1923), à une coupe du monde de football (1958), à la retraite de Jean (1986), à la fermeture de la fosse où Jean travaillait en dernier (mars 1989) et à l'écriture du roman Germinal par Émile Zola (1800 et quelques). Concernant cette dernière instance,

il est possible d'instancier par siècle et de situer l'écriture du roman au 19ème siècle avec un *E61 Time Primitive*. Il serait également possible de vérifier l'information et d'instancier cette création d'œuvre *E65 Creation* avec une période d'écriture plus précise encore (écrit d'avril 1884 à janvier 1885). Le choix de rester sur les connaissances transmises par l'entretien implique deux choses : de potentiels erreurs sur les datations ou la perte d'information si le choix est fait de ne pas instancier.

Trois instances ont également pu être relevées concernant *E52 Time-Span* reliées à la durée de l'école des cadres (4 ans), l'âge de Jean à sa retraite qui est également l'âge maximal pour être mineur (50 ans) et la durée de la carrière de Jean dans les mines (36 ans).

Deux types de connaissances se distinguent de cette instanciation : d'une part, des connaissances liées à la personne interrogée, à sa famille et à ses activités que nous ne pouvions pas forcément connaître et d'autre part, des connaissances historiques faisant partie de la culture générale, concernant des personnes ou des actions connues et ayant eu un impact sur un grand nombre de personne.

8.2 Les objets matériels

Dans TerminoMines, les objets matériels sont définis comme des appareils, des accessoires, des équipements, des outils, des machines, des pièces qui forment des ensembles ou des parties d'ensembles composés de matière et qui sont assemblés d'une certaine façon afin qu'un acteur les utilisent pour accomplir une tâche précise dans le cadre du travail à la mine. Ils ont comme caractéristiques de pouvoir être portés ou de se mettre en mouvement (ce qui distingue par exemple le bâti, qui est également fait de matière par exemple, le bâti d'une machine à extraction). Parmi ceux-ci, la *molette* est une 'poulie dominant le puits et destinée à l'enroulement des câbles de la cage', la *patte* est le 'bout extrême d'un câble', le *pourcheau d'fosse* est un 'contrepoids en fonte utilisé dans un plan incliné à voie simple ou dans un puits à une seule cage', le *cuffat* est l''ancêtre de la cage, tonneau qui descendait les mineurs au fond du puits et qui remontait le charbon'.

Dans ThesoMines, la branche *matériel* regroupe tous ces termes selon la même logique que les termes sous *métier*, c'est-à-dire par tâche ou fonction. Les termes trop spécifiques désignant des petites parties d'un tout comme *patte* ne sont pas insérés sauf s'ils sont souvent mentionnés dans les corpus, ils sont alors insérés en EP d'un terme plus générique comme le terme *araïot* et ses variantes qui sont une 'tige de fer solide, à poignée en boucle, que l'ouvrier mineur engageait dans la roue d'un wagonnet pour le contrôler.' Certains termes comme *cuffat*, nous l'avons vu, désignent un objet qui était utilisé à une certaine période et a été remplacé par un autre, ici la *cage*. Ce n'est pas le même objet mais ils ont la même fonction, ici de transport des hommes ou du matériel. Le terme *cage* étant le plus récent, celui-ci est préféré et *cuffat* est inséré en EP. ThesoMines ne contient néanmoins pas de mention à cette évolution historique.

OntoMines décrit et instancie ces objets sous la classe *E22 Man-Made Object*. Contrairement aux acteurs, entités temporelles et spatiales, la classe *E22* est étudiée en détail en fonction des propriétés sémantiques de chacun des concepts. Ainsi, 12 classes spécifiques à *E22* ont pu être relevées grâce à

l'étude des concepts de ThesoMines. OntoMines permet en effet de faire la différence entre le concept de *chapelle*, celui de *gamelle* ou encore celui de *repas* (concept désignant la nourriture et non le moment ni le temps du repas) qui ont comme point commun le fait de pouvoir être identifiés par une appellation (*E41*), d'être reliées à un descripteur de thésaurus (*E55*), de pouvoir s'insérer dans un espace (*E92*), d'être composées de chose matérielle (*E18*), d'avoir ou d'avoir eu pour localisation un lieu (*E53*) et d'avoir comme projection temporelle une durée (*E52*). En revanche, le concept de *chapelle*, contrairement à celui de *gamelle* et de *repas* peut être le support d'un objet symbolique (*E90*). Ceci permet notamment de relier le concept de *chapelle* au type de culte.

Le témoignage de Jean permet d'enregistrer et de décrire 22 instances pour la classe *E22*. Celles-ci sont représentées sur la figure 70. Ces instances portent toutes la forme du concept de ThesoMines associé. Elles sont en réalité toutes reliées à une classe *E41 Appellation* qui rapporte la forme du terme tel qu'il est inscrit dans le témoignage. La particularité de ces instances est qu'elles sont toutes peu précises c'est-à-dire qu'elles ne sont pas inscrites dans une spatialité ou dans une temporalité énoncée par Jean. Cependant des déictiques tels qu'« ici » peuvent permettre d'identifier des lieux, à condition de savoir où le témoignage a été tourné. Nous savons que celui-ci a été tourné dans l'ancien vestiaire à Arenberg ce qui permet de comprendre une phrase telle que : « c'est le souvenir rien qu'ici en regardant en haut vous voyez pas à chaque molette qu'il y avait là des vestons, des godasses, des casques, les bancs numérotés moi quand je viens ici je regarde et je crois qu'il y a toujours les bancs numérotés, qu'il y a tout le temps les chaînes qui sont pendues c'est encore gravé c'est encore dans la mémoire ». Ceci permet par exemple de situer des objets (veston, chaussure, casque) dans un lieu, objets qui ne sont actuellement plus là comme nous pouvons le voir sur la figure 69.

Figure 69 : Vestiaire de la fosse d'Arenberg – Juin 2018

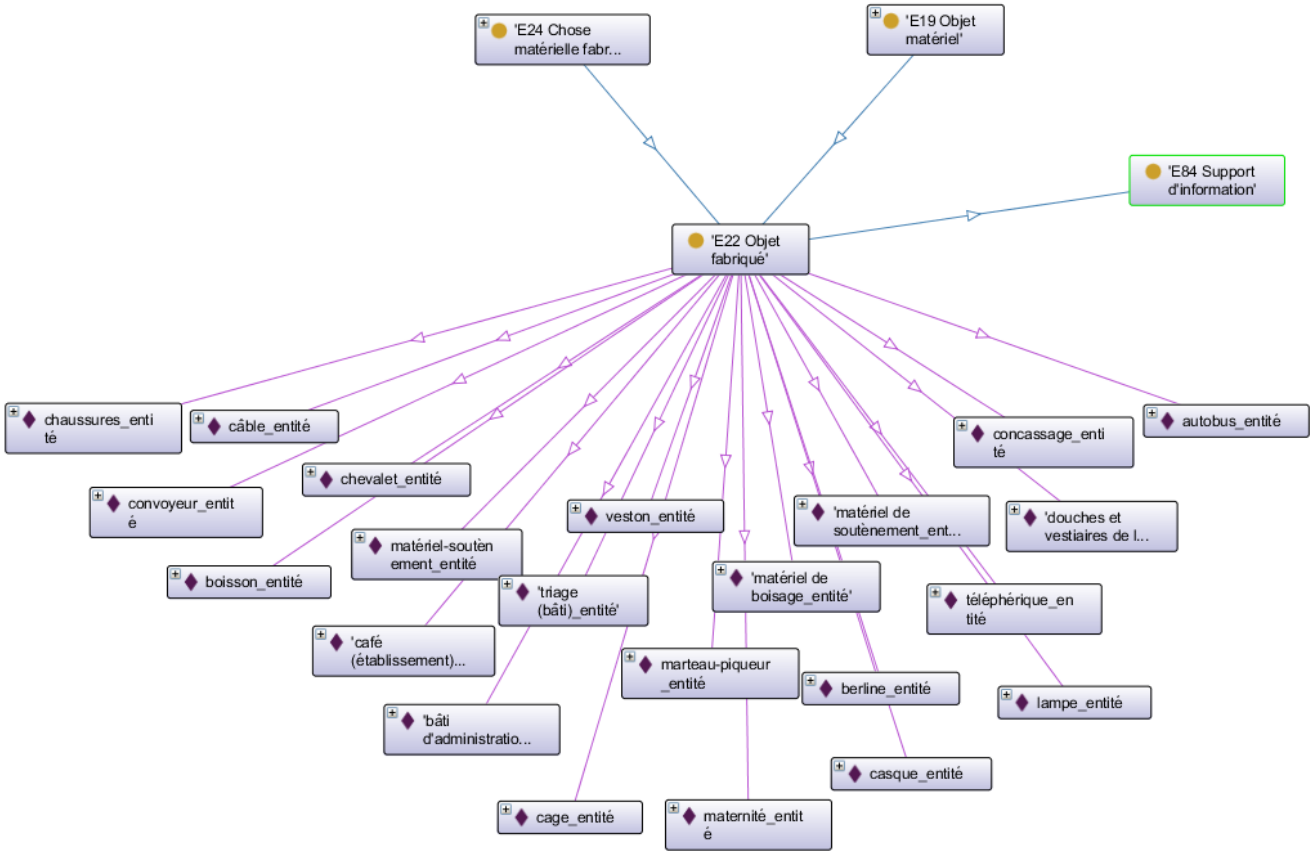


Figure 70 : Instances de *E22 Man-Made Object* issues du témoignage de Jean. Les flèches violettes représentent la relation d'instanciation, les bleues de hiérarchie. Schéma réalisé avec OntoGraph sur Protégé

8.3 Les activités

Dans TerminoMines, les activités sont représentées par des verbes d'action ou des verbes substantivés représentant des opérations ou des tâches particulières réalisées par les acteurs du domaine de l'activité minière. Cela peut être effectuer un ouvrage comme *plancher* qui est 'poser la petite voie ferrée destinée aux wagonnets', caler quelque chose comme *abloquache* qui est l''action d'abloquer, de caler un montant provisoire', appliquer une technique comme *alfranquir* qui est 'couper droit le front de taille', empêcher ou prévenir quelque chose comme *schistification* qui est l''opération de neutralisation des poussières charbonneuses par projection de produit crayeux' ou encore manœuvrer comme *tourteler* qui est 'actionner à la main la manivelle d'un treuil ou d'un ventilateur'.

Dans ThesoMines, les activités sont regroupées sous la branche *opération*. Ce sont ces termes qui sont repris sous les branches *matériel*, *métier* et *bâti* car en fonction des tâches à effectuer, des appellations de personne, de matériel et de bâti spécifiques sont utilisées. Par exemple, le descripteur *opération-triage* sous *opération* est associé aux descripteurs *matériel-triage* sous *matériel*, *métier-triage* sous *métier* et *bâti-triage* sous *bâti*. Les relations d'association permettent de relier, grâce aux termes génériques, les termes spécifiques *tamis*, *tapis-roulant*, *criblage* ou *criblache*, *triage*, *cafu*, *cafut*, *cribleuse*, *cul à gaillette*, *trieur*, *trieuse* ensemble tout en gardant une granularité fine concernant l'information issue de la terminologie.

Dans OntoMines, celles-ci sont instanciées par la classe *E7 Activity*. De la même manière que la classe *E22*, la classe *E7* est étudiée en détail et nous relevons neuf manières différentes de décrire ses instances (Cf. annexe 4 sous *E7_Activity*). Nous distinguons par exemple les concepts *abattage* (schéma 11), du concept de *sanction* (schéma 12), de *briquet* (schéma 13), d'*arrêt-exploitation* ou *conflit militaire* (schéma 14), de *chômage* (schéma 15), d'*activité domestique* ou *bal* (schéma 16), d'*avantage social* (schéma 17), de *condition-femme* (schéma 18) et de *risque sanitaire* (schéma 33). Tandis que tous ces concepts peuvent être décrits par une appellation (*E41*), être reliés à un descripteur de ThesoMines (*E55*), avoir lieu sur ou à l'intérieur d'une chose matérielle (*E18*), être influencés par une instance de classe d'OntoMines (*E1*) et être réalisés par un acteur (*E39*), ils diffèrent selon les propriétés listés dans le tableau 34.

Tableau 34 : Propriétés du modèle OntoMines qui diffèrent selon le type de concept pour la classe *E7 Activity*

Propriété : Rang	**Schémas**								
	(11)	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**	**(17)**	**(18)**	**(33)**
P7 took place at : E53 Place		ok	ok	ok	ok	ok	ok	ok	ok
P10 falls within E92 : Spacetime Volume								ok	
P11 had participant : E39 Actor		ok		ok	ok	ok	ok	ok	
P16 used specific object : E70 Thing	ok	ok	ok		ok	ok			ok
P19 was intended use of : E71 Man-Made Thing									ok
P21 had general purpose : E55 Type				ok	ok	ok	ok	ok	
P32 used general technique : E55 Type	ok	ok							
P117 occurs during : E2 Temporal Entity	ok			ok	ok	ok	ok	ok	
P134 continued : E7 Activity						ok	ok	ok	
P161 has spatial projection : E53 Place						ok			

L'instanciation du témoignage de Jean avec OntoMines nous a permis de relever 40 instances (figure 71) pour la classe *E7*.

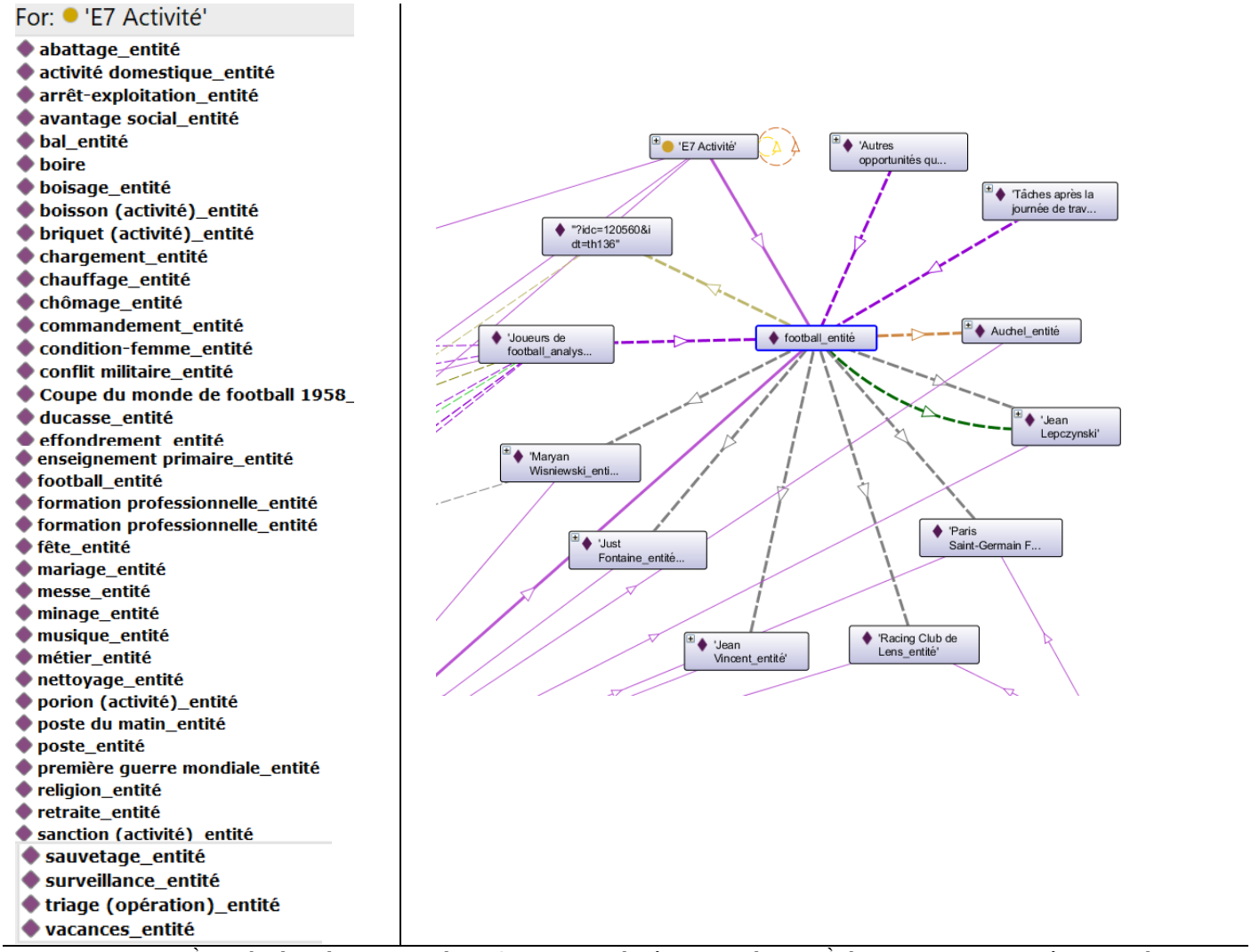


Figure 71 : À gauche, liste des instances de E7 Activity issue du témoignage de Jean. À droite, environnement sémantique de l'instance *football*. Captures d'écran issues de Protégé avec OntoGraph.

Chacune des instances est reliée au moins à un descripteur de ThesoMines. En moyenne les instances sont reliées à 3.1 instances (huit au maximum). Les liens aux huit instances sont en réalité effectués grâce à quatre propriétés différentes (avec six utilisations de la propriété *P11 had participant*), en gris sur le schéma de la figure 71. Nous l'avons vu plus haut, la finesse et la richesse de l'instanciation dépend de la délimitation du domaine et de la désambiguïsation des instances. Sur le schéma, les flèches violettes en pointillées représentent le lien entre la segmentation en tour de parole du témoignage et le contenu, ici l'instance *football*. Trois segments font donc référence au concept *football* que nous avons intitulés thématiquement : « joueurs de football », « autres opportunités que la mine », et « tâches après la journée de travail ». Il est difficile de valider le modèle d'OntoMines à partir de l'instanciation d'un seul entretien mais celle-ci a permis de proposer des enrichissements : la propriété *P15 was influenced by* et son inverse *P15i influenced* a par exemple été ajoutée à tous les schémas suite à la phase d'instanciation pour signifier l'influence d'une activité sur une autre et a ainsi permis de représenter l'influence du concept de fermeture des exploitations sur celui de mutation, de l'effet du mariage sur le fait d'avoir une maison des mines ou le fait d'avoir une retraite complète grâce à celui de mutation.

Pour conclure, la finesse de description varie en fonction des caractéristiques des SOC : les instances ou concepts feuilles de ThesoMines n'ont pas la précision des instances d'OntoMines. En effet, le premier ne traite pas ou peu les entités nommées, tandis que le deuxième est tout particulièrement dédié à leur description. Néanmoins, le thésaurus et l'ontologie possèdent tous deux la possibilité de décrire ou d'indexer différents niveaux d'analyse du domaine (de 2 à 5 niveaux pour le thésaurus en comptant le premier niveau et jusqu'à 6 niveaux pour les classes d'OntoMines sans compter la racine et jusqu'à 2 niveaux de spécificité pour les propriétés), ce que ne permet pas la terminologie. Cette dernière met sur le même plan de spécificité tous les termes du domaine alors que le thésaurus décide de niveaux hiérarchiques entre ceux-ci, les termes trop spécifiques n'étant soit pas inclus dans la structure ou insérés comme des employés pour de concepts plus génériques. Enfin, l'ontologie instanciée (OntoMines) fait le pont entre la spécificité du domaine et la généralité des classes pour permettre l'interopérabilité avec d'autres domaines. Les différents niveaux d'analyse permettent de multiples points de vue sur le domaine et par conséquent, augmentent les chances de correspondre aux attentes des utilisateurs.

CHAPITRE 9. Expressivité des formalismes

Ce chapitre étudie la complémentarité des SOC dans leur capacité à représenter exhaustivement le domaine grâce à leur formalisme. Les formalismes sont des moyens de concevoir et de représenter les connaissances. Ici, les formalismes des SOC sont des langages de représentation des connaissances. Ceux-ci se caractérisent par leur syntaxe (signes ou règles qui structurent les assertions) et par leur sémantique, donc la manière dont les assertions sont interprétées (GDT[157]). L'expressivité est la capacité des assertions à être interprétées par leur seule forme. Nous étudions ici le pouvoir d'expressivité sémantique qu'entretiennent les unités de chaque SOC entre elles. Il s'agit plus précisément de comprendre la différence d'expressivité des formalismes des trois SOC pour montrer leur complémentarité à l'aide de l'étude fine du témoignage de Jean. L'étude en question consiste à organiser les connaissances issues de la transcription grâce aux trois SOC. Les limites de leur expressivité sont également discutées.

9.1 La structure de la terminologie

En premier lieu, nous souhaitons représenter la structure des données terminologiques identifiées dans le témoignage. Nous indexons la transcription grâce aux termes de TerminoMines. Pour cela, nous choisissons de marquer automatiquement la transcription à partir du module d'indexation de Word et d'utiliser comme index un fichier Word réalisé avec les entrées des 20 ressources lexicales.

Les résultats montrent que sur 7 493 mots contenus dans la transcription (calculés par Word, un mot étant une suite de caractères distingués par un séparateur de type espace) 47 occurrences (correspondant à 50 mots selon Word) de TerminoMines apparaissent dans le document. Une analyse à partir des définitions permet de mettre de côté 14 occurrences dont le sens n'est pas spécifique au domaine minier. Parmi ceux-ci, le terme *coupe*, qui est un synonyme de travail dans le domaine minier est marqué par l'index dans l'expression « *coupe du monde 58 »* ou encore le terme *avot* défini dans TerminoMines comme une '*longue corde ou câble utilisé pour attacher du matériel devant descendre ou monter un plan incliné ou un puits*' est employé dans le corpus comme le verbe avoir à la troisième personne en dialecte « *on avot le double poste* ». En effectuant la transcription, nous avons en effet souhaité rester au plus proche du langage du mineur et transcrire le dialecte.

33 occurrences correspondent donc au domaine de la mine. De ces 33 occurrences, 8 sont reliées à des variantes ou synonymes (par exemple, *cage* a pour variante *cache*, *chevalet* a pour variante *chevalement*, *cafus* a pour synonyme *mahus*, *terril* a pour variante *terri*). Le lien entre les occurrences et leurs variantes est noté seulement dans quelques ressources lexicales par les termes « renvoi » ou « voir », « synonyme » ou encore « se dit aussi » que seul l'humain peut interpréter comme une relation. Il n'y a

[157] GDT Langage de représentation des connaissances

pas d'information sur le fait que ces termes expriment une relation sémantique pour la machine. L'expressivité du formalisme est donc faible.

Afin de visualiser ces résultats, nous avons réalisé la figure 72 qui correspond à une représentation des données terminologiques et documentaires (en orange) portées par l'étude de la transcription du témoignage de Jean au regard de TerminoMines. Cette représentation relie les occurrences (en orange et en majuscule sur la figure 72) à leur définition issue de sources (ressources lexicales), leur variante et renvoi. Par souci de lisibilité, nous n'avons représenté ici que six occurrences sur les 33 identifiées dans le témoignage et n'avons pas reporté toutes les définitions présentes dans les ressources lexicales. Les liens permettent de relier des données linguistiques entre elles (notés *définition*, *variante*, *renvoi*) ou des données documentaires (notés *contient*, celui-ci représente le lien d'indexation entre un document et son contenu (ici le contenu est représenté par des termes et leur contexte), *source* indique la ressource lexicale fournissant la *définition*). Les liens entre chaque donnée sont réalisés par interprétation humaine c'est-à-dire qu'aucune couche sémantique sur la nature des liens n'est disponible dans TerminoMines.

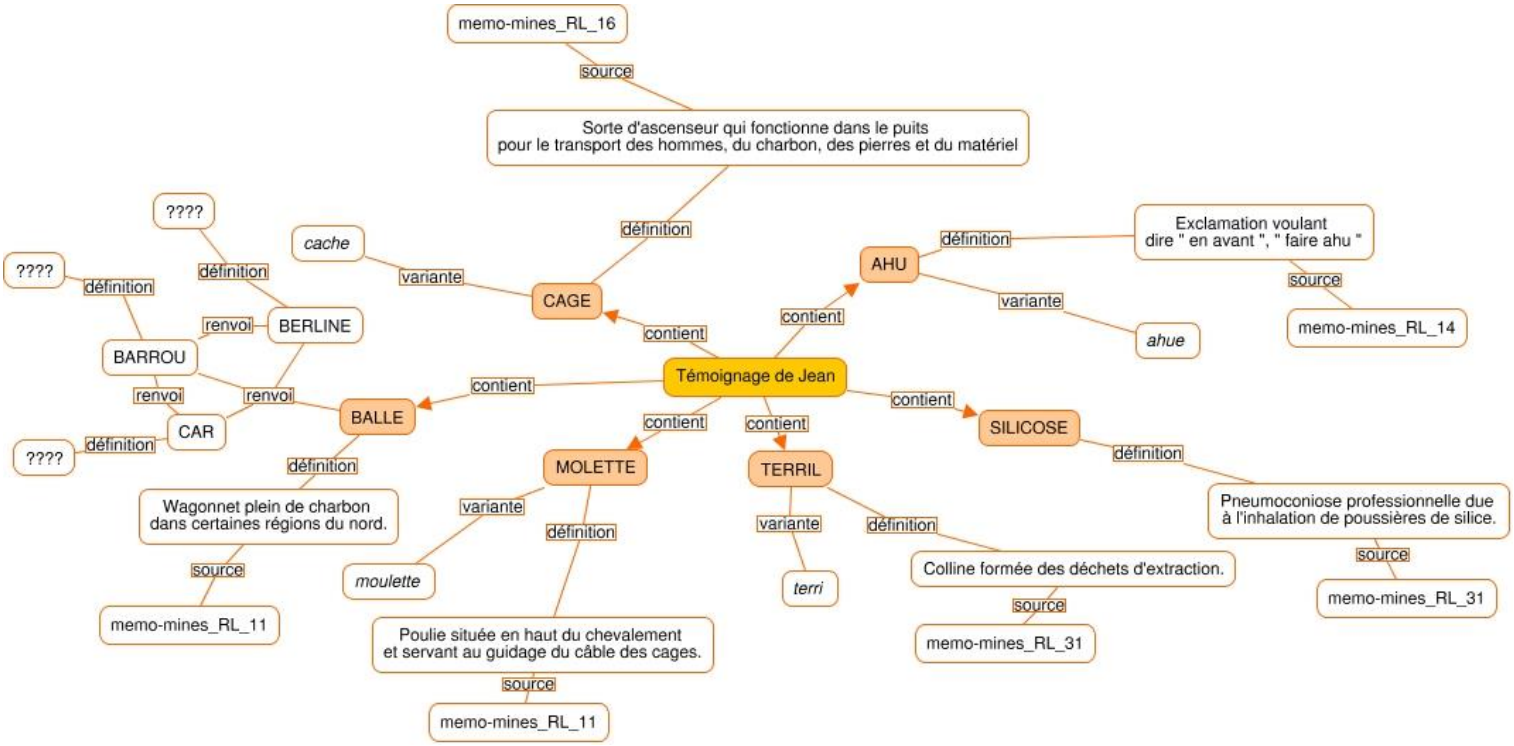


Figure 72 : Termes contenus dans le témoignage de Jean reliés à leurs données terminologiques

Au niveau documentaire, TerminoMines pourrait être utilisée pour indexer les ressources lexicales, et les documents, de la même manière qu'une indexation plein texte, mais la structure « à plat » des données de TerminoMines (il n'y a pas de lien hiérarchique explicités entre les termes) produirait un phénomène de surindexation ou une granularité d'indexation trop fine pouvant mener à du bruit documentaire.

9.2 Les relations hiérarchiques du thésaurus

Il s'agit maintenant de comparer l'expressivité des données terminologiques par rapport aux données du thésaurus.

Des 33 occurrences terminologiques identifiées par TerminoMines dans le témoignage de Jean, 23 sont incluses dans le thésaurus dont 15 comme descripteurs et 8 comme non-descripteurs. Parmi les 10 occurrences non incluses dans ThesoMines : *caillou*, *clayats*, *exploitation*, *mine*, *noir*, *pendage*, *risque*, *taper*, *ahu* et *triage*. Ces occurrences ont été rejetées car considérées comme trop génériques pour indexer (*mine*) ou appartenant au domaine de la géologie (*caillou, clayats*) ou trop ambiguës (*noir, risque, triage, exploitation*) ou ne correspondant pas à un substantif (*taper*). Les termes ambigus sont désambiguïsés avant d'être ajoutés dans ThesoMines : *noir* correspondant à *charbon* est inséré comme EP d'*énergie fossile*, *risque* est divisé en *risque naturel* ou *risque sanitaire*, *triage* est divisé en *bâti-triage*, *opération-triage* et *métier-triage*, exploitation en *méthode-exploitation*, *histoire-exploitation*, *arrêt-exploitation*. Les verbes sont rattachés conceptuellement à un substantif, ici *taper* qui signifie 'abattre le charbon' est inséré comme EP d'*abattage*. La figure 73 représente la figure 72 complétée par les informations de ThesoMines (en vert) concernant une des occurrences, *cage*, reliée comme concept à une hiérarchie de concept et se situe au troisième niveau d'une hiérarchie de quatre sous *matériel*, *matériel-extraction*, *tour d'extraction*.

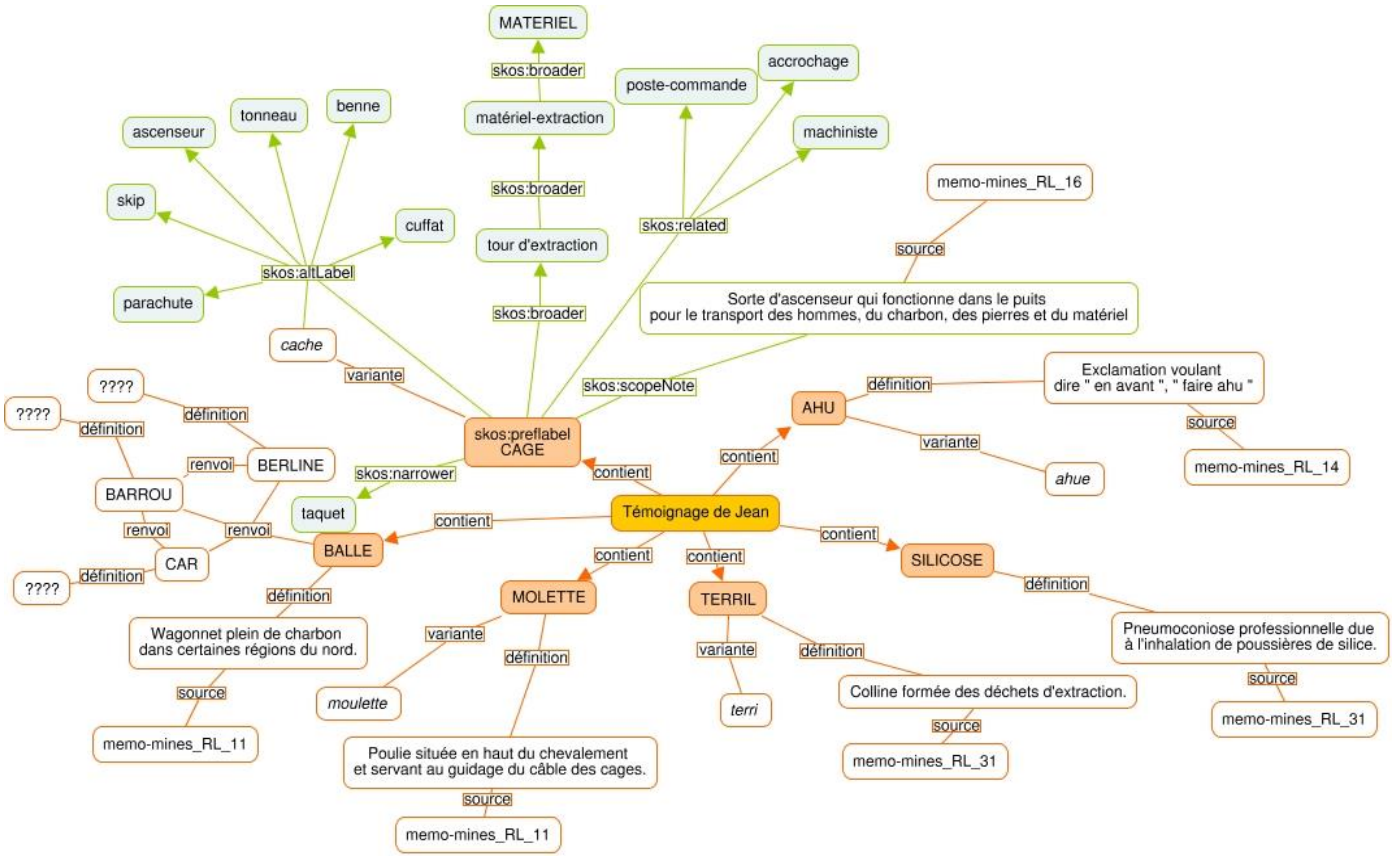


Figure 73 : Extrait des données de TerminoMines et ThesoMines contenues dans le témoignage de Jean

L'indexation des ressources lexicales ou du témoignage par ThesoMines par ce qui devient un « terme-concept » exprime en réalité ce qui apparaît sur le graphe de la figure 73. Les relations hiérarchiques, relations typiques de thésaurus, sont de plusieurs types dans ThesoMines. Néanmoins le format SKOS, considéré comme un modèle simple, ne « se soucie pas de différencier divers types de relations hiérarchiques » (Hadi 2012, 227) et relie les concepts spécifiques aux concepts génériques et inversement grâce à *skos:broader* et à *skos:narrower*. La différence hiérarchique de formalisme entre

les termes de têtes (ou plutôt concepts de tête) et les autres est que les concepts de tête sont notés *skos:hasTopConcept* et tous les concepts spécifiques sont des *skos:Concept*. Concernant cette relation hiérarchique, *cage* est un composant de la tour d'extraction (relation partitive que nous pouvons qualifier de structure), la *tour d'extraction*, fait partie du *matériel d'extraction* (relation partitive) et le *matériel d'extraction* est du *matériel* (relation générique). Afin de marquer la différence entre les termes issus de TerminoMines et ceux qui n'y figurent pas (car issu de l'étape de structuration et de conceptualisation), nous nommons les premiers *termes* et les deuxièmes *termes-concepts*. Le tableau 35 présente quelques exemples de relations hiérarchiques dans ThesoMines dont la spécificité n'est pas différenciée dans le formalisme.

Tableau 35 : Exemples de relations hiérarchiques dans ThesoMines

	Relation entre termes	**Relation entre terme-concept et terme**	**Relation entre termes-concepts**
Partitive TSP	galerie TSP accrochage tour d'extraction TSP cage cage TSP taquet	condition-enfant TSP galibot cité minière TSG coron chantier TSP déblai TSP étage	environnement naturel TSP énergie matériel-extraction TSP machine-extraction TSP tour d'extraction
Générique TSG		substance dangereuse TSG grisou opération TSG abattage bâti TSG bâti-civil TSG bâti-transport	matériel TSG matériel-extraction statut-mineur TSG avantage social
Instance TSI			fête TSI nouvel an événement religieux TSI Sainte-Barbe explosion TSI coup-grisou site et paysage TSI zone tampon classement TSI patrimoine mondial

Par exemple, *condition-enfant* n'est pas un terme issu de TerminoMines, tandis que *galibot* l'est[158]. La distinction entre les termes issus de TerminoMines et les termes-concepts permet de remarquer que les relations entre termes de TerminoMines sont partitives, que les relations entre terme-concept et termes de TerminoMines sont partitives et génériques et que les relations entre termes-concepts sont partitives, génériques et d'instance. Ceci permet notamment de montrer la faible structuration de TerminoMines par rapport à ThesoMines et de l'augmentation d'expressivité de ce dernier par rapport au premier, grâce à la conceptualisation. En revanche, au niveau du formalisme cette expressivité n'est pas maximale.

[158] Le premier a été formé sur l'analyse conceptuelle du terme et ce qu'il représentait comme réalité : la condition des enfants, au même titre que la condition féminine (condition-femme) ou la condition des retraités (condition-retraités).

9.3 Le réseau de l'ontologie

L'ontologie est un réseau, c'est-à-dire un « système de représentation des connaissances prenant la forme d'un graphe composé de nœuds représentant des objets, des concepts ou des événements reliés par des arcs qui décrivent les relations entre les nœuds » (GDT, OQLF 2005). Ce type de représentation, exprimé par un formalisme spécifique, est ici questionné par rapport à la capacité du CIDOC CRM à exprimer la connaissance de notre domaine.

9.3.1 Une couverture sémantique aussi générique que spécifique

Comme le mentionnent Hammache et Ahmed-Ouamer, une indexation par ontologie augmente la pertinence de la recherche d'information, « car l'ontologie est un formalisme qui offre une couverture sémantique plus large et appropriée pour un domaine donné [qu'une indexation par thésaurus] et permet également d'adjoindre à ce modèle d'indexation des inférences par l'exploitation de la sémantique des liens existants entre les concepts de l'ontologie » (Hammache et Ahmed-Ouamer 2006, 903). Cette section évalue la capacité du CIDOC CRM à représenter le patrimoine minier en étudiant finement son instanciation, sa couverture sémantique en partant d'un exemple, celui de l'occurrence *cage* identifiée dans le témoignage. La figure 74 représente l'ajout d'une couche sémantique issue du modèle d'OntoMines en bleu (lui-même issu de CIDOC CRM v.6.2.1). OntoMines permet de décrire et d'instancier des *E31 Document* qui correspondent ici au média du témoignage de Jean, à ses segmentations en tour de parole et aux ressources lexicales (encadré au fond jaune sur la figure 74) et de les relier à leur contenu avec la propriété *P70 documents* dont le rang peut être tout élément décrit par OntoMines qui est *E1 CRM Entity*.

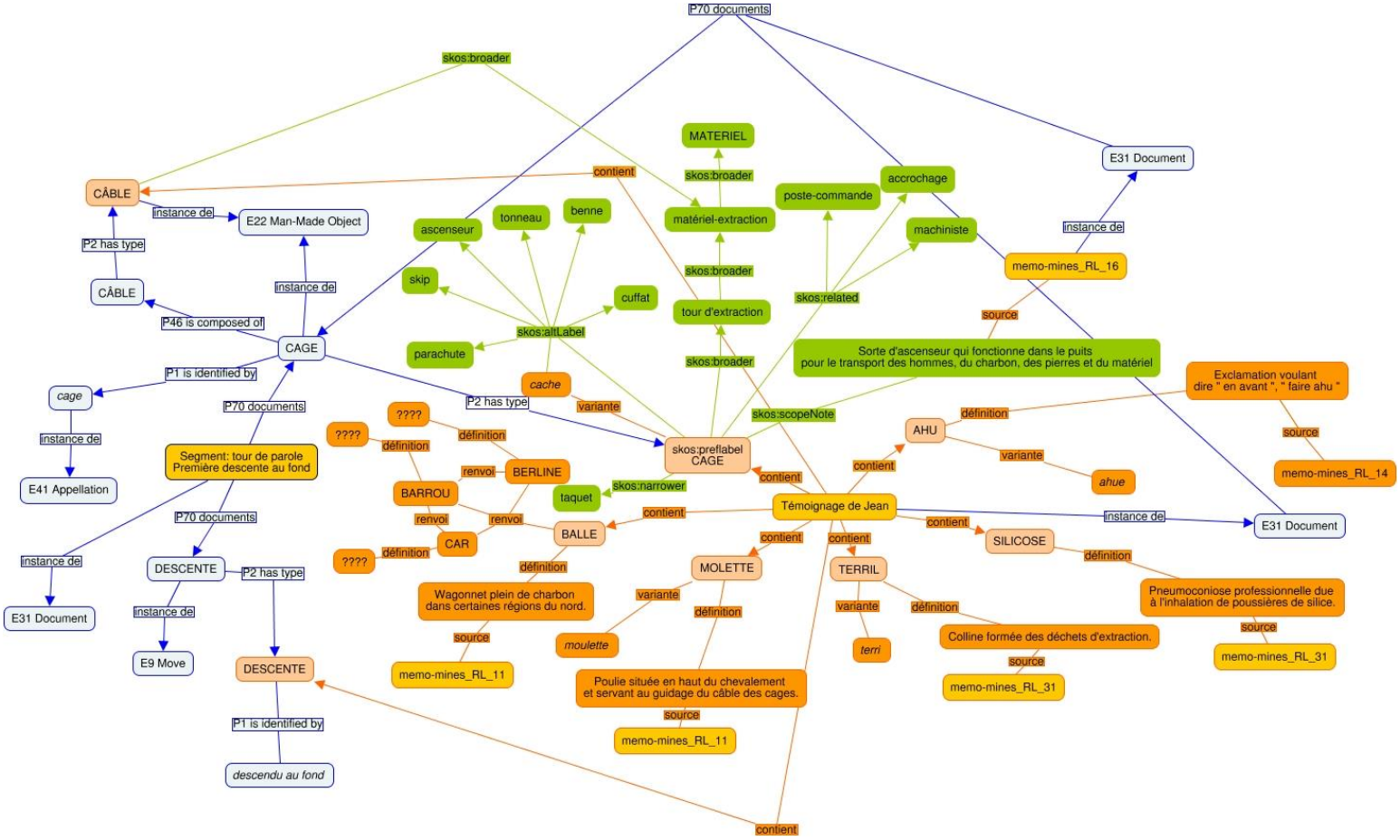


Figure 74 : Extrait des données d'OntoMines, TerminoMines et ThesoMines contenues dans le témoignage de Jean. Disponible à https://cmapscloud.ihmc.us:443/rid=1XBQ4P7XD-1GV9CSL-1DTD4R

Ainsi, l'instance *cage* de *E22 Man-Made Object* peut être reliée au descripteur *cage* de ThesoMines, au document du témoignage de Jean (et au segment du document) mais aussi à la ressource lexicale *memo-mines_RL_16*. Le lien au segment du document vidéo permet de replacer le terme *cage* dans son contexte, et d'augmenter l'expressivité du formalisme de l'ontologie avec la prononciation du terme (audio), les émotions de la personne interviewée et l'environnement filmé (ce n'est pas le cas pour le témoignage de Jean dont le fond est vert mais d'autres documents représentent des objets ou des lieux de la mine. C'est par exemple le cas pour le témoignage d'Edmond qui présente la cage sur une maquette à Arenberg et explique comment et avec quel type de bois celles-ci sont guidées à l'intérieur du puits (figure 75).

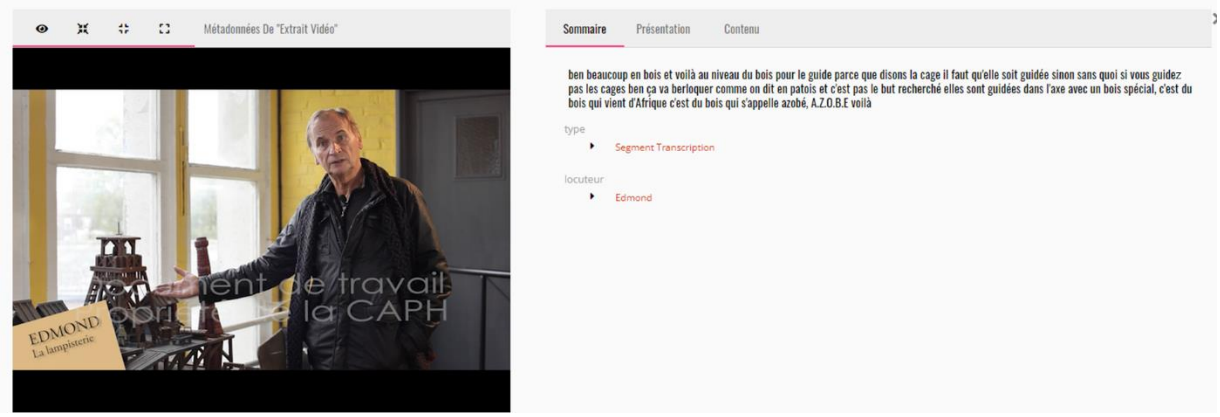


Figure 75 : Maquette représentant une cage et explications par un ancien mineur, Edmond à Arenberg. Document déposé sur la plateforme OKAPI.

Le contexte du document étudié *Première descente au fond* permet de relier l'instance *E22 cage* à l'instance *E22 câble* grâce à une relation sémantique de structure, relation sémantique absente dans ThesoMines. Dans ce SOC, les concepts *cage* et *câble* sont en effet spécifiques à *matériel-extraction* mais n'entretiennent pas de relation entre eux. Au sein du modèle OntoMines la relation hiérarchique partitive *P46 is composed of (forms part of)* permet en effet de relier deux *E18 Physical Thing* (père de *E22 Man-Made Object*). Quant à la relation hiérarchique d'instance, elle est notée par la relation qui unit l'instance *cage* et la classe *E22* et d'une manière générale toutes les instances à leur classe. Enfin la relation hiérarchique générique est décrite par les propriétés ayant pour rang *E55 Type* et la propriété *P127 has broader term (has narrower term)* qui permet de réaliser la même relation qu'effectue SKOS avec *skos:broader* et *skos:narrower*. En revanche, tandis que SKOS traite de concept avec skos:Concept, le CIDOC CRM nomme « *term* » les instances de *E55 Type*. Il semblerait que les deux formalismes soient sur ce point contradictoires.

Nous venons de le voir, le formalisme du CIDOC CRM permet donc d'exprimer aussi bien des connaissances très spécifiques liées à des objets décrits lors de témoignages que des documents et le lien entre ceux-ci. Par rapport au formalisme de ThesoMines, celui d'OntoMines augmente le pouvoir d'expressivité grâce à la décomposition de relations sémantiques spécifiques. Certaines propriétés ont en effet des sous-propriétés (qui ont entre elles un lien de relation générique, par exemple *P1 is identified by (identifies)* a comme sous-propriété *P102 has title (is title of)*, ceci permettant de préciser le type de classe, une appellation (*E41 Appellation*) dans le premier cas ou un titre (*E35 Title*) dans le deuxième cas, *E41* étant générique à *E35*. Cette couche ou colle sémantique découpée en généricité soulève l'hypothèse que la richesse d'expression des ontologies se situe dans leur capacité à formaliser des expressions verbales, donc à se rapprocher du langage naturel.

9.3.2 Limites sémantiques du modèle et ajustements nécessaires

Dans le cadre de l'instanciation au niveau le plus fin, nous avons pu relever des limites à l'expressivité du modèle par rapport au domaine à décrire et quelques ambiguïtés nécessitant d'expliciter clairement les choix effectués.

Concernant les fluides et les gaz (comme le *grisou*, par exemple, qui est un gaz contenu dans certaines couches de houille qui se dégage lors de l´abattage, qui est inflammable et qui est à l'origine de nombreuses catastrophes dans le bassin minier), le CIDOC CRM précise pour la classe *E18 Physical Thing* que « *The CIDOC CRM is generally not concerned with amounts of matter in fluid or gaseous states* ». Ceci peut être une limite mais nous avons fait le choix d'utiliser tout de même cette classe pour l'instancier car les propriétés permettent de décrire des réalités importantes, par exemple qu'une explosion a été influencée par du grisou dans un lieu ou encore la composition du grisou. Jean ne mentionne pas celui-ci, nous illustrons donc avec d'autres extraits d'entretiens :

Un membre de l'association des amis de la mine mentionne lors de la visite : « le grisou c'est du méthane c'est la bouteille de butane que vous avez chez vous c'est le même gaz ». Ceci peut être décrit par :

E18 Physical Thing (*grisou*) : *P46 is composed of* : *E18 Physical Thing* (*méthane*)

Ensuite, Edmond relie la qualité du charbon au danger, plus celui-ci est gras avec des matières volatiles, plus il y a de poussière dans l'air et comme le charbon est composé de grisou, lui-même composé de méthane, il s'enflamme plus facilement : « […] mais c'est du charbon maigre de très bonne qualité et dans le Pas-de-Calais c'est du charbon gras avec beaucoup de matière volatile, d'accord, qui dit beaucoup de matière volatile, plus de danger avec le grisou déjà. Comme vous le savez en 1906 à Courrières, à Courrières, coup de grisou 1099 victimes […] »

Le CIDOC CRM permet de décrire la cause du coup de grisou, le lieu touché par celui-ci, la date de l'événement, le fait que le coup de grisou a provoqué mais ne permet pas de décrire le nombre de victimes à moins de n'instancier chacune des 1099 victimes ce qui peut être soulevé comme limite.

E7 Activity (coup-grisou) : P15 was influenced by : E18 Physical Thing (grisou)

E7 Activity (coup-grisou) : P7 took place at : E53 Place (Courrières)

E7 Activity (coup-grisou) : P1 is identified by : E50 Date (1906)

E69 Death (mort) : P117 occurs during : E7 Activity (coup-grisou)

Ensuite, lorsque deux instances ont été identifiées dans un corpus oral et que ces deux sont reliées entre elles sémantiquement, il faut pouvoir identifier le sens exact transmis par la relation (qui est d'ailleurs parfois implicite). Or, le doute peut s'installer lorsque le sens porté par une propriété est quasi-équivalent et non exactement équivalent (de la même façon qu'il existe la notion de quasi-synonymes et synonymes exacts). Le sens proche porté par certaines propriétés sélectionnées lors du premier travail peut paraître ambigu. C'est le cas, par exemple, avec la *P49 has former or current keeper* et la *P51 has former or current owner*. Si l'on traduit littéralement, la *P49* mentionne la notion de « gardien », alors que la *P51* mentionne un « propriétaire » (légal d'après le CIDOC CRM). Dans le contexte muséal, ceci se comprend : un organisme conserve un objet dans ses locaux dont il n'est pas le propriétaire légal. Imaginons qu'un marteau-piqueur appartienne à un ancien mineur et qu'il l'ait prêté à un musée pour une exposition. Le musée devient le gardien de l'objet, tandis que l'ancien mineur reste son propriétaire

légal. Prenons maintenant l'extrait de l'entretien de Jean suivant : « […] on accepte d'être muté dans le Nord à une condition d'avoir une maison et ben croyez-moi quand on arrive de là-bas et qu'on a une maison à nous des Houillères […] ». Ici, nous relevons la mention à l'habitat minier avec « maison », la mention d'appartenance avec « à nous », et une deuxième mention d'appartenance avec « des Houillères ». Lorsque l'on a connaissance du contexte, nous savons que les Houillères étaient propriétaires de maisons qu'elles mettaient à disposition des anciens mineurs en échange de leur travail à la mine. Dans ce cas, la notion de propriétaire est ambigüe et elle ne peut réellement se superposer à l'exemple d'un objet dans un musée. Nous précisons donc que la *P49* lie l'objet à un occupant et la P51 à un propriétaire légal. Ce qui est décrit :

E27 Site (maison) : P49 has former or current keeper (is former or current keeper of) : E74 Group (Jean)

E27 Site (maison) : P51 has former or current owner (is former or current owner of) : E21 Person (Houillères)

Le CIDOC CRM possède une souplesse qui permet d'adapter la description à des domaines plus précis, en revanche, il est primordial d'expliciter clairement les choix effectués lorsqu'une ambiguïté se présente. Ceci peut en effet avoir des conséquences sur l'échange ou la comparaison des données finales et implique des problèmes d'interopérabilité au niveau sémantique : même si des instances se situent sous la même classe, les choix d'instanciation peuvent mener à avoir des instances qui ne peuvent être comparées car elles ne sont pas de même nature.

Ensuite, le modèle ne permet pas de déclarer la négation d'un énoncé.

Par exemple, Jean déclare : « ah au fond non moi j'ai pas connu, […], non de notre temps euh j'ai pas connu de cafus au fond, mais peut-être qu'il y en avait avant […] ». La connaissance transmise par cet extrait est : lors de la carrière de Jean (1950-1986), les trieuses (cafus) ne travaillaient pas au fond. Le modèle ne permet pas d'exprimer ceci, mais permet d'exprimer un sens équivalent grâce à l'affirmation contraire. Lors de la carrière de Jean, les cafus travaillaient au jour, *au jour* étant le contraire de *au fond*. Le formalisme du CIDOC CRM permet finalement de décrire cette connaissance :

E7 Activity (carrière de Jean) : P14 carried out by : E39 Actor (Jean)

E7 Activity (carrière de Jean) : P160 has temporal projection : E52 Time-Span : P81 ongoing throughout : E61 Time Primitive (1950-1986)

E7 Activity (activité -triage- réalisée par les cafus lors de la carrière de Jean) : P14 carried out by : E74 Group (cafus)

E7 Activity (triage) : P160 has temporal projection : E52 Time-Span : P81 ongoing throughout : E61 Time Primitive (1950-1986)

E7 Activity (triage) : P161 has spatial projection : E53 Place (au jour)

ou E7 Activity (triage) : P7 took place at : E53 Place (au jour)

E53 Place (au jour) : P87 is identified by : E48 Place Name (carreau de la fosse n°6 des mines de Marles)

La possibilité du CIDOC CRM d'exprimer cette finesse dans la description sémantique des connaissances montre son réel potentiel. Une bonne connaissance du formalisme est toutefois nécessaire afin de pouvoir inférer la connaissance entre classes et propriétés génériques et spécifiques lors de l'instanciation. Même si nous avons choisi certaines propriétés et classes du CIDOC CRM pour former le modèle OntoMines, nous n'avons pas arrêté le niveau sémantique, c'est-à-dire que si nos schémas préconisent d'utiliser telle propriété pour l'instanciation de la classe *E18 Physical Thing*, l'instanciation peut se réaliser grâce à toute instance des fils de *E18*, ce qui augmente considérablement les possibilités d'instancier : par les instances de ses fils *E19 Physical Object*, *E24 Physical Man-Made Thing*, *E26 Physical Feature* et toutes les instances de leurs fils (donc des classes *E20, E22, E21, E84, E25, E78, E84, E27*). Ainsi, comme le montre la figure 76, la classe *E6 Destruction* peut décrire la destruction de puits, de sites miniers, d'animaux, de personnes, d'objets, de supports d'information, de collections.

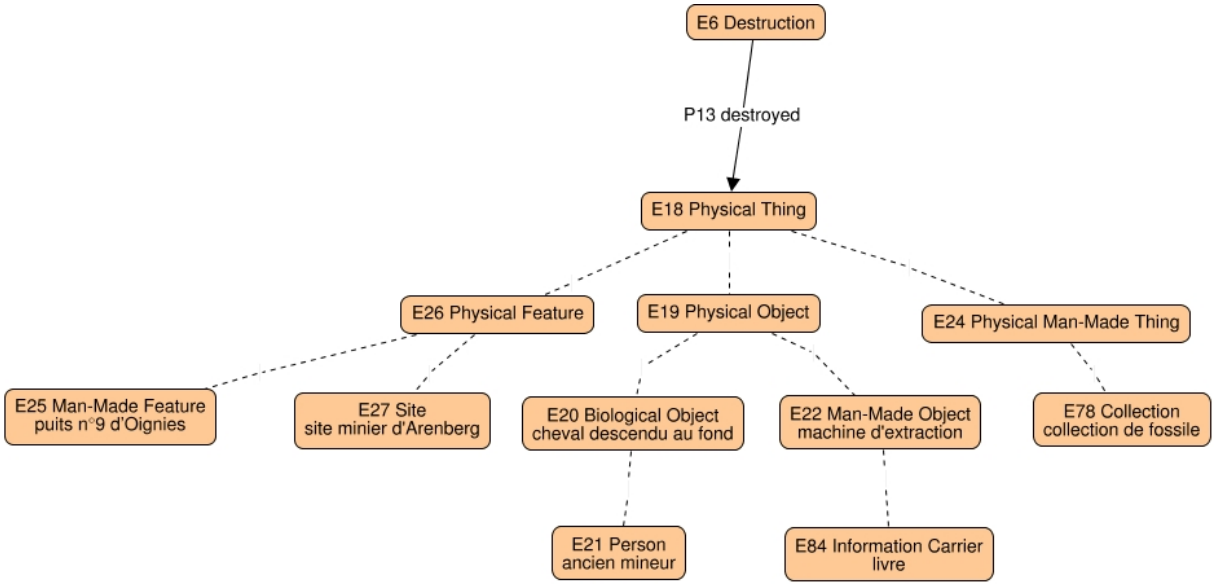


Figure 76 : Instanciation de l'activité de destruction des *E18 Physical Thing* par héritage de classes

Enfin, le langage naturel est tel qu'il peut désigner un objet par son matériau (dans les mines, un *bois* désigne le *matériel de soutènement* mais aussi le *matériau* utilisé pour soutenir le soutènement). L'instanciation permet de désambiguïser. D'un côté, on a un *E22 Man-Made Object* (*bois* dont le descripteur est *matériel-boisage* dans ThesoMines) et de l'autre on a un *E57 Material* (*bois*). Avec le CIDOC CRM, il est possible de décrire un déplacement des instances de *E22*, mais pas de *E57*. Il n'est donc pas possible de décrire le déplacement d'un matériau, mais l'objet fait de matière. Ainsi, l'extrait du témoignage de Jean : « ben au début on transportait les matériaux pour les vieux c'est-à-dire on leur apportait les bois, les étançons et après ils nous donnaient les marteaux-piqueurs pour nous former sur le tas » entre en contradiction avec le modèle, Jean mentionnant le transport de matériau. Ceci est instancié :

E9 Move (transport) : P25 moved : E22 Man-Made Object (matériel-boisage, étançons…)

E9 Move (transport) : P14 carried out by (performed) : E39 Actor (mineur)

E22 Man-Made Object (matériel-boisage, étançons…) : P2 has type : E57 Material (bois)

E11 Modification (soutènement) : P126 employed : E57 Material (bois)

L'instanciation de propos oralisés à l'aide d'un modèle oblige à remettre en question la « vérité » des propos, ce ne sont pas les matériaux mais les objets composés de matériau que Jean transportait avec ses collègues.

Pour conclure, étudier l'expressivité des formalismes des SOC, donc leur capacité à représenter du sens et des connaissances par l'interprétation de leur seule forme montre que même le système ayant le formalisme le plus riche sémantiquement (OntoMines) ne permet parfois pas de dépasser l'implicite de l'interprétation ou se heurte une nouvelle fois à l'ambiguïté de la langue naturelle.

CHAPITRE 10. Interopérabilité

Ce chapitre étudie l'interopérabilité des trois SOC, c'est-à-dire la capacité qu'ils ont à pouvoir échanger ou non des informations entre eux. Selon la norme ISO 25964[159], l'interopérabilité se définit comme la « capacité de deux ou plusieurs systèmes à échanger des informations et à utiliser les informations qui ont été échangées avec une perte minimale de contenu ». Hudon définit l'interopérabilité comme la « capacité qu'ont certains agents, services, systèmes et applications informatiques d'échanger des données, de l'information et des connaissances en préservant l'intégrité et la pleine signification de celles-ci » (Hudon 2013, s.p.). Afin de préserver ce qu'Hudon nomme « intégrité » et « pleine signification », il est nécessaire d'étudier l'interopérabilité syntaxique des SOC d'une part et leur interopérabilité sémantique d'autre part. La première concerne le langage des données utilisé. L'interopérabilité syntaxique permet l'échange de données via un format de données commun défini pour l'encodage, le décodage et la représentation des données grâce aux normes et standards du web sémantique tels que RDFS, SKOS et OWL (Zeng 2019). La seconde concerne le contexte des données. L'interopérabilité sémantique permet en effet l'échange de données via des contextes qui sont prédéfinis et livrés dans les vocabulaires des SOC (Zeng 2019). Tandis que Zeng étudie comment l'ISO 25964 traite de l'interopérabilité des thésaurus avec les autres types de SOC tels que les « *classification schemes, taxonomies, subject heading schemes, ontologies, terminologies, name authority lists, and synonym rings* » (Zeng 2019), nous nous intéressons plus particulièrement à l'interopérabilité sémantique portée par l'utilisation du modèle ontologique du CIDOC CRM.

10.1 Interopérabilité syntaxique

10.1.1 ThesoMines et SKOS

Pour représenter ThesoMines, nous avons choisi le langage SKOS qui est un langage formel permettant une représentation structurée des concepts de thésaurus, des classifications ou autre type de vocabulaire contrôlé. Outre la structuration des concepts entre eux, ce langage détermine leur rattachement à des instances lexicales (descripteurs et non-descripteurs). La figure 77 est un extrait du fichier SKOS de ThesoMines qui illustre le concept de puits d'aérage (*puits-aérage*) représenté par l'identifiant suivant :

https://opentheso.huma-num.fr/opentheso/?idc=120410&idt=th136.

[159] ISO 25964-2:2013 Information et documentation — Thésaurus et interopérabilité avec d'autres vocabulaires — Partie 2: Interopérabilité avec d'autres vocabulaires.

```
<rdf:Description rdf:about="https://opentheso.huma-num.fr/opentheso/?idc=120410&amp;idt=th136">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="fr">puits-aérage</skos:prefLabel>
    <skos:altLabel xml:lang="fr">entrée d'air</skos:altLabel>
    <skos:altLabel xml:lang="fr">intrée d'air</skos:altLabel>
    <skos:altLabel xml:lang="fr">puits d'aérage</skos:altLabel>
    <skos:altLabel xml:lang="fr">puits d'entrée d'air</skos:altLabel>
    <skos:altLabel xml:lang="fr">puits de retour d'air</skos:altLabel>
    <skos:altLabel xml:lang="fr">retour d'air</skos:altLabel>
    <skos:altLabel xml:lang="fr">r'tour d'air</skos:altLabel>
    <skos:related rdf:resource="https://opentheso.huma-num.fr/opentheso/?idc=120962&amp;idt=th136"/>
    <skos:broader rdf:resource="https://opentheso.huma-num.fr/opentheso/?idc=120383&amp;idt=th136"/>
    <skos:inScheme rdf:resource="https://opentheso.huma-num.fr/opentheso/th136"/>
    <opentheso:memberOf rdf:resource="https://opentheso.huma-num.fr/opentheso/?idg=G1220&amp;idt=th136"/>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2020-02-04</dcterms:created>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2020-05-16</dcterms:modified>
    <dcterms:identifier>120410</dcterms:identifier>
    <dcterms:description>152879##120369##120383##120410</dcterms:description>
    <skos:scopeNote xml:lang="fr">Indexe le puits d'entrée d'air et celui de retour d'air en tant que ventilation naturelle.
    Pour la ventilation forcée, voir ventilateur dans matériel-aérage.</skos:scopeNote>
```

Figure 77 : Extrait du fichier SKOS de ThesoMines v1.0

Chaque concept de ThesoMines possède donc un identifiant et est caractérisé par un descripteur *skos:prefLabel* (ici *puits-aérage*) et éventuellement par des non-descripteurs *skos:altLabel* (ici *entrée d'air*, *intrée d'air* etc.).

La publication de ThesoMines sur le web grâce à la plateforme Opentheso, qui fournit les données en SKOS, permet d'aligner facilement celles-ci avec celles d'autres vocabulaires ou thésaurus en SKOS et de rendre donc ThesoMines interopérable au niveau syntaxique. Cette interopérabilité facilite l'enrichissement de la ressource et par conséquent, la vision sur le domaine. Il est par exemple possible d'aligner un concept avec celui d'un autre vocabulaire grâce à *skos:exactMatch* (décrivant une équivalence totale), *skos:closeMatch* (décrivant une équivalence partielle), *skos:narrowMatch* (décrivant une équivalence plus spécifique), *skos:relatedMatch* (décrivant une équivalence associée), *skos:broadMatch* (décrivant une équivalence générique). Ainsi, il est possible de relier le concept de *puits* de ThesoMines au concept *Puits de mine* de RAMEAU ou au concept *shaft* dans Wikidata (figure 78). Enfin, il est également possible de fusionner ThesoMines avec un thésaurus ou un microthésaurus[160] plus spécifique, comme un thésaurus sur les minerais ou la géologie à la seule condition qu'il soit lui-même en SKOS.

[160] Subdivision d'un thésaurus.

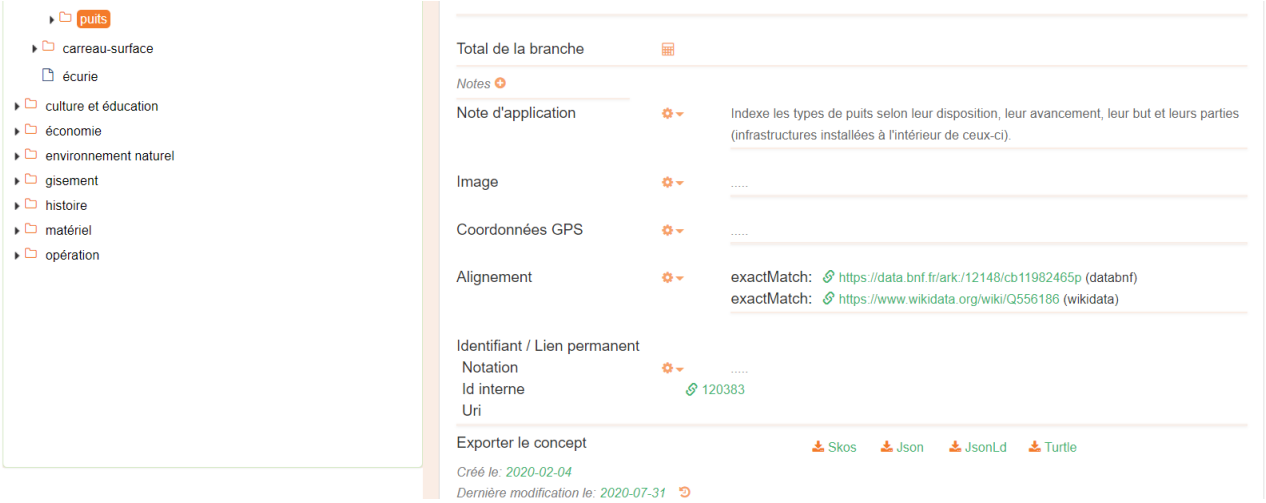


Figure 78 : Alignement du concept puits avec un concept distant dans ThesoMines

Enfin, l'utilisation d'un format standard permet de pouvoir comparer les SOC et leurs données, d'utiliser différents éditeurs de thésaurus (comme Tematres[161], OpenTheso, Ginco[162]) sans perdre d'information lors d'un passage de l'un à l'autre et puis, grâce au web sémantique, de décrire les données contenues dans un thésaurus à l'intérieur d'une ontologie.

10.1.2 TerminoMines et TBX

TerminoMines n'étant pas un livrable mais une base pour la conception de ThesoMines et OntoMines, il ne nous a pas paru nécessaire de la décrire grâce à un standard. Cependant, il aurait été possible d'envisager, si le temps nous l'avait permis, que l'enrichissement de celle-ci soit décrit avec un standard pour permettre l'interaction entre ThesoMines (donc avec OntoMines, puisque ThesoMines et OntoMines sont interopérables à travers la mise en lien des instances de l'ontologie avec les concepts de ThesoMines) et TerminoMines. Pour cela, nous avons vu en 6.1 que le standard TBX ou *TermBase eXchange* pouvait être utilisé. Celui-ci est définit comme « *the international standard for representing and exchanging information from termbases* »[163]. En général, une ressource terminologique est organisée en entrées de concepts, dont chacune comprend un ou plusieurs termes désignant un concept particulier. Une ressource lexicale est organisée en entrées lexicales, chacune d'entre elles comprenant un ou plusieurs sens d'un élément lexical particulier (un mot ou une phrase). Une entrée de concept contenant plusieurs termes peut être divisée en plusieurs entrées lexicales, une par terme, et plusieurs entrées lexicales associées au même concept peuvent être combinées en une seule entrée de concept. TBX permet de faire la différence entre chacune de ces données et fournit un vocabulaire à utiliser. Concernant TerminoMines, les champs conservés peuvent être mis en relation simplement avec la syntaxe de TBX : ainsi les entrées sont des *term*, l'identifiant correspond à un *termEntry* ou *ConceptEntry*, le domaine (du patrimoine minier) peut être noté *subjectField* etc. Le tableau 36 ci-

[161] https://www.vocabularyserver.com/
[162] https://culturecommunication.github.io/ginco/
[163] https://www.tbxinfo.net/ et https://www.iso.org/standard/45797.html

dessous reprend les champs de TerminoMines et les aligne avec les catégories de TBX en citant les définitions de chacune.

Tableau 36 : Alignement des champs de TerminoMines avec les catégories de TBX

Champs disponibles dans TerminoMines	**ISO 30042:2019 (TBX)**	**Définition**[164]
Identifiant	*termEntry ou Concept Entry*	*The <tbx:termEntry> element is the root element for a terminological entry, which contains all the elements, and their attributes, that describe a concept and all the terms that denote that concept, potentially in several languages.* *The <tbx:termEntry> element is often referred to as the "concept level" of the entry. At this level, only external-type references (references that point to locations outside of the document) plus the various <tbx:langSet> elements are allowed.*
Entrée	*term*	*The <tbx:term> element encloses one representation of a concept by a sign which denotes it. This sign can be a textual "term", a symbol, or a graphic. In glossaries and other terminology resources, designations are primarily "terms", which is why the element <tbx:term> is used for all three types.*
Domaine	*subjectField*	*A field of special knowledge, such as information technology or leather industry. Specifying a subject field can help to clarify the field in which a particular term or concept is used.*
Source bibliographique	*source*	*According to ISO 10241-1, if a terminological entry, language section, term, definition, non-verbal representation (such as a symbol) or example has been taken from another standard, or another authoritative source, the source should be indicated. Enter the name of the source document in the <tbx:source> element. If the text taken from the source was modified in any way, this should be indicated in an accompanying note. The data or data category to which the source applies may be also be indicated in a note, for instance, to specify that the source applies to an example or to the definition.*
Langue/Dialecte	*langSet*	*The <tbx:langSet> is a container element that holds all the information in a terminological entry pertaining to one language, including all the TIG elements (terms and associated information) for that language.*

En 2019, Reineke et Romary étudient le pont possible entre TBX et SKOS qui sont deux formalismes régis par des modèles de données similaires, à savoir la description de concepts (plutôt que de mots). D'après les auteurs, le problème d'interopérabilité concerne la granularité de la description des informations et la flexibilité d'instanciation. Tandis que TBX est clair et approfondi, avec une grande flexibilité d'instanciation, SKOS fournit un formalisme de description relativement plat, rigide et partiellement ambigu (Reineke et Romary 2019, 27). Aux auteurs d'ajouter : « *Consequently, the difference in the diversity of instantiable information units does not affect data integrity when mapping from SKOS to TBX, but generally leads to data impoverishment when mapping from TBX to SKOS.* »

10.1.3 OntoMines et OWL/RDFS

Nous l'avons vu, RDF est le langage de base du web sémantique qui vise à munir le web d'un modèle de données ayant une structure de graphe. Toute information en RDF est donc représentée en triplet dont un sujet, un prédicat et un objet. L'une des syntaxes (ou sérialisations) de ce langage est RDF/XML. RDFS est un langage qui permet de décrire les classes et les propriétés pour des ressources RDF. OWL,

[164] Définitions fournies à https://www.niso-sts.org/TagLibrary/niso-sts-TL-1-0-html/

nous l'avons vu, est un langage fondé sur XML qui permet de représenter une ontologie. Au niveau syntaxique, OWL et RDFS sont privilégiés. Ils permettent en effet une interopérabilité avec les moteurs d'inférence du web sémantique. Certains moteurs d'inférence ne peuvent raisonner qu'au niveau des classes et des propriétés alors que d'autres permettent également de raisonner sur les instances de classes. Outre la possibilité d'être interopérables avec les moteurs d'inférence, OWL et RDFS permettent une organisation technique favorisant l'interrogation de la ressource via tout mécanisme de type graphe de connaissances.

Or, le choix du CIDOC CRM par le projet MémoMines implique l'enregistrement, la description des données grâce à des standards du web sémantique et notamment en RDF, RDFS et en OWL. La figure 79 illustre un extrait du fichier RDF/XML d'OntoMines instanciée. RDF permet de préciser le type de classe, ici E7 Activity, mais aussi le fait que l'instance créée est un type de connaissance partagée dans l'ontologie générale d'OKAPI (ina.fr/core#CommonKnowledge), RDFS permet de nommer l'instance par une chaîne de caractères, ici *avantage social_entité* ou de préciser que la classe *E71 Man-Made Thing* du CIDOC CRM est une sous-classe de *E18 Physical Thing*. Enfin, ns5 désigne l'URL du CIDOC CRM http://www.cidoc-crm.org/cidoc-crm/ et exprime la sémantique, les propriétés *P17*, *P1*, *P9* reliées à des ressources RDF qui sont des instances ou des concepts de ThesoMines (pour la *P2*).

```
<rdf:Description rdf:about="http://www.memomines.fr/avantage_entite">
  <rdf:type rdf:resource="http://www.cidoc-crm.org/cidoc-crm/E7_Activity" />
  <rdf:type rdf:resource="http://www.ina.fr/core#CommonKnowledge" />
  <rdfs:label xml:lang="fr">avantage social_entité</rdfs:label>
  <ns5:P17_was_motivated_by rdf:resource="http://www.campus-AAR.fr/resource_217773642" />
  <ns5:P17_was_motivated_by rdf:resource="http://www.campus-AAR.fr/resource_1296570981" />
  <ns5:P1_is_identified_by rdf:resource="http://www.campus-AAR.fr/resource_261988516" />
  <ns5:P2_has_type rdf:resource="https://opentheso.huma-num.fr/opentheso/?idc=120474&amp;idt=th136" />
  <ns5:P9_consists_of rdf:resource="http://www.memomines.fr/retraite_entite" />
```

Figure 79 : Extrait d'OntoMines, instance « avantage social »

Cet extrait montre la possibilité de décrire l'ontologie au sein d'autres ontologies (ici celle de l'INA).

10.2 Interopérabilité sémantique

10.2.1 Entre SOC

Le modèle OntoMines a été construit en faisant correspondre les concepts de ThesoMines aux classes du CIDOC CRM, rendant les deux SOC interopérables au niveau sémantique. Par exemple, OntoMines permet de passer du descripteur *bâti* à la classe *E22 Man-Made Object* ou du descripteur *cimetière* à la classe *E27 Site*. Nous l'avons vu, ce passage est effectué grâce à la propriété *P2 has type* dans OntoMines. Le tableau 37 illustre la mise en correspondance des descripteurs de ThesoMines (troisième colonne) à la classe correspondante dans le CIDOC CRM (deuxième colonne) et à un *rdfs:label* (première colonne) c'est-à-dire à la forme que l'instance prendra lors de la visualisation des données[165].

[165] « *rdfs:label is an instance of rdf:Property that may be used to provide a human-readable version of a resource's name* »
https://www.w3.org/TR/rdf-schema/#ch_label

Tableau 37 : Passage des concepts de ThesoMines à OntoMines		
rdfs:label	**URI OntoMines**	**cidoc-crm:P2_has_type**
abatteur_(personne)_E	http://www.cidoc-crm.org/cidoc-crm/E21_Person	https://opentheso.huma-num.fr/opentheso/?idc=120825&idt=th136
abatteur_(activité)_E	http://www.cidoc-crm.org/cidoc-crm/E7_Activity	https://opentheso.huma-num.fr/opentheso/?idc=120825&idt=th136
abatteux_(appellation)_E	http://www.cidoc-crm.org/cidoc-crm/E41_Appellation	https://opentheso.huma-num.fr/opentheso/?idc=120825&idt=th136
about_(personne)_E	http://www.cidoc-crm.org/cidoc-crm/E21_Person	https://opentheso.huma-num.fr/opentheso/?idc=120829&idt=th136
about_(activité)_E	http://www.cidoc-crm.org/cidoc-crm/E7_Activity	https://opentheso.huma-num.fr/opentheso/?idc=120829&idt=th136
accident de terrain_(appellation)_E	http://www.cidoc-crm.org/cidoc-crm/E41_Appellation	https://opentheso.huma-num.fr/opentheso/?idc=120899&idt=th136
accident de transport_(appellation)_E	http://www.cidoc-crm.org/cidoc-crm/E41_Appellation	https://opentheso.huma-num.fr/opentheso/?idc=120823&idt=th136

Le travail de mise en correspondance entre descripteurs et classes du modèle OntoMines est disponible en annexe 4. L'information concernant la différence de spécificité entre les concepts de ThesoMines par rapport au domaine et la généralité des classes d'OntoMines (CIDOC CRM) reste disponible puisque grâce aux identifiants uniques (syntaxe), le passage d'un concept de ThesoMines à une classe d'OntoMines ne modifie pas la structure de départ. Les liens, dont la nature est donnée grâce au niveau sémantique du modèle du CIDOC CRM, sont tout d'abord effectués entre les instances puis, par héritage, entre les classes, ce qui permet d'avoir à la fois une description fine du domaine et une description plus large et deux niveaux d'interopérabilité sémantique.

10.2.2 Entre les ressources

L'hétérogénéité des ressources classificatoires utilisées pour représenter le domaine a nécessité un alignement au niveau sémantique et conceptuel menant aux termes-concepts de ThesoMines : l'utilisateur doit en effet pouvoir retrouver un même contenu à partir de plusieurs terminologies différentes, ThesoMines participe en ce sens à l'interopérabilité sémantique entre les ressources classificatoires (Cf 5.3, figure 37).

10.2.3 Entre les domaines

Nous l'avons vu, le modèle du CIDOC CRM est largement étudié au sein de la communauté scientifique travaillant sur le patrimoine culturel. Lors de la conception du modèle d'OntoMines, le choix de ne pas étendre le modèle du CIDOC CRM a été fait dans l'optique de pouvoir échanger des données entre différents domaines. Pour illustrer cette interopérabilité, nous avons vu que le choix du CIDOC CRM dans deux projets culturels tels que le textile et la mine pouvait permettre d'étudier ceux-ci sur un même niveau de granularité (Kergosien et *al.* 2019) notamment au niveau géographique grâce à la classe *E53 Place*. Par exemple, la mention à la ville de Roubaix (*E53 Place*) dans le domaine minier dans l'entretien

de Jean (Cf. 8.1.2 Extrait (4)), et la mention à Roubaix (*E53 Place*) dans le domaine du textile (Kergosien et *al.* 2019) pourrait permettre d'étudier le lien entre les mines avec les trieuses allant travailler dans les filatures à Roubaix et des événements qui se sont déroulés à Roubaix dans le cadre du textile telle que l'*Exposition internationale de Roubaix* (*E5 Event*). Ceci serait seulement réalisable à condition d'avoir des entités temporelles concordantes (l'exposition internationale de Roubaix a eu lieu en 1911, le travail des trieuses dans les filatures pourrait correspondre temporellement puisqu'il est mentionné dans un document de la fresque de l'INA[166] : « Le travail souterrain des femmes a été interdit à partir des 1874 et dès la fin du siècle, elles ne représentent plus que 5,17 % du personnel total de la Compagnie des mines d'Aniche ». On peut d'ailleurs remarquer que cet événement est mentionné dans un document sur le site des archives nationales du monde du travail (ANMT)[167] qui présente le synopsis d'une exposition sur « *Mines et mineurs entre réalité et imaginaire*. *Deux siècles d'archives privées et publiques de l'Ancien Régime à Charbonnages de France* ». Cet exemple illustre tout à fait le rôle du CIDOC CRM : « *The primary role of the CIDOC CRM is to serve as a basis for mediation of cultural heritage information and thereby provide the semantic "glue" needed to transform today's disparate, localized information sources into a coherent and valuable global resource* »[168].

CONCLUSION DE LA PARTIE 3

La partie 3 consistait à évaluer les SOC selon quatre critères : la couverture informationnelle, la granularité, l'expressivité de leurs formalismes et l'interopérabilité. Grâce à ces critères nous avons notamment pu montrer en quoi les trois SOC étaient différents mais complémentaires et ce qu'impliquait leur conception. Concernant leur couverture informationnelle, nous avons vu dans le chapitre 7 l'importance de considérer l'hétérogénéité des données qui les compose afin de comprendre en détail leur taux de recouvrement. Le recouvrement peut représenter une partie de l'interopérabilité entre chaque SOC au niveau sémantique. ThesoMines étant le SOC ayant subi le plus de test (utilisation des descripteurs pour indexer, test de la souplesse de la structure, enrichissement), il est considéré comme le SOC le plus complet et donc le plus saturé. En raison de l'aspect chronophage de la conception des SOC, seule l'analyse d'un document a permis l'instanciation d'OntoMines. Plus qu'une ontologie de domaine, le modèle instancié est une ontologie de document et nous ne pouvons affirmer qu'elle représente tout le domaine. En revanche, la finesse d'instanciation permet d'illustrer la richesse de celui-ci et de soulever certaines questions quant à l'utilisation des ontologies pour la sauvegarde du patrimoine culturel.

Nous avons ensuite vu dans le chapitre 8 en quoi le niveau de granularité différait selon le type de SOC grâce à l'analyse de plusieurs entités et classes telles que les acteurs, les entités spatiales, les entités temporelles, les objets matériels et les activités. Nous avons montré la différence entre les entités

[166] https://fresques.ina.fr/memoires-de-mines/fiche-media/Mineur00171/ouvrieres-du-textile-du-bassin-minier.html
[167] https://archives-nationales-travail.culture.gouv.fr/ Document disponible à : https://archives-nationales-travail.culture.gouv.fr/Mediatheque/Images/Expo-mines2/Synopsis-de-l-exposition
[168] http://www.cidoc-crm.org/

nommées précises (ayant une référence exacte au sein d'un système prédéfini) et les liens qui pouvaient être établis entre elles et les autres entités pour créer du sens et des rapprochements entre différents extraits d'un même entretien. L'étude de la granularité autorise des niveaux d'analyse à plusieurs niveaux (connaissance précise ou connaissance générale). Au niveau de l'expressivité des formalismes, le chapitre 9 montrait que les SOC expriment tous les trois des points de vue différents sur le domaine : chacun ayant une sémantique propre, plus ou moins formalisée. L'expressivité du formalisme de l'ontologie est en effet plus élevée que celle du thésaurus et celle du thésaurus est plus élevée que celle de la terminologie. Tandis que le thésaurus est structuré hiérarchiquement, l'ontologie organise ses concepts dans un graphe avec des relations sémantiques et de subsomption. Enfin le chapitre 10 montre que les SOC sont interopérables au niveau syntaxique et au niveau sémantique entre eux, avec d'autres ressources et d'autres domaines, d'une part grâce à l'utilisation des standards du web sémantique qui sont interopérables entre eux (SKOS, OWL, RDFS), d'autre part grâce à la méthodologie de construction des SOC à partir du système conceptuel du domaine du patrimoine minier.

CONCLUSION

Les apports de la thèse sont la réalisation de trois systèmes d'organisation des connaissances : une terminologie (TerminoMines), un thésaurus (ThesoMines) et une ontologie (OntoMines), une méthodologie de conception généralisable et l'apport d'une réflexion sur leurs enjeux conceptuels.

La problématique de notre thèse s'inscrivait dans le cadre du projet MémoMines qui avait pour but de valoriser la mémoire minière, actuellement en danger pour plusieurs raisons. Les raisons qui intéressent nos travaux étaient d'une part, la disparition progressive des anciens mineurs, et d'autre part, la dispersion sur le territoire des ressources la documentant. Dans ce cadre, notre problématique souhaitait répondre à la question suivante :

> Comment représenter exhaustivement le patrimoine houiller de l'ancien bassin du Nord–Pas-de-Calais au vu de l'hétérogénéité de sa documentation et de ses ressources classificatoires, de la dispersion de celles-ci et en prenant en compte les différents points de vue de ses acteurs ?

Nous avons montré en quoi le champ de l'organisation des connaissances fournissait des approches pertinentes pour répondre à ce type de problématique et nous avons notamment souhaité étudier l'approche par analyse de domaine (Hjørland 2008). Le but était de considérer le patrimoine houiller comme un domaine de connaissance. Un domaine peut en effet être représenté par un système conceptuel auquel appartiennent un ensemble de concepts définis ou à définir, par l'ensemble des termes techniques utilisés par des personnes qui co-construisent le domaine (que nous appelons des acteurs), l'ensemble des mots de la langue générale permettant de créer du sens dans les écrits et les oraux réalisés par ces personnes identifiées, l'ensemble des catégorisations lexicales et des classes sémantiques.

Nous avons vu que la représentation d'un domaine grâce à une approche en organisation des connaissances pouvait être effectuée par des systèmes d'organisation des connaissances spécifiques. Ceux-ci sont très hétérogènes, plus ou moins formalisés et la diversité de leurs usages est croissante (Zacklad 2018).

Nous avons fait l'hypothèse que pour représenter les connaissances d'un domaine nous avions besoin de le décrire, de définir le langage utilisé par ses acteurs, de structurer et désambiguïser ce langage, d'indexer la documentation disponible pour y avoir accès et enfin de trouver des formalismes très expressifs permettant le raisonnement pour enrichir les différentes visions sur le domaine. Nous avons également fait l'hypothèse qu'une approche hybride, ascendante et descendante, telle que préconisée par Bourigault, Aussenac-Gilles, et Charlet 2004, était une façon de représenter le plus exhaustivement possible cette connaissance.

L'utilisation de cette méthodologie a donc été appliquée pour réaliser simultanément trois systèmes prenant en compte les contraintes décrites ci-dessus : une terminologie bilingue français-picard, un thésaurus et une ontologie.

Il s'est avéré que l'analyse de ressources lexicales hétérogènes représentatives du langage des mineurs de fond pouvait être utilisée pour la définition des concepts feuilles d'un thésaurus du patrimoine minier et pour la structuration de ses termes à l'aide des définitions. Nous avons cependant montré que la couverture informationnelle des ressources lexicales (TerminoMines) n'était pas complète par rapport au domaine. Il manquait en effet le vocabulaire de l'après-mine, période que nous avions définie temporellement (1990 jusqu'à aujourd'hui). Un apport d'autres ressources, documentaires et classificatoires plus larges, identifiées grâce à la rencontre des acteurs du domaine permettant l'enrichissement de la structure du thésaurus, jusqu'à sa stabilisation (ThesoMines v1.0) a été nécessaire.

Le thésaurus est considéré comme le système le plus abouti, c'est-à-dire, utilisable en l'état par tous. Il a en effet été déposé en accès libre sur une plateforme ouverte[169] (Opentheso). Sa conception a suivi certaines règles et standards qui permettent de le nommer thésaurus, il ne peut, en revanche, être considéré comme « fini ». Peut-on d'ailleurs parler de la fin de conception d'un SOC ? Pour répondre à cette question, nous avons introduit la notion de saturation, qui est l'état d'un SOC tel que son enrichissement n'apporte plus de nouveaux concepts. Nous avons vu que celle-ci dépendait du niveau de granularité des SOC. Est-ce que les limites de conception d'un SOC sont définies par les utilisateurs ou par les concepteurs ? ThesoMines n'ayant pas encore d'utilisateurs, il est difficile, même s'il a été conçu pour et qu'il est opérationnel, de prouver qu'il correspondra à toutes leurs attentes. Nous savons, de plus, qu'un SOC représentant un domaine en activité se doit d'être mis-à-jour en fonction des évolutions et des nouvelles connaissances. Cette caractéristique soulève un autre problème, celui de sa maintenance et de ses coûts. Un SOC non mis-à-jour est destiné à perdre son utilité première. Il ne sert plus à organiser les documents, les représentations de documents, les œuvres et les concepts (Hjørland 2008). En revanche, il peut acquérir un nouveau statut et devenir un objet d'étude, représentatif d'un mode de pensée, d'une époque, de méthodes de conception, d'une culture. Les SOC et leur approche de conception sont en effet des témoins d'un travail intellectuel, ils sont le résultat de raisonnements collectifs et d'applications théoriques. Toute l'importance de décrire les nouveaux usages et méthodes de conception des SOC entre dans ce cadre.

Nous avons ensuite montré en quoi le formalisme du thésaurus permettait une représentation plus expressive que la terminologie grâce à sa hiérarchie, mais manquait néanmoins d'expressivité au niveau sémantique : les termes-concepts de thésaurus peuvent en effet être reliés entre eux par des relations d'association, des relations hiérarchiques et d'équivalence, mais pas sémantiquement. Un autre système devait alors être construit, un système plus expressif et permettant, si possible de raisonner. Suite à sa première stabilisation, le thésaurus était une base solide pour effectuer ce travail et certaines méthodologies avaient montré l'intérêt et les limites d'une construction d'ontologie à partir de thésaurus (Mothe et Hernandez 2006 ; Hubert, Mothe, et Will 2009 ; Gandon et Passant 2015). Il fallait néanmoins changer de point de vue par rapport à l'organisation des concepts du thésaurus : celle-ci ne suivait pas celle des classes du modèle standard que nous devions utiliser, le CIDOC CRM, modèle destiné à la

[169] https://opentheso.huma-num.fr/opentheso/?idt=th136

documentation scientifique du patrimoine culturel. Le choix de ce modèle avait été fait par le projet MémoMines pour permettre l'échange de données (interopérabilité) entre les nouveaux systèmes créés et ceux en vigueur. La principale difficulté dans l'alignement des termes-concepts de ThesoMines aux classes du CIDOC CRM est due au fait que le système et le modèle possèdent deux orientations conceptuelles propres, nous obligeant parfois à forcer la sémantique en faisant correspondre des termes-concepts à des classes du CIDOC CRM partiellement équivalentes sémantiquement, au risque d'exprimer des liens entre instances qui ne correspondent pas aux définitions des classes et propriétés du modèle. Il est également parfois arrivé de ne pas pouvoir représenter certains termes désignant des objets du patrimoine immatériel, ceci montrant les limites du modèle. Pour des raisons pratiques liées à la description de domaines spécifiques, nous avons vu que le modèle pouvait être étendu, en revanche, l'extension doit suivre impérativement sa logique conceptuelle, ses principes au risque de perdre en interopérabilité. La question de l'interopérabilité des données avec l'utilisation du CIDOC CRM avait déjà été soulevée par Szabados et Letricot, les auteures mentionnant : « L'ontologie est relativement facile à comprendre et à manipuler – en particulier dans la phase de modélisation de l'information durant laquelle elle fournit, pour la réflexion, un éventail de concepts liés au domaine culturel – mais on constate qu'elle permet de modéliser une même information de plusieurs façons différentes. Dans ce cas, même si les données sont bien enrichies sémantiquement, qu'en est-il effectivement de l'interopérabilité de ces données ? » (Szabados et Letricot 2012, 13). L'importance de documenter la méthodologie et la finesse des choix, entre dans ce cadre pour limiter les problèmes d'interopérabilité.

Ensuite, pour valider le modèle ainsi construit (modèle OntoMines), une étape d'instanciation devait être réalisée afin de vérifier que les réalités du domaine portées par le discours d'un ancien mineur pouvaient être représentées. Nous avons montré quelques limites inhérentes au formalisme du CIDOC CRM au regard de la description du patrimoine culturel immatériel : il ne permet par exemple pas d'exprimer certains concepts appartenant à la connaissance sensible tels que les émotions, les odeurs ou des concepts plus spécifiques au domaine : les gaz et fluides, ceux-ci faisant partie du patrimoine minier. Dans ce cadre, l'apport des autres SOC, et notamment du thésaurus a été montré, son enrichissement pouvant pallier des manques au niveau de la sémantique du modèle. Un autre point important est la souplesse du modèle, véritable volonté des concepteurs (Boeuf 2012), permettant de représenter des informations contradictoires. Nous avons illustré celle-ci. Néanmoins, l'interopérabilité syntaxique est primordiale, c'est-à-dire que l'échange entre les données d'un SOC et l'autre doit se faire sans perte de contenu. Les différents moyens employés pour capter toutes les connaissances d'un domaine posent questions au niveau de l'interrogation des données par les utilisateurs. Quelles connaissances pourront-ils interroger et comment ?

L'aspect chronophage mais qualitatif d'une analyse manuelle est un aspect qui a été mentionné dans des travaux de conception de ressources termino-ontologiques à partir de textes comme ceux de Bourigault, Aussenac-Gilles, et Charlet 2004. L'apport, ici, n'est pas seulement la qualité ou l'efficacité des systèmes, mais l'importante réflexion conceptuelle associée au travail de conception, d'indexation ou

d'instanciation manuelle, notamment concernant la granularité de description. Concernant cette granularité, l'instanciation à partir du contenu d'une transcription de type témoignage a montré ses limites : le modèle ontologique du CIDOC CRM n'est pas destiné à instancier toute la sémantique contenue dans un document mais les métadonnées associées (ce qui présuppose en amont une première structuration des données du document). La représentation des connaissances d'un domaine ne peut s'arrêter à l'instanciation d'un modèle ontologique de haut-niveau, d'où la nécessité de compléter les points de vue grâce à d'autres ressources (terminologie, thésaurus mais également d'autres SOC).

Il était aussi question de comprendre en quoi l'instanciation par ontologie (grâce à l'interopérabilité des SOC) permettait de multiplier les points de vue d'analyse sur un même document. Nous souhaitions plus précisément montrer l'apport des SOC dans l'analyse de l'architecture documentaire (Hjørland 2008) et l'impact de cette approche sur le niveau de granularité d'instanciation et d'indexation. Celle-ci montre d'intéressantes perspectives dans l'étude du document, en réduisant le caractère subjectif de l'indexation. En effet, il n'est plus question de passer d'une lecture subjective d'un document à une indexation, elle-même tout à fait subjective, mais de pouvoir montrer les paliers de réflexion intermédiaires permettant d'arriver à une indexation finale. L'instanciation permet le repère des formes en contexte, formes pouvant être reliées à plusieurs niveaux de granularité sur des plans conceptuels différents (concept de thésaurus ou classe de l'ontologie). Comme le mentionne Pincemin, « [s]i l'option la plus courante consiste à étudier le corpus à travers son lexique, on peut tout aussi bien le voir par le biais d'autres descriptions (ex. catégories grammaticales) et d'autres paliers (le morphème ou son approximation par des tri-grams, le tour de parole,…) » (Pincemin 2011, 265). Les ontologies peuvent montrer si le découpage d'une unité documentaire est pertinente ou non. Il était alors intéressant d'étudier la pertinence du découpage du document pour l'indexation. En effet, comme le mentionne Pincemin, « la variation du choix de la partition (découpage du corpus en parties) est un moyen de rendre compte de catégories métatextuelles […] » (Pincemin 2011, 265). Nous avons montré que le découpage en tour de parole pour un entretien, s'il peut être pertinent pour l'indexation, n'y est cependant pas pour l'instanciation, le discours n'étant pas forcément linéaire, pouvant être interrompu par l'interviewer ou par l'environnement. En revanche, l'ontologie permet de rapprocher des concepts issus de plusieurs segments, permettant une autre lecture de l'entretien et dépassant le découpage linéaire. Par extension, les limites (de document, de format, de langue) peuvent être dépassées par l'ontologie.

Nous avons enfin étudié la représentation visuelle du graphe conceptuel réalisé à la suite de l'instanciation du modèle d'OntoMines et avons remarqué comme Pietriga et Lee (2009) le défaut dans la lisibilité des résultats lorsque toutes les données sont affichées dans le graphe. La présentation de ceux-ci doit pouvoir se faire d'une manière dynamique et par niveau de granularité ou par type de concept : affichage seul des descripteurs de thésaurus, affichage seul des instances, affichage seul des classes, des propriétés etc., afin de réellement pouvoir visualiser les différences de granularité d'analyse d'un même document. Néanmoins, nous avons montré que la représentation d'une instanciation par graphe conceptuel pouvait permettre de montrer le rapprochement de concepts, ou au contraire leur

éloignement en fonction des algorithmes utilisés. Pour notre part, l'algorithme de représentation utilisé donnait de la valeur au nombre de liens entrants et sortants d'un même nœud, agrégeant les nœuds les plus représentés sémantiquement au centre du graphe et éloignant les autres, ceci pouvant permettre de prendre une décision dans la phase d'indexation.

Pour conclure, la réalisation simultanée des trois systèmes a permis de vérifier nos hypothèses. En effet, l'enrichissement de l'un implique l'enrichissement de l'autre, ce qui augmente la couverture informationnelle du domaine et les points de vue sur celui-ci. Concernant ce point, nous avons vu que le choix de standards recommandés par le *World Wide Web Consortium* pour la formalisation des SOC, SKOS pour ThesoMines, OWL et RDFS pour OntoMines augmente leur capacité à être interopérables entre eux, avec d'autres ressources et d'autres domaines. Ceci augmente également la couverture du domaine, notamment grâce à l'alignement de concept issus de ThesoMines et d'autres vocabulaires en SKOS ainsi que l'identification de mêmes instances ou de mêmes classes d'un domaine à un autre, dans le cas d'OntoMines, permettant un futur raisonnement. La finesse de description du domaine a été étudiée et le choix de proposer plusieurs niveaux hiérarchiques dans ThesoMines et OntoMines s'est révélé être un bon compromis entre spécificité du domaine et conceptualisation : il y a autant de points de vue différents que de niveaux de spécificité. Enfin, nous avions fait l'hypothèse que la prise en compte des entités nommées tels que les acteurs, les entités spatiales et temporelles permettait de rendre effective la représentation du domaine. Nous avons vu que ces catégorisations pouvaient être décrites de différentes manières selon chaque SOC et qu'elles avaient besoin d'être liées à d'autres catégories telles que les activités ou les objets afin d'être pleinement significatives.

D'autre part, ce travail offre d'intéressantes perspectives techniques, conceptuelles et méthodologiques. Les premières concernent principalement l'enrichissement des SOC pour leur maintenance mais aussi et surtout pour poursuivre l'objectif initial de valorisation de la mémoire minière. Les deuxièmes interrogent l'interopérabilité culturelle et l'interculturalité, l'évaluation des ontologies et l'étude des systèmes d'organisation des connaissances comme outils documentaires de médiation, trois axes permettant de renouveler les approches en organisation des connaissances.

Les perspectives techniques concernent : (1) l'indexation des ressources lexicales (TerminoMines) avec les descripteurs de ThesoMines afin de les mettre en valeur sur le web, (2) l'instanciation d'OntoMines grâce à d'autres entretiens pour valider le modèle et étudier plus en profondeur le lien entre la pertinence de la granularité de découpage d'un document transcrit en tour de parole et sa granularité d'indexation, (3) l'enrichissement des alignements avec d'autres concepts en SKOS provenant de vocabulaires distants pour agrandir la couverture du domaine, (4) l'enrichissement de TerminoMines grâce aux instances d'OntoMines et réalisation des alignements avec le standard TBX afin d'obtenir une complète interopérabilité, (5) la description du domaine avec d'autres modèles ontologiques de haut-niveau (par exemple celui d'Europeana, *Europeana Data Model* ou FRBRoo)[170], (6) l'illustration de la terminologie et le thésaurus à l'aide de photographies ou de dessins afin d'augmenter leur expressivité, (7) une

[170] http://www.cidoc-crm.org/Resources/the-edm-model

réflexion sur la transférabilité de la méthode pour d'autres territoires en transition écologique, comme la Pologne et pour cela enrichir les SOC avec d'autres langues, (8) l'évaluation des SOC par les professionnels de la documentation du patrimoine minier, avec l'indexation d'un véritable fond documentaire sur le patrimoine minier.

Les perspectives conceptuelles et méthodologiques concernent la prise en compte des SOC selon un point de vue communicationnel. Le concept d'interopérabilité est lié à celui d'interculturalité, comme le mentionne Al Sahyouni Bou Fadel, « la communication interculturelle est un ensemble de compétences requises pour une interaction réussie entre des personnes de cultures différentes. Précisément, c'est la capacité de comprendre, d'analyser les différences d'une autre culture, de s'y adapter, d'y évoluer et d'atteindre ces objectifs dans cette différence » (Al Sahyouni Bou Fadel 2016, 2). Ces auteurs étudient « les pratiques et les enjeux de l'interopérabilité culturelle ainsi que la place qu'elle occupe aujourd'hui dans la société d'information marquée par les activités de partage et d'échange interculturel visant l'interopérabilité culturelle » (Al Sahyouni Bou Fadel 2016, 2). Dans cette lignée, il pourrait être intéressant d'étudier l'interopérabilité culturelle des SOC, c'est-à-dire leur capacité à échanger des connaissances entre une même communauté de pratiques possédant des langues et des cultures différentes par exemple, ou bien étudier les relations interculturelles entre des cultures différents mises en contact sur un même territoire afin de montrer l'enrichissement culturel contenues dans les interactions portées par celles-ci.

Une deuxième perspective est d'étudier l'évaluation des ontologies, problématique incontournable, ne possédant pas encore de cadre fédérateur, mais faisant l'objet de nombreuses propositions. L'évaluation consiste à s'assurer « que l'ontologie modélise correctement le domaine du monde réel pour lequel elle a été développée » (Zemmouchi-Ghomari, Deghmani, et Meghnous 2017). Ben Abbès, Zargayouna, et Nazarenko mentionnent trois difficultés : « la complexite des structures ontologiques », « la variabilité de la référence » qui se révèle être « généralement très dépendante du domaine modélisé et de l'utilisation prévue » ou encore les « limites des mesures classiques telles que la précision et le rappel qui reposent sur des jugements de pertinence binaires alors qu'on veut pouvoir rendre compte d'une pertinence graduée, une classe pouvant ne correspondre que partiellement a un concept de la référence » (Ben Abbès, Zargayouna, et Nazarenko 2010, 2). La solution proposée par ces auteurs pour palier ces difficultés est de « décomposer le problème de l'évaluation des ontologies en sous-problèmes suivant la nature des données manipulées » (Ben Abbès, Zargayouna, et Nazarenko 2010, 3). Ils montrent notamment la procédure d'évaluation, qui prend en entrée une liste de classes sémantiques et une ontologie de référence et qui fournit en sortie une mesure de pertinence pour chacune des classes sémantiques et au regard de l'ontologie de référence ». Zemmouchi-Ghomari, Leila, Faiza Deghmani et Aya Meghnous proposent quant à eux la génération d'un questionnaire à partir d'une ontologie pour l'évaluer (Zemmouchi-Ghomari, Deghmani, et Meghnous 2017).

Les approches de validation des ontologies peuvent être classées selon différentes dimensions :

> « *those based on comparing the ontology to a "golden standard" (which may itself be an ontology; e.g. Maedche et Staab, 2002); those based on using the ontology in an application and evaluating the results (e.g. Porzel & Malaka, 2004); those involving comparisons with a source of data (e.g. a collection of documents) about the domain to be covered by the ontology (e.g. Brewster et al., 2004); those where evaluation is done by humans who try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc. (e.g. Lozanotello and Gómez-Pérez, 2004).* » (Brank, Grobelnik, et Mladenić 2009, 1)

Les auteurs déclinent ces dimensions en plusieurs niveaux d'analyse : « *Lexical, vocabulary, or data layer* », « *Hierarchy or taxonomy* », « *Other semantic relations* », « *Context or application level* », « *Syntactic level* » et « *Structure, architecture, design* ». Il sera intéressant de considérer ces différentes dimensions et niveaux d'analyse pour étudier les enjeux conceptuels de l'évaluation des ontologies.

Une dernière perspective est d'étudier les SOC comme outils de médiation. Ils peuvent en effet être étudiés sous l'angle de la médiation culturelle et documentaire et ce, pour plusieurs raisons : ils permettent à différentes communautés de partager le même vocabulaire (dans une même langue ou dans différentes langues), ils représentent les connaissances (concepts, contenu de document, objets) à partir de différents points de vue. Dans le cadre de la médiation culturelle, ils peuvent, de ce fait, faciliter la réalisation d'une exposition, en proposant des contenus adaptés au public. En tant que SOC, la terminologie peut, par exemple, être étudiée dans le cadre de traduction pour des touristes étrangers (Muñoz 2011) et la pratique terminologique « peut être abordée à travers ses réalisations concrètes (index, thesaurus, ontologies, réseaux sémantiques…) comme un dispositif de médiation des savoirs et des connaissances » (Blampain, Thoiron, et Campenhoudt 2006, 161). Le thésaurus peut être étudié comme un outil capable de permettre « *index librarians reference and information service librarians and users of a specialized information system to share the same vocabulary.* » (Dias, Moreira, et Alves 2020). Dans certains projets, comme celui d'HyperThésau, le thésaurus est « construit comme un outil de médiation entre des vocabulaires « locaux » ou « maison » et des vocabulaires documentaires plus généraux. » (Perrin 2020, résumé).

Dans le web de données et dans le web sémantique, les ontologies sont des bases aux systèmes et outils de médiation : « Le web de données est une réponse à ce besoin et devient en soi un outil de médiation au sens premier du terme, c'est-à-dire une façon de transmettre du sens et du contenu. » (O. Rousseau et *al.* 2014). « *Ontologies have been proposed as mediators as they can offer unified access to the heterogeneous resources » (*Hadzic et *al.* 2009, 100). Les ontologies peuvent répondre au problème de l'hétérogénéité des sources de données, problème que peuvent avoir les médiateurs. Les ontologies utilisées pour médier entre différentes bases de données, sont appelées ontologie de médiation.

Les SOC peuvent jouer d'intermédiaires entre les documents, les objets, les concepts et les actions de médiation : les documents indexés par les SOC peuvent être utilisés dans différents types d'actions de médiation, « publications papier, expositions muséales ou virtuelles, mais aussi interventions dans le

cadre de l'enseignement universitaire ou encore dans celui d'événements nationaux comme les Journées européennes du patrimoine ou la Nuit des musées. » (Issenmann 2011). Ils peuvent être étudiés comme outils numériques pour l'accès au patrimoine et sa médiation : **«** La construction historique de « document » montre que ce dernier est le produit d'actions de médiation, lesquelles sont réalisées par des activités comme la sélection, la représentation, le classement, les expositions. Des systèmes et services et d'autres activités constituent des dispositifs documentaires qui fonctionnent comme proposition de parcours des sujets dans le monde de l'information. Ainsi, la notion classique de document et ses dérivations fondent l'opérationnalisation des activités citées. » (Ortega et Saldanha 2017). Enfin, comme le soulignent Guimarães et *al.*, « *[…] a KOS is the product of historical, cultural, and social factors, whose concepts reflect the divisions of scientific/cultural/social labor in knowledge domains* (Albrechtsen et Jacob 1998 ; McTavish, Neal, et Wathen 2011). *From this perspective, KOS would be understood as political tools for a reliable mediation* » (Guimarães et *al.* 2019, 48). Ceci ouvre de réelles perspectives d'évaluation des SOC au regard de ce nouveau point de vue.

BIBLIOGRAPHIE

La norme bibliographique utilisée est issue du logiciel Zotero et correspond à Chicago manual of style 17th edition (author-date, french).

Bibliographie scientifique

Addis, Matthew, Mike Boniface, Simon Goodall, Paul Grimwood, Sanghee Kim, Paul Lewis, Kirk Martinez et Alison Stevenson. 2003. « SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling ». Dans *The Semantic Web - ISWC 2003*, édité par Dieter Fensel, Katia Sycara et John Mylopoulos, 582-596. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer. https://doi.org/10.1007/978-3-540-39718-2_37.

AFNOR. 1981. « NF Z47-100; Règles d'établissement des thésaurus monolingues ».

Agostinelli, Serge. 2012. « Connaissance : pseudo-concept partiellement opératoire ». *Études de communication - langages, information, médiations*, n° 39 (décembre). Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication de l'Université Lille 3 : 65-76. https://doi.org/10.4000/edc.3883.

Al Sahyouni Bou Fadel, Reine. 2016. « L'interopérabilité culturelle et l'interculturalité ». *Communication. Information médias théories pratiques* 34 (1). Editions Nota bene. https://doi.org/10.4000/communication.6636.

Albrechtsen, Hanne et Elin Jacob. 1998. « The Dynamics of Classification Systems as Boundary Objects for Cooperation in the Electronic Library ». *Library Trends* 47 (septembre) : 293-312.

Allen, Charles Geoffry et Jean Aitchison, éd. 1984. *Bibliographie des vocabulaires, thésaurus, vedettes-matières et systèmes de classification mono et multilingues dans le domaine des sciences sociales*. Rapports et documents de sciences sociales 54. Paris : Unesco.

Almeida, Bruno, Nuno Freire et Daniel Monteiro. 2021. « The Development of the ROSSIO Thesaurus: Supporting Content Discovery and Management in a Research Infrastructure ». Dans *Proceedings of the 17th Italian Research Conference on Digital Libraries*, 138-146.

Amar, Muriel. 2000. « Les fondements théoriques de l'indexation: une approche linguistique ». Paris, France : ADBS éditions.

Archambault, Michèle. 2012. « Faculté imaginale du roman et fonction de médiation : vers la construction d'un domaine info-littéraire ». Thèse de doctorat en Sciences de l'information et de la communication, Lyon 2. http://www.theses.fr/2012LYO20086.

Arsenault, Clément, Lyne Da Sylva, Dominic Forest, Yves Marcoux et Dominique Maurel. 2009. « La recherche d'information ». Dans *Introduction aux sciences de l'information*, édité par Jean-Michel Salaün et Clément Arsenault, Presses de l'Université de Montréal PUM, 101-158. Montréal, Canada.

Arsenault, Clément et Alireza Noruzi. 2012. « Les Relations Bibliographiques Entre Une Œuvre et Une Autre Œuvre Selon Le Modèle FRBR : La Perspective Canadienne ». Dans *L'organisation Des Connaissances : Dynamisme et Stabilité*, édité par Widad Mustafa El Hadi, 105-117. Hermès science et Lavoisier. http://eprints.rclis.org/20294/.

Auclaire, Florence. 1985. « L'Elaboration d'un dictionnaire technique trilingue français-allemand-anglais : domaine minier ». Thèse de doctorat en Linguistique, Sorbonne Université. http://www.theses.fr/1985PA040081.

—. 1990. « Thésaurus de l'exploitation minière souterraine et à ciel ouvert ». BRGM. http://infoterre.brgm.fr/rapports/RR-31244-FR.pdf.

Auger, Pierre, Louis-Jean Rousseau, Rosita Harvey et Jean-Claude Corbeil. 1978. *Méthodologie de la recherche terminologique*. Québec, Canada : Office de la langue française.

Auger, Pierre, Louis-Jean Rousseau et Québec. Office de la langue française. 1973. *Lexique anglais-français de l'industrie minière. 1, l'exploitation*. Québec, Canada : Ministère de l'éducation.

Aussenac-Gilles, Nathalie, Mustapha Baziz et Nathalie Hernandez. 2006. « Ontologies pour la recherche d'information : importance de la dimension terminologique ». Dans *Terminologie et accès à l'information spécialisée.*, édité par Widad Mustafa El Hadi, 211-234. Techniques et traités des sciences et techniques de l'information. Hermes science publications : Lavoisier. http://www.irit.fr/publis/IC3/Aussenac-traiteWidad_2006.pdf.

Avilés-F., Alicia, Clara Castro-G., Ceresa-O. Nora, Ivonne Clerc-M., Evelyn Con-LK, María Teresa Cortés-C., Lidia Ellis-M., María Loreto Gatica-C. et Sofía Opitz-V. 1997. « Tesauro de la Minería y Materias Afines ». *REDIMIN Revista Digital Minera*. REDMIN. http://repositorio.conicyt.cl/bitstream/handle/10533/205771/TESAURIO_DE_LA_MINERIA_Y_MATERIAS_AFINES.pdf?sequence=1&isAllowed=y.

Baader, Franz, Diego Calvanese, Deborah McGuinness, Peter Patel-Schneider et Daniele Nardi. 2003. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press.

Babik, Wieslaw, H. Peter Ohly et Karsten Weber. 2017. *Theorie, Semantik Und Organisation von Wissen*. Ergon Verlag.

Bachimont, Bruno. 2000. « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances. » Dans *Ingénierie des connaissances: évolutions récentes et nouveaux défis*, 305-323. https://www.hds.utc.fr/~bachimon/dokuwiki/_media/fr/ontologie-icbook.pdf.

Bachimont, Bruno, Fabien Gandon, Gautier Poupeau, Bernard Vatant, Raphaël Troncy, Stéphane Pouyllau, Ruth Martinez, Michèle Battisti et Manuel Zacklad. 2011. « Enjeux et technologies : des données au sens ». *Documentaliste-Sciences de l'Information* Vol. 48 (4) : 24-41.

Baget, Jean-François, Étienne Canaud, Jérôme Euzenat et Mohand Saïd-Hacid. 2003. « Les langages du web sémantique ». Report. https://hal.inria.fr/hal-00906616.

Beau, Francis. 2012. « L'organisation des connaissances au cœur de la démarche scientifique. Organiser une mémoire pour comprendre et savoir, puis agir et décider avec sagesse ». *Études de communication - langages, information, médiations*, n° 39 (décembre). Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication de l'Université Lille 3 : 77-103. https://doi.org/10.4000/edc.3885.

—. 2015. « L'organisation des connaissances au cœur du système d'information ». *Communication management* Vol. 12 (1). ESKA : 13-30.

Béjoint, Henri et Philippe Thoiron. 2000. *Le sens en terminologie*. Lyon, France : Presses Universitaires Lyon.

Ben Abbès, Sarra, Haifa Zargayouna et Adeline Nazarenko. 2010. « Évaluation de classes sémantiques pour la construction d'ontologies ». Dans *21èmes 21es Journées Francophones d'Ingénierie des Connaissances*, édité par Sylvie DESPRES, 297-308. Nîmes, France : Ecole des Mines d'Alès. https://hal.archives-ouvertes.fr/hal-00487726.

Ben Smida, Kaouther. 2016. « Production of first domain ontology in CIDOC CRM format from heterogeneous metadata ». Mémoire de Master, Tunisia : University of Jendouba.

Benneteu, Brigitte, Sabine Boudou-Ourliac, Myriam Devalette et Tarn. Conservation des musées. 2010. *Les mots de la mine: petit abécédaire pour mineur du XXIe siècle*. Albi, France : Éd. un Autre reg'art.

Bernard, Michel. 1994. *De quoi parle ce livre ? élaboration d'un thésaurus pour l'indexation thématique d'oeuvres littéraires*. Travaux de linguistique quantitative. Paris : Honoré Champion.

Biagetti, Maria Teresa. 2020. « Ontologies (as knowledge organization systems) ». Édité par Birger Hjørland et Claudio Gnoli. *Encyclopedia of Knowledge Organization*. https://www.isko.org/cyclo/ontologies.

Binding, Ceri, Keith May et Douglas Tudhope. 2008. « Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM ». Dans *Research and Advanced Technology for Digital Libraries*, édité par Birte Christensen-Dalsgaard, Donatella Castelli, Bolette Ammitzbøll Jurik et Joan Lippincott, 280-290. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer. https://doi.org/10.1007/978-3-540-87599-4_30.

Blampain, Daniel, Philippe Thoiron et Marc van Campenhoudt, éd. 2006. *Mots, termes et contextes: actes*. Actualité scientifique. Paris (France) : Archives contemporaines.

Bliss, Henry Evelyn. 1929. *The organization of knowledge and the system of the sciences*. New York, Etats-Unis d'Amérique : H. Holt and Company.

—. 1933. *The organization of knowledge in libraries and the subject-approach to books*. New York, Etats-Unis d'Amérique : H.W. Wilson.

Boeuf, Patrick Le. 2012. « De la sémantique des inventaires aux musées en dialogue : la modélisation CIDOC CRM ». https://hal-bnf.archives-ouvertes.fr/hal-00807664.

Bohas, Georges. 2016. *L'illusion de l'arbitraire du signe*. Rennes, France : Presses universitaires de Rennes.

Borgo, Stefano, Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari et Luc Schneider. 2002. « WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE Ontology ». *CSPREA Computer Science Preprint Archive* 2002 (11) : 74-110.

Borst, Willem Nico. 1997. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Enschede, NL, Pays-Bas : Centre for Telematics and Information Technology.

Boudhir, Bechir. 2016. « La normalisation des TICE : facette terminologique et application à des situations multilingues ». Thèse de doctorat en Sciences de l'information et de la communication, Université Paris 8. http://www.theses.fr/2016PA080055.

Boulogne, Arlette et Institut national des techniques de la documentation. 2004. *Vocabulaire de la documentation*. Paris, France : ADBS éditions.

Bourigault, Didier, Nathalie Aussenac-Gilles et Jean Charlet. 2004. « Construction de ressources terminologiques ou ontologiques à partir de textes Un cadre unificateur pour trois études de cas ». *Revue d'intelligence artificielle* 18 (1) : 87-110. https://doi.org/10.3166/ria.18.87-110.

Boutin, Eric, Pei Liu, Stéphane Goria, Philippe Dumas et Amos David. 2008. « Les Sciences de l'Information - Communication vu à travers le thésaurus Rameau : reflet de la réalité ou interprétation du monde ». Dans *Actes du colloque de la SFSIC*, 1-12. https://archivesic.ccsd.cnrs.fr/sic_00827350.

Bovio, Jean. 1906. *Vocabulaire technique des mineurs du Nord et du Pas-de-Calais*. Douai, France : Quantin-Dégez.

Brachman, Ronald J. 1977. « What's in a Concept: Structural Foundations for Semantic Networks ». *International Journal of Man-Machine Studies* 9 (2) : 127-152. https://doi.org/10.1016/S0020-7373(77)80017-5.

Brank, Janez, Marko Grobelnik et Dunja Mladenić. 2009. « A survey of ontology evaluation techniques », janvier.

Broughton, Vanda. 2006a. *Essential Thesaurus Construction*. London : Facet Publishing.

—. 2006b. « The need for a faceted classification as the basis of all methods of information retrieval ». Édité par Andy Dawson et David Brown. *Aslib Proceedings* 58 (1/2). Emerald Group Publishing Limited : 49-72. https://doi.org/10.1108/00012530610648671.

—. 2009. « Facet analysis as the theoretical basis of vocabulary tool construction and subject representation in real and digital environments ». Dans *Content architecture: exploiting and managing diverse resources: proceedings of the first national conference of the United Kingdom chapter of the International Society for Knowedge Organization (ISKO)*. London.

—. 2011. « Facet analysis as a tool for modelling subject domains and terminologies. » Dans *Classification and ontology: formal approaches and access to knowledge: proceedings of the International UDC Seminar*, par Aida Slavic et Edgardo Civallero, 207-228. The Hague, The Netherlands : Würzburg: Ergon Verlag.

—. 2017a. « Faceted Classification as the Basis of All Information Retrieval ». présenté à UDC Seminar 2017 Faceted Classification Today: Theory, Technology and End Users, London. http://seminar.udcc.org/2017/files/VBroughton_UDCSeminar2017_slides.pdf.

—. 2017b. « La classification à facettes comme théorie générale pour l'organisation des connaissances ». *Les Cahiers du numerique* Vol. 13 (1) : 25-48.

Broughton, Vanda, Joacim Hansson, Birger Hjørland et Maria J. López-Huertas. 2005. « Knowledge Organization ». Dans *European curriculum reflections on education in library and information science.*, édité par Leif Lorring et Leif Kajberg, 133-148. Danemark : Royal School of Library and Information Science. https://www.repository.unipr.it/handle/1889/1704.

Bui, Thi Minh Phung. 2003. « La structuration sémantique des contenus des documents audiovisuels selon les points de vue de la production ». Thèse de doctorat en Sciences de l'Information et de la Communication, Université Paris 8. https://tel.archives-ouvertes.fr/tel-00259261.

Buzan, Tony et Barry Buzan. 2006. *The Mind Map Book*. Pearson Education.

Cabré i Castellví, Maria Teresa, Monique C. Cormier, John Humbley et Jean-Claude Préfacier Corbeil. 1998. *La terminologie: théorie, méthode et applications*. Ottawa, Canada, France : Les Presses de l'Université d'Ottawa.

Cabré, Maria Teresa. 2000. « Sur la représentation mentale des concepts: bases pour une tentative de modélisation ». Dans *Le sens en terminologie*, par Henri Béjoint et Philippe Thoiron, 21-39. Lyon, France : Presses Universitaires Lyon.

Candolle, Augustin Pyramus de, Jean-François-Pierre Deterville et Jean-Louis-Étienne-François Martel. 1813. *Théorie élémentaire de la botanique, ou exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux*. A Paris, Chez Déterville, libraire, rue Hautefeuille, n° 8. 1813, France.

Canivet, Isabelle. 2011. *Bien rédiger pour le web : Stratégie de contenu pour améliorer son référencement naturel Ed. 2*. Eyrolles. https://univ-scholarvox-com.ressources-electroniques.univ-lille.fr/book/88805885.

Caron, Philippe, Rodolphe Defiolle et Marie-Hélène Lay. 2019. *L'enjeu des métadonnées dans les corpus textuels*. Rivages linguistiques. Presses universitaires de Rennes. https://hal.archives-ouvertes.fr/hal-02484671.

Carvalho, Sara. 2018. « Proposition d'organisation des connaissances dans le domaine de l'Endométriose : une question de terminologie ». http://www.theses.fr. 21 septembre 2018. http://www.theses.fr/s96951.

Cavalié, Étienne. 2019. *L'indexation matière en transition: de la réforme de Rameau à l'indexation automatique*. Paris, France : Éditions du Cercle de la librairie.

Centre liégeois d'archives et de documentation de l'industrie charbonnière (CLADIC). s. d. « Thésaurus ». Blegny-Mine.

CERCHAR. 1976. « Classification des fiches analytiques ». Groupement de documentation des industries extractives : CERCHAR.

Chalon, Paul-Frédéric. 1909. *Manuel du mineur: recherches de mines et leur exploitation*. Paris, France : Librairie polytechnique Ch. Béranger.

Charlet, Jean, Bruno Bachimont et Raphaël Troncy. 2004. « Ontologies pour le Web sémantique ». *Revue I3, numéro Hors Série « Web sémantique »*, 43-63.

Charlet, Jean, Philippe Laublet et Chantal Reynaud. 2003. « Action spécifique 32 CNRS/STIC sur le Web Sémantique. Rapport final ».

Chaumier, Jacques. 2007. « Les ontologies ». *Documentaliste-Sciences de l'Information* 44 (1) : 81-83. https://doi.org/10.3917/docsi.441.0081.

Chevalier, Aline. 2008. « Pourquoi les sites web sont-ils difficiles à utiliser ? Quelques éléments de réponse ». Dans *Ergonomie des documents électroniques*, 177-203. Presses Universitaires de France. https://www-cairn-info.ressources-electroniques.univ-lille.fr/ergonomie-des-documents-electroniques--9782130559924-page-177.htm.

Chorosz, G. et Société de l'industrie minérale. 1987. *Vocabulaire des termes consacrés à l'étude de l'atmosphère minière: français, anglais, allemand, espagnol*. Saint-Étienne France, France : Société de l'industrie minérale.

Chovaux, Olivier. 2001. « Football minier et immigration. Les limites de l'intégration sportive dans les années trente ». *Staps* no 56 (3) : 9-18.

—. 2006. *Identités et représentations du football nordiste au premier XXe siècle*. Autrement. https://www-cairn-info.ressources-electroniques.univ-lille.fr/le-football-dans-nos-societes--9782746707894-page-24.htm.

Classification Research Group. 1969. *Classification and Information Control: Paper Presenting the Work of the Classification Research Group during 1960-1968.* London : The Association.

Clavier, Viviane. 2014. « L'organisation des connaissances au prisme du langage, du texte et du discours. Un parcours en recherche d'information. » Habilitation à diriger des recherches, Université Grenoble-Alpes. https://hal.archives-ouvertes.fr/tel-01208271.

—. 2019. « Organisation des connaissances, données et algorithmes : place aux visualisations ! » Dans *12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?*, 1-12. Montpellier, France. https://hal.archives-ouvertes.fr/hal-02306622.

Condamines, Anne. 1994. « Terminologie et représentation des connaissances ». *Didaskalia (Paris)*, n° 5 (janvier) : 35-51. https://doi.org/10.4267/2042/23235.

Conférence permanente du bassin minier. 1997. « Synthèse du livre blanc ». https://www.missionbassinminier.org/wp-content/uploads/Synthese_Livre_blanc_CPBM_Acte_1.pdf.

Corbion, Jacques. 2016. « LE SAVOIR ... FER / Glossaire du Haut-Fourneau / 5e édition ». Dictionnaire. LE SAVOIR... FER Glossaire du Haut-Fourneau. 2016. http://savoir.fer.free.fr/.

Cordereix, Pascal. 2016. « Comment indexer les corpus oraux ? » *Histoire Epistémologie Langage* 38 (2). EDP Sciences : 101-113. https://doi.org/10.1051/hel/2016380208.

Courbières, Caroline. 2012. « Virtualité, représentation, signification : approche de la complexité documentaire ». *Études de communication - langages, information, médiations*, n° 39 (décembre). Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication de l'Université Lille 3 : 103-116. https://doi.org/10.4000/edc.4100.

Couzinet, Viviane. 2006. « Les connaissances au regard des sciences de l'information et de la communication : sens et sujets dans l'interdiscipline. » Dans *Semaine de la connaissance*, 1 : 1-6. Université de Nantes.

—. 2011. *Janus ou le langage documentaire : union des deux faces d'un métier*. *Professeur-documentaliste*. Éducagri éditions. https://www-cairn-info.ressources-electroniques.univ-lille.fr/professeur-documentaliste--9782844448408-page-173.htm.

—. 2012. « L'organisation des connaissances au regard des Sciences de l'information et de la communication, une exception française ? » Dans *L'organisation des connaissances : dynamisme et stabilité*, par Widad Mustafa El Hadi, 35-48. Paris : Hermes science publications : Lavoisier.

Dalbin, Sylvie. 2003. « Journée d'étude ADBS. La modélisation : pourquoi l'intégrer dans les systèmes d'information documentaire ? » *Documentaliste-Sciences de l'Information* 40 (3). A.D.B.S. : 226-231.

—. 2007. « Thésaurus et informatique documentaires ». *Documentaliste-Sciences de l'Information* 44 (1) : 42-55. https://doi.org/10.3917/docsi.441.0042.

Dalbin, Sylvie, Emmanuelle Bermès, Antoine Isaac, Romain Wenz, Yann Nicolas, Tayeb Merabti, Anila Angjeli, et *al.* 2011. « Approches documentaires : priorité aux contenus ». *Documentaliste-Sciences de l'Information* Vol. 48 (4) : 42-59.

Daloz, Amélie. 2019. « Méthodologie de validation et d'enrichissement d'une ontologie minière fondée sur le CIDOC CRM ». Dans *Corpus et archives numériques*. MESHS Lille.

Daloz, Amélie et Stéphane Chaudiron. 2019. « Méthodologie de conception d'un thésaurus du domaine minier ». Dans *Actes du 21ème Colloque international sur le document numérique (CIDE 21) : La numérisation info-documentaire*, édité par Jacquemin B. et Ghenima M. (dir.), 11-23. Djerba, Tunisia : Europia. https://hal.archives-ouvertes.fr/hal-02568840.

Dauby, Jean. 1979. *Le livre du « rouchi », parler picard de Valenciennes*. Amiens, France : Musée de Picardie.

—. 1983. *Complément au Livre du rouchi: parler picard de Valenciennes*. Amiens, France : Musée de Picardie.

Davallon, Jean et Yves Jeanneret. 2004. « La fausse évidence du lien hypertexte ». *Communication & Langages* 140 (1) : 43-54. https://doi.org/10.3406/colan.2004.3266.

Delengaigne, Xavier et Luis Garcia. 2018. *Organisez votre temps avec le Mind Mapping : Sortez la tête du guidon ! Ed. 2*. Dunod. https://univ-scholarvox-com.ressources-electroniques.univ-lille.fr/book/88859286.

Dépelteau, François. 2010. *La démarche d'une recherche en sciences humaines: De la question de départ à la communication des résultats*. De Boeck Supérieur.

Desfriches Doria, Orélie. 2012. « Contribution de la classification à facettes pour l'organisation des connaissances dans les organisations ». *Études de communication - langages, information, médiations*, n° 39 (décembre) : 173-200. https://doi.org/10.4000/edc.3889.

Dias, Danilo Camargo, Walter Moreira et Rachel Cristina Vesu Alves. 2020. « A representação temática de imagens digitais da NASA no Flickr: as contribuições dos sistemas de organização do conhecimento ». *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação* 18 (mai) : e020011-e020011. https://doi.org/10.20396/rdbci.v18i0.8658560.

Díaz-Corona, Dayany, Javier Lacasta, Miguel Ángel Latre, F. Javier Zarazaga-Soria et Javier Nogueras-Iso. 2019. « Profiling of Knowledge Organisation Systems for the Annotation of Linked Data Cultural Resources ». *Information Systems* 84 (septembre) : 17-28. https://doi.org/10.1016/j.is.2019.04.008.

Diemert, Benjamin. 2012. « Description sémantique de documents audiovisuels structurés ». Thèse de doctorat en Technologies de l'information et des systèmes, Université de Technologie de Compiègne. https://www.theses.fr/2012COMP2064.

Direction Départementale Des Territoires Et De La Mer Du Nord. 2017. « Plan de Prévention des Risques Miniers (PPRM) Denain – Haveluy – Lourches. Pièce n°1 : Note de présentation. » http://www.hauts-de-france.developpement-durable.gouv.fr/IMG/pdf/1-pprm_note_de_presentation_denaisis_vap_18122017.pdf.

Djambian, Caroline. 2010. « Valorisation d'un patrimoine documentaire industriel et évolution vers un système de gestion des connaissances orienté métiers ». Thèse de doctorat en Sciences de l'Information et de la Communication, Université Jean Moulin Lyon 3. https://tel.archives-ouvertes.fr/tel-00483442.

Djioua, Brahim. 1999. « DISCC : Un outil de construction et d'utilisation d'une Base de Connaissances Sémantico-Cognitives des verbes ». Dans *TALN 1999*. Cargèse. http://lalic.paris-sorbonne.fr/PUBLICATIONS/1998-1999/Djioua/Taln99.pdf.

Djurić, Dragan, Dragan Gašević et Vladan Devedžić. 2005. « Ontology Modeling and MDA ». *Journal of Object Technology* 4 (1) : 109-128. https://doi.org/10.5381/jot.2005.4.1.a3.

Doerr, Martin. 2003. « The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata ». *AI Magazine* 24 (septembre) : 75-92.

—. 2009. « Ontologies for Cultural Heritage ». Dans *Handbook on Ontologies*, édité par Steffen Staab et Rudi Studer, 463-486. International Handbooks on Information Systems. Berlin, Heidelberg : Springer. https://doi.org/10.1007/978-3-540-92673-3_21.

Doria, Orélie Desfriches. 2013. « La classification à facettes pour la gestion des connaissances métier : méthodologie d'élaboration de FolkClassifications à facettes ». Thèse de doctorat en Sciences de l'Information et de la Communication, Conservatoire national des arts et metiers - CNAM. https://tel.archives-ouvertes.fr/tel-01015606.

Drame, Khadim. 2014. « Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical ». Thèse de doctorat en Informatique et Santé, Université de Bordeaux. https://tel.archives-ouvertes.fr/tel-01166042.

Du Château, Stefan. 2010. « Simplicius, système d'aide au management des connaissances pour le patrimoine culturel ». Thèse de doctorat en Informatique, Université Jean Moulin Lyon 3. http://www.theses.fr/2010LYO31085.

Dubois, Guy. 1981. *2000 mots du patois de chez nous*. France.

—. 1988. *Les Dits du coeur. pour n'point perde el' Nord, ni l' Pas-d'Calais 2 2*. Lillers] (BP 4, 62190) : G. Dubois.

—. 1993. *Ces bonnes histoires qu'on appelait des cafougnettes Folklore des mines du nord et du pas de calais Jules Mousseron*. Presses de la SCIE Bully Les Mines.

—. 2019. *Parlez Chti !: dictionnaire français-chti*. CLERMONT FERRAND, France : CENTRE FRANCE LIVRES.

Dubois, Guy et Jean-Marie Minot. 1991. *Histoire des Mines du Nord et du Pas-de-Calais: (des origines à 1939-45). Tome 1*. Bruxelles : Scripto Publishing.

—. 1992. *Histoire des Mines du Nord et du Pas-de-Calais (de 1946 à 1992): Album-souvenir, Tome 2*. Bruxelles : Scripto Publishing.

Dubuc, André Éditeur scientifique Préfacier. 2010. *Héros ou martyrs: identités, conditions et statuts des mineurs du Nord-Pas-de-Calais*. Lewarde, France : Centre historique minier.

Dufort, Diane. 2016. « Outils sémantiques d'aide à la conception de jeux pervasifs pour la médiation culturelle ». Thèse de doctorat en Sciences de l'information et de la communication, Université de Franche-Comté. http://www.theses.fr/2016BESA1020.

Eco, Umberto et Jean-Marie Klinkenberg. 1988. *Le signe: histoire et analyse d'un concept*. Bruxelles, Belgique : Éditions Labor.

Ertzscheid, Olivier et Gabriel Gallezot. 2006. « Etude exploratoire des pratiques d'indexation sociale comme une renégociation des espaces documentaires. Vers un nouveau big bang documentaire ? » Dans *Document numérique et société*, édité par Ghislaine Chartron et Evelyne Broudoux. Sciences et techniques de l'information. ADBS Éditions.

Fadili, Hammou et Malika Chakiri. 2014. « Conception et peuplement d'une ontologie modélisant la notion de contexte enrichie par les fonctions lexicales pour la détection du sens dans le texte Parler du Maroc central ». *ⴰⵙⵉⵏⴰⴳ-Asinag*, n° 9 : 147-170.

Fagnoni, Édith. 2015. « L'industrie versus patrimonialisation : questionnement autour d'une transition vers une économie culturelle ». *Bulletin de l'association de géographes français. Géographies* 92 (92-4) : 556-564. https://doi.org/10.4000/bagf.1110.

Favier, Laurence. 2017. « Les humanités numériques et l'évolution des infrastructures de recherche. Quels enjeux pour l'organisation des connaissances ? » Dans *Systèmes d'organisation des connaissances et humanités numériques: Actes du 10ème colloque ISKO France 2015*, édité par Emmanuelle Chevry Pébayle. Strasbourg : ISTE Group.

Favier, Laurence et Widad Mustafa El Hadi. 2013. *L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir ?* https://hal.univ-lille.fr/hal-01312645.

Felicetti, Achille et Francesca Murano. 2021. « Ce Qui Est Ecrit et Ce Qui Est Parle. CRMtex for Modelling Textual Entities on the Semantic Web ». *Semantic Web* 12 (2). Amsterdam : Ios Press : 169-180. https://doi.org/10.3233/SW-200418.

Feyler, François. 2007. « De la compatibilité à l'interopérabilité en matière de repérage d'information pertinente ». *Documentaliste-Sciences de l'Information* 44 (1) : 84-92. https://doi.org/10.3917/docsi.441.0084.

Fondin, Hubert. 2005. « La Science de l'information ou le poids de l'histoire ». *Les Enjeux de l'information et de la communication* Varia 2005 (06/1). GRESEC : 35-54.

Francis, Élie et Odile Quesnel. 2007. « Indexation collaborative et folksonomies ». *Documentaliste-Sciences de l'Information* 44 (1) : 58-63. https://doi.org/10.3917/docsi.441.0058.

Gala, Alexander Pinto-De la, Yudith Cardinale, Irvin Dongo et Regina Ticona-Herrera. 2021. « Towards an ontology for urban tourism ». Dans *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 1887–1890. SAC '21. New York, NY, USA : Association for Computing Machinery. https://doi.org/10.1145/3412841.3442142.

Gandon, Fabien, Catherine Faron-Zucker et Olivier Corby. 2012. *Le Web sémantique: comment lier les données et les schémas sur le web ?* Paris, France : Dunod.

Gandon, Fabien et Alexandre Passant. 2015. « Folksonomies, Thésaurus et Ontologies : trois artefacts combinés dans la structuration des données du Web », juin. Presses universitaires de Rennes. https://doi.org/10.4000/books.pur.88593.

Gangemi, Aldo, Nicola Guarino, Claudio Masolo et Alessandro Oltramari. 2003. « Sweetening WORDNET with DOLCE ». *AI Magazine* 24 (3) : 13-13. https://doi.org/10.1609/aimag.v24i3.1715.

Gasnier, Marina. 2011. *Patrimoine industriel et technique: perspectives et retour sur 30 ans de politiques publiques au service des territoires*. Édité par technologiques et scientifiques Laboratoire de recherche sur les choix industriels et France. Ministère de la culture et de la communication. Direction générale des patrimoines. Lyon, France : Lieux dits Editions.

—. 2019. « Réflexion épistémologique sur le patrimoine industriel : de la pluridisciplinarité à l'interdisciplinarité ». *Revue d'histoire des sciences* Tome 72 (2). Armand Colin : 309-347.

Gaudin, François. 2005. « La socioterminologie ». *Langages*, n° 157 : 81-93. https://doi.org/10.3917/lang.157.0081.

Gauthier, Benoît. 2010. *Recherche sociale. de la problématique à la collecte des données*. Québec, Canada : Presses de l'université de Québec.

Gheorghita, Inga. 2011. « Ressources lexicales au service de recherche et d'indexation des images ». Dans *Actes des 13e Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*. Montpellier, France : Association pour le Traitement Automatique des Langues.

—. 2014. « Construction automatique de hiérarchies sémantiques à partir du Trésor de la Langue Française informatisé (TLFi) : application à l'indexation et la recherche d'images ». Thèse de doctorat en Linguistique Informatique, Université de Lorraine. https://hal.univ-lorraine.fr/tel-01751588.

Ghio, Alain. 1997. « Achile : un dispositif de décodage acoustico-phonétique et d'identification lexicale indépendant du locuteur à partir de modules mixtes ». Thèse de doctorat en Sciences et Technologies de l'Information, Université de droit, d'économie et des sciences d'Aix - Marseille III. https://hal.archives-ouvertes.fr/tel-01663493.

Gicquel, Florence et Lafarge. 2007. « Un glossaire peut cacher une ontologie... » *Documentaliste-Sciences de l'Information* 44 (1) : 64-64. https://doi.org/10.3917/docsi.441.0064.

Gnoli, Claudio. 2020. *Introduction to Knowledge Organization*. Facet Publishing.

Gómez, A., A. Moreno, J. Pazos et A. Sierra-Alonso. 2000. « Knowledge Maps: An Essential Technique for Conceptualisation ». *Data & Knowledge Engineering* 33 (2) : 169-190. https://doi.org/10.1016/S0169-023X(99)00050-6.

Granger, Gilles-Gaston. 1960. *Pensée formelle et sciences de l'homme*. Paris, France : Aubier-Montaigne.

—. 1979. *Langages et épistémologie*. Paris, France : Klincksieck.

Grawitz, Madeleine. 2000. *Méthodes des sciences sociales*. Paris, France : Dalloz.

Groupe de Normalisation documentaire ICOM/CIDOC, Nick Crofts, Martin Doerr, Stephen Stead, Tony Gill et Matthew Stiff. 1999. « Définition du Modèle Conceptuel de Référence du CIDOC (CRM) ». Traduit par Patrick Le Boeuf. http://old.cidoc-crm.org/.

Groupe de travail « Terminologie et documentation ». 2014. « Recommandations relatives à la terminologie. 3ème édition. CST – Conférence des Services de traduction des États européens. » Dans *Recommandations relatives à la terminologie. 3ème édition. CST – Conférence des Services de traduction des Etats européens.*, 1-121. https://www.bk.admin.ch/bk/fr/home/documentation/langues/publications-en-terminologie.html.

Groussier, Marie-Line et Claude Rivière. 1996. *Les mots de la linguistique: lexique de linguistique énonciative*. Gap, France.

Gruber, Thomas R. 1993. « A Translation Approach to Portable Ontology Specifications ». *Knowledge Acquisition* 5 (2) : 199-220. https://doi.org/10.1006/knac.1993.1008.

—. 1995. « Toward Principles for the Design of Ontologies Used for Knowledge Sharing? » *International Journal of Human-Computer Studies* 43 (5) : 907-928. https://doi.org/10.1006/ijhc.1995.1081.

Guarino, Nicola, Daniel Oberle et Steffen Staab. 2009. « What Is an Ontology? » Dans *Handbook on Ontologies*, édité par Steffen Staab et Rudi Studer, 1-17. International Handbooks on Information Systems. Berlin, Heidelberg : Springer. https://doi.org/10.1007/978-3-540-92673-3_0.

Guimarães, José Augusto Chaves, D. Grant Campbell, Suellen Oliveira Milani et Helber Holland. 2019. « Cultural Biases in Knowledge Organization Systems: A Discussion Regarding International University Rankings ». *NASKO* 7 (1) : 48-62. https://doi.org/10.7152/nasko.v7i1.15621.

Guiraud, Pierre. 1978. *Patois et dialectes français*. Paris, France : Presses Universitaires de France.

Hadi, Widad Mustafa El, éd. 2012. *L'organisation des connaissances: dynamisme et stabilité*. Traité des sciences et techniques de l'information. Paris : Hermes science publications : Lavoisier.

Hadzic, Maja, Janusz Kacprzyk, Pornpit Wongthongtham, Tharam Dillon et Elizabeth Chang. 2009. *Ontology-Based Multi-Agent Systems*. Berlin, Heidelberg, Allemagne : Springer Berlin Heidelberg.

Hammache, Arezki et Rachid Ahmed-Ouamer. 2006. « Système d'inférence pour une indexation de documents basée sur une ontologie de domaine ». Dans *Actes du XXIVeme Congres INFORSID 2006*, 895-910.

Hernandez, Nathalie. 2005. « Ontologies de domaine pour la modélisation du contexte en Recherche d'information ». Thèse de doctorat en Informatique, Université Paul Sabatier - Toulouse III. https://tel.archives-ouvertes.fr/tel-00371323.

Hirst, Graeme. 2009. « Ontology and the Lexicon ». Dans *Handbook on Ontologies*, édité par Steffen Staab et Rudi Studer, 269-292. International Handbooks on Information Systems. Berlin, Heidelberg : Springer. https://doi.org/10.1007/978-3-540-92673-3_12.

Hjørland, Birger. 2008. « What is Knowledge Organization (KO)? » *Knowledge Organization* 35 (juillet). https://doi.org/10.5771/0943-7444-2008-2-3-86.

—. 2017. « Reviews of Concepts in Knowledge Organization ». *Knowledge Organization* 44 (6) : 436-464. https://doi.org/doi.org/10.5771/0943-7444-2017-6.

Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Cambridge, Etats-Unis : Digital Library Federation, Council on Library and Information Resources, 1755 Massachusetts Ave. https://eric.ed.gov/?id=ED440657.

Holzem, Maryvonne. 1999. « Terminologie et documentation: pour une meilleure circulation des savoirs ». Paris, France : ADBS.

Hubain, Raphaël, Seth Van Hooland et Ruben Verborgh. 2016. « Classification automatisée : rêve ou réalité ? Analyse critique de l'usage du text mining pour la conception de vocabulaires contrôlés ». *I2D - Information, donnees documents* 53 (2) : 70-79.

Hubert, Gilles, Josiane Mothe et Pierre Etienne Will. 2009. « Produire une ontologie à partir d'un thésaurus – méthode et outil ». Dans *Actes du séminaire VSST 2009 [En ligne]*, 1-14. Nancy, France. https://atlas.irit.fr/PIE/VSST/VSST-2009.html.

Hudon, Michèle. 2006. « Le thésaurus, au carrefour des sciences de l'information et de la terminologie ». Dans *Terminologie et accès à l'information spécialisée*, par Widad Mustafa El Hadi, 71-98. Paris, France : Lavoisier : Hermes science publ.

—. 2013. *Analyse et représentation documentaires: Introduction à l'indexation, à la classification et à la condensation des documents*. PUQ.

—. 2018. « Du bon usage des facettes : un linguiste revisite la théorie de Ranganathan ». Dans *Fondements épistémologiques et théoriques de la science de l'information-documentation: actes du 11e colloque ISKO France 2017, 11 et 12 juillet 2017, Siège de l'Unesco, Paris*, par Widad Mustafa El Hadi et Colloque ISKO France, 70-83. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord : Iste éditions.

—. 2020. « Facet ». *Knowledge Organization* 47 (4) : 320-333. https://doi.org/10.5771/0943-7444-2020-4-320.

Hudon, Michèle et Widad Mustafa El Hadi. 2010. « Organisation des connaissances et des ressources documentaires ». *Les Cahiers du numerique* Vol. 6 (3) : 9-38.

—. 2017. « Introduction. La classification à facettes revisitée. De la théorie à la pratique ». *Les Cahiers du numerique* Vol. 13 (1) : 9-24.

Hyvönen, Eero. 2012. *Publishing and using cultural heritage linked data on the Semantic Web*. San Rafael, Etats-Unis d'Amérique : Morgan & Claypool.

INICHAR. 1965. « Documentation par fiches. » Ixelles-Bruxelles : Annales des mines de Belgique.

Issenmann, Delphine. 2011. « Collections universitaires et patrimoine culturel - L'inventaire du patrimoine de l'Université de Strasbourg ». *In Situ. Revue des patrimoines*, n° 17 (novembre). Ministère de la culture et de la communication, direction générale des patrimoines. https://doi.org/10.4000/insitu.9988.

Jackendoff, Ray. 1983. *Semantics and Cognition*. MIT Press.

Jacob, Christian. 2014. *Qu'est-ce qu'un lieu de savoir ?* Encyclopédie numérique. Marseille : OpenEdition Press. http://books.openedition.org/oep/423.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann et Mathieu Bastian. 2014. « ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software ». *PLOS ONE* 9 (6). Public Library of Science : e98679. https://doi.org/10.1371/journal.pone.0098679.

Kassel, Gilles. 1999. « PHYSICIAN is a role played by an object, whereas SIGN is a role played by a concept ». Dans *Proceedings of the IJCAI'99 Workshop on Ontologies and Problem-Solving Methods : Lessons Learned and Future Trends*, 61-69.

Kergosien, Eric, Marie-Noëlle Bessagnet, Maguelonne Teisseire, Joachim Schöpfel, Amin Farvardin, Stéphane Chaudiron, Bernard Jacquemin, et *al.* 2017. « Méthodologie pour identifier les terrains d'étude dans des corpus scientifiques ». *Document numerique* 20 (2). Lavoisier : 11-30.

Kergosien, Eric, Amélie Daloz, Stéphane Chaudiron et Bernard Jacquemin. 2019. « Le CIDOC-CRM, un modèle ontologique pour représenter les connaissances du bassin minier des Hauts-de-France. Premiers résultats. » Dans *Actes du 6èmeHyperheritage International Symposium (HIS 6), La fabrique du patrimoine à l'ère numérique*. Institut d'études avancées de Paris : Europia Productions.

Kergosien, Eric, Bernard Jacquemin, Marta Severo et Stéphane Chaudiron. 2015a. « The TECTONIQ project for valorization of digital textile industrial heritage in North of France ». Dans *Industrial Heritage in the Twenty-First Century, New Challenges. Proceedings of the XVIth International TICCIH Congress 2015*, 15 p. Lille, France : The International Committee for the Conservation of the Industrial Heritage (TICCIH). https://hal.archives-ouvertes.fr/hal-01358478.

—. 2015b. « Vers l'interopérabilité des données hétérogènes liées au patrimoine industriel textile », novembre. https://hal.univ-lille3.fr/hal-01281716.

Kolonja, Ljiljana, Ranka Stanković, Ivan Obradović, Olivera Kitanović et Aleksandar Cvjetić. 2016. « Development of terminological resources for expert knowledge: a case study in mining ». *Knowledge Management Research & Practice* 14 (4). Taylor & Francis : 445-456. https://doi.org/10.1057/kmrp.2015.10.

La Barre, Kathryn. 2010. « A Semantic (Faceted) Web? » *Les Cahiers du numerique* 6 (3) : 103-131.

Lacour, Philippe. 2004. « L'oubli de la sémantique dans le programme cognitiviste : réflexions sur l'oeuvre de François Rastier ». *Texto !* http://www.revue-texto.net/Inedits/Lacour_LOubli.html.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. 2 vol. Stanford (Calif.), Etats-Unis d'Amérique : Stanford university press.

Larouzière, Béatrice de Bégon de. 2013. « Les langages documentaires au service de la valorisation des collections muséales : quels enjeux ? Le cas du musée Rodin ». Other, Institut national des techniques de la documentation du CNAM. https://memsic.ccsd.cnrs.fr/mem_00944577.

Lateur, Marius et Mario Roques. 1951. *Lexique du parler populaire et ouvrier des régions minières d'Artois: près de 6000 mots, avec leurs prononciation et traduction en français, exemples, dictons, verbes*. Paris, France : Librairie Ricour et Chevillet.

Laubé, Sylvain. 2016. « Patrimoine(s), artefact(s), modélisation des connaissances et humanités numériques ». Dans *Séminaire de recherche « Histoire technique et humanités numériques »*. https://hal.archives-ouvertes.fr/hal-01624824.

Le Bœuf, Patrick. 2013. « Le modèle conceptuel de référence du CIDOC : de la sémantique des inventaires aux musées en dialogue ». *Culture & Musées* 22 (1). Persée - Portail des revues scientifiques en SHS : 89-111. https://doi.org/10.3406/pumus.2013.1753.

Le Coadic, Yves-François. 1984. « La science de l'information : aspects structurels et institutionnels d'une nouvelle interdiscipline ». *Bulletin des bibliothèques de France (BBF)*, n° 2 : 168-172.

Lebon, André. 1976. *Les mots de la mine: lexique dialectal de la profession de mineur au début du 20ème siècle dans la région de Denain*. France.

L'Homme, Marie-Claude. 2002. « Fonctions lexicales pour représenter les relations sémantiques entre termes ». *Traitement automatique de la langue* 43 (janvier) : 19-41.

—. 2004. *La terminologie: principes et techniques*. Paramètres. Montréal : Les Presses de l'université de Montréal.

—. 2005. « Sur la notion de « terme » ». *Meta : journal des traducteurs / Meta: Translators' Journal* 50 (4). Les Presses de l'Université de Montréal : 1112-1132. https://doi-org.ressources-electroniques.univ-lille.fr/10.7202/012064ar.

Machado, Luís Miguel Oliveira. 2021. « Ontologies in Knowledge Organization ». *Encyclopedia* 1 (1). Multidisciplinary Digital Publishing Institute : 144-151. https://doi.org/10.3390/encyclopedia1010015.

Malaisé, Véronique. 2005. « Méthodologie linguistique et terminologique pour la structuration d'ontologies différentielles à partir de corpus textuels ». Thèse de doctorat en Linguistique, Université Paris-Diderot - Paris VII. https://tel.archives-ouvertes.fr/tel-00162575.

Maniez, Jacques. 1999. « Des classifications aux thesaurus: du bon usage des facettes ». *Documentaliste – Sciences de l'information* 36 (4-5) : 249-262.

—. 2002. *Actualité des langages documentaires: les fondements théoriques de la recherche d'information*. Paris, France : ADBS Éditions.

—. 2007. « Rupture ou continuité ? » *Documentaliste-Sciences de l'Information* 44 (1) : 12-16. https://doi.org/10.3917/docsi.441.0012.

Mathieu-Colas, Michel. 2007. « Domaines et classes sémantiques ». *Verbum (Presses Universitaires de Nancy)* 29 (1-2) : 11.

McTavish, Jill R., Diane Rasmussen Neal et C. Nadine Wathen. 2011. « Is What You See What You Get? Medical Subject Headings and Their Organizing Work in the Violence Against Women Research Literature ». *KO KNOWLEDGE ORGANIZATION* 38 (5). Nomos Verlagsgesellschaft mbH & Co. KG : 381-397. https://doi.org/10.5771/0943-7444-2011-5-381.

Mehta, Dinesh P et Sartaj Sahni, éd. 2005. *Handbook of data structures and applications*. CHAPMAN & HALL/CRC.

Menon, Bruno. 2007. « Les langages documentaires ». *Documentaliste-Sciences de l'Information* 44 (1) : 18-28. https://doi.org/10.3917/docsi.441.0018.

Messaoudi, Tommy. 2017. « Proposition d'une ontologie de domaine dédiée à l'annotation d'images spatialisées pour le suivi de la conservation du patrimoine culturel bâti ». Thèse de doctorat en Conception, Ecole nationale supérieure d'arts et métiers - ENSAM. https://pastel.archives-ouvertes.fr/tel-01628391.

Mission Bassin minier. 2015. « Bassin Minier du Nord-Pas de Calais. Patrimoine mondial de l'UNESCO ». https://www.bassinminier-patrimoinemondial.org/wp-content/uploads/2015/03/Synth%C3%A8se-du-dossier-dinscription.pdf.

Mongaudon, Jean-Pierre. 1988. *A la lueur de nos lampes*. Noisy le Grand, France : J.-P. Mongaudon.

Morel, Juliette. 2018. « L'Atlas Historique du Limousin. Questions et méthodes vers un système de données spatio-temporelles contributif ». Dans *Sagéo 2018*. Montpellier, France. https://hal.archives-ouvertes.fr/hal-02096627.

Moseley, Christopher, éd. 2010. *Atlas of the World's Languages in Danger*. 3ème. Paris : UNESCO Publishing. http://www.unesco.org/culture/en/endangeredlanguages/atlas.

Mothe, Josiane et Nathalie Hernandez. 2006. « TtoO: une méthodologie de construction d'ontologie de domaine à partir d'un thésaurus et d'un corpus de référence ». Rapport interne IRIT. Toulouse : Institut de Recherche en Informatique de Toulouse (IRIT).

Mousseron, Jules. 1994. *Oeuvres complètes*. 4 vol. Denain, France : G. Cattiaux.

Mousseron, Jules et Lucien Jonas. 1933. *Mes dernières berlines: Moeurs et coutumes du Pays Minier avec les nouvelles aventures de Cafougnette : poésies patoises*. Lille; Paris : J. Tallandier.

Moutat, Audrey. 2019. « Le commentaire œnologique ou comment déguster le vin par les mots ». *Lexis. Journal in English Lexicology*, nº 13 (avril). Université Jean Moulin Lyon 3 : 1-19. https://doi.org/10.4000/lexis.3026.

Müller, Bertrand. 2008. « Archives, documents, données : problèmes et définitions ». *Gazette des archives* 212 (4) : 35-44. https://doi.org/10.3406/gazar.2008.4509.

Muñoz, Isabel Durán. 2011. « Tourist Translations as a Mediation Tool: Misunderstandings and Difficulties. » *Cadernos de Traduçao* 1 (27). Universidade Federal de Santa Catarina (UFSC) : 29-49.

Murguia, Eduardo Ismael et Rodrigo De Sales. 2013. « PRÁTICAS DISCURSIVAS NA ORGANIZAÇÃO DO CONHECIMENTO Otlet, Bliss e ISKO ». Dans *Informação e/ou Conhecimento: as duas faces de Jano*, 446-461. Universidade do Porto.

Navigli, Roberto et Paola Velardi. 2008. « From Glossaries to Ontologies: Extracting Semantic Structure from Textual Definitions ». *Ontology Learning and Population* 167 (juin) : 71-87.

Neches, Robert, Richard E. Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator et William R. Swartout. 1991. « Enabling Technology for Knowledge Sharing ». *AI Magazine* 12 (3) : 36-36. https://doi.org/10.1609/aimag.v12i3.902.

Obradović, Ivan, Ranka Stanković, Olivera Kitanović et Ljiljana Kolonja. 2012. « Towards a Mining Equipment Ontology ». Dans *Proceedings of the 12th International Conference « Research and Development in Mechanical Industry » (RaDMI-2012)*, édité par Predrag Dašić. Vol. 2. Vrnjačka Banja, Serbia : SaTCIP Publisher Ltd.

Ohly, H. Peter. 2016. « The Future of Knowledge Organization and Information Organization ». *PTINT Praktyka i Teoria Informacji Naukowej i Technicznej* 24 (2-3) : 63-70.

Ortega, Cristina et Gustavo Saldanha. 2017. « La notion de document d'Otlet à Meyriat et les propositions néodocumentalistes ». *Sciences de la société*, nº 100 (janvier). Presses Universitaires du Mirail. https://doi.org/10.4000/sds.5945.

Otlet, Paul. 1903. « The science of bibliography and documentation. » Dans *International organization and dissemination of knowledge: selected essays of Paul Otlet.*, par W. Boyd Rayward, 71-86. Amsterdam : Elsevier.

—. 1934. *Traité de documentation: le livre sur le livre, théorie et pratique*. Bruxelles, Belgique : Editiones Mundaneum.

Pamart, Anthony. 2013. « Vers une sauvegarde numérique du patrimoine industriel et minier, et mineur ». Plateforme d'études et de partage. *DNArchi* (blog). 2013. http://dnarchi.fr/analyses/vers-une-sauvegarde-numerique-du-patrimoine-industriel-minier-et-mineur/.

Paquette, Gilbert. 2002. *Modélisation des connaissances et des compétences : Un langage graphique pour concevoir et apprendre*. Presses de l'Université du Québec. https://univ-scholarvox-com.ressources-electroniques.univ-lille.fr/book/88801305.

Paquette, Gilbert, Jacqueline Bourdeau, France Henri, Josianne Basque, Michel Leonard et Marcelo Maina Cirta. 2003. « Construction d'une base de connaissances et d'une banque de ressources pour le domaine du téléapprentissage ». *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation* 10 (1). Persée - Portail des revues scientifiques en SHS : 29-56. https://doi.org/10.3406/stice.2003.856.

Péchoin, Daniel. 2001. « Thésaurus documentaire, thésaurus idéologique et dictionnaire analogique ». Dans *Repérer, formaliser, traduire les concepts philosophiques: colloque de Besançon, 23-24 juin 1999*, édité par Jean-Pierre Cotten, Benoît Hufschmitt, Centre de documentation et bibliographie philosophiques et Université de Franche-Comté. Département de philosophie, 87-118. Besançon, France : Presses universitaires franc-comtoises.

Perrin, Emmanuelle. 2020. « Les données de l'archéologie au prisme des humanités numériques : thésaurus et interopérabilité sémantique ». https://hal.archives-ouvertes.fr/hal-02865148.

Picard, Anne-Claire Le. 2014. « Le cycle de vie d'une ontologie : évaluation de l'ontologie du domaine de la Toxicologie Nucléaire ». Other, Conservatoire national des arts et métiers - CNAM ; Institut National des Techniques de la Documentation - INTD. https://memsic.ccsd.cnrs.fr/mem_01128938.

Pietriga, Emmanuel et Ryan Lee. 2009. « Langages et outils pour la visualisation et la manipulation de données du Web sémantique ». *Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques* 28 (2) : 173-197. https://doi.org/10.3166/tsi.28.173-197.

Pincemin, Bénédicte. 2011. « Sémantique interprétative et textométrie– Version abrégée ». *Corpus*, n° 10 (novembre) : 259-269. https://doi.org/10.4000/corpus.2121.

Polity, Yolla, Gérard Henneron et Rosalba Palermiti, éd. 2005. *L'organisation des connaissances: approches conceptuelles, Actes du 4è congrès ISKO-France, 3 et 4 juillet 2003*. Paris, Budapest, Turin : Harmattan.

Prévot, Laurent, Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci et Alessandro Oltramari. 2010. « Ontology and the lexicon: A multidisciplinary perspective ». Dans *Ontology and the Lexicon: A Natural Language Processing Perspective*, 3-24. https://doi.org/10.1017/CBO9780511676536.002.

Quillian, Ross M. 1968. « Semantic memory ». Dans *Semantic Information Processing*, 216-270. Cambridge : MIT press.

Ramos, Margarida. 2020. « Knowledge Organisation and Terminology: Application to Cork ». Thèse de doctorat en STIC Informatique, Universidade Nova de Lisboa ; Université Savoie Mont-Blanc. https://hal.archives-ouvertes.fr/tel-03106436.

Ramos, Margarida, Rute Costa, Christophe Roche, Sara Carvalho et Bruno Almeida. 2017. « A terminological and ontology-based approach to e-dictionaries construction: The cases of cork, endometriosis, and Islamic archaeology ». Dans *Workshop TOTh 2017: Terminology and Ontologies: Theory and Applications, within the scope of « Terminology and e-Dictionaries »*. Castelló, Spain. https://hal.archives-ouvertes.fr/hal-02140063.

Ranganathan, Shiyali Ramamrita. 1976. *Colon Classification*. Bombay : Asia Publishing House.

—. 1984. *New Trends in International Librarianship: S.R. Ranganathan Festschrift to Mark the Platinum Jubilee of the Madras University Library*. Allied Publishers.

Rastier, François. 1996. « Le terme : entre ontologie et linguistique ». *Texto !*, n° 7 : 35-65.

—. 1997. « Herméneutique matérielle et sémantique des textes ». Dans *Herméneutique : textes, sciences*, édité par Jean-Michel Salanskis, François Rastier et Ruth Scheps. Paris : Presses Universitaires de France.

—. 2004. « Ontologie(s) ». *Revue d'intelligence artificielle* 18 (1) : 15-40.

—. 2005. « Enjeux épistémologiques de la linguistique de corpus ». Dans *La linguistique de corpus*, par Geoffrey Williams, 31-46. Rivages linguistiques. Rennes, France : Presses universitaires de Rennes.

Reineke, Detlef et Laurent Romary. 2019. « Bridging the Gap between SKOS and TBX ». *Edition - Die Fachzeitschrift Für Terminologie* 19 (2). https://hal.inria.fr/hal-02398820.

Rey, Alain. 1979. *La terminologie: noms et notions*. Paris, France : PUF.

Reymonet, Axel, Jérôme Thomas et Nathalie Aussenac-Gilles. 2007. « Modélisation de Ressources Termino-Ontologiques en OWL ». Dans *Journées Francophones d'Ingénierie des Connaissances (IC 2007)*, édité par F. Trichet, 169-180. Grenoble, France : Cépaduès Editions. https://hal.archives-ouvertes.fr/hal-00365888.

Ribaud, Vincent et Patrick Le Boeuf. 2010. « Un wiki sémantique basé sur le CIDOC CRM ». Dans *Actes VSST 2010*, 1-12. https://hal.univ-brest.fr/hal-00630579.

Rivier, Alexis. 1990. « Construction des langages d'indexation: aspects théoriques ». *Construction des langages d'indexation: aspects théoriques* 27 (6) : 263-274.

Roche, Christophe. 2005. « Terminologie et ontologie ». *Langages* n° 157 (1) : 48-62.

—. 2007. « Dire n'est pas concevoir ». Dans *Ingénierie des Connaissances*. https://hal.archives-ouvertes.fr/hal-00509928.

—. 2008. « Le terme et le concept : fondements d'une ontoterminologie », février.

Roche, Christophe, Marie Calberg-Challot, Luc Damas et Philippe Rouard. 2009. « Ontoterminology: A new paradigm for terminology ». Dans *International Conference on Knowledge Engineering and Ontology Development*, 321-326. Madeira, Portugal. https://hal.archives-ouvertes.fr/hal-00622132.

Roche, Christophe et Maria Papadopoulou. 2020. « Terminologie et ontologie pour les humanités numériques : le cas des vêtements de la Grèce antique ». *Humanités numériques*, n° 2 (juin). Humanistica. https://doi.org/10.4000/revuehn.462.

Roget, Peter Mark. 1982. *Roget's thesaurus of English words and phrases*. Édité par Susan M. Lloyd. Harlow, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord : Longman, 1982.

Rousseau, Louis-Jean. 1974. « Le lexique anglais-français de l'industrie minière de l'Office de la langue française ». *Meta* 19 (4) : 189–196. https://doi.org/10.7202/004004ar.

Rousseau, Odile, Gonzague Gauthier, Alexandra Saemmer, Nolwenn Tréhondart, Marie Després-Lonnet, Rémi Mathis, Katell Briatte, Emmanuelle Bermès et Omer Pesquer. 2014. « 3. De la documentation à la médiation ». *Documentaliste-Sciences de l'Information* Vol. 51 (2) : 56-67.

Ruelle, Pierre et Samüel Jadon. 1953. *Le vocabulaire professionnel du houilleur borain: étude dialectologique*. Bruxelles, Belgique : Palais de l'Académie.

Salvador, Xavier-Laurent. 2016. *XML pour les linguistes*. Paris, France : l'Harmattan.

Schöch, Christof. 2013. « Big? Smart? Clean? Messy? Data in the Humanities ». *Journal of Digital Humanities* 2 (3) : 2-13.

Servicio Nacional de Geología y Minería et Sociedad nacional de minería. 1997. « Tesauro de la minería y Materias Afines ». Redimin: Santiago.

Souza, Renato Rocha, Douglas Tudhope et and Maurício Barcellos Almeida. 2012. « Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems ». *KNOWLEDGE ORGANIZATION* 39 (3) : 179-192. https://doi.org/10.5771/0943-7444-2012-3-179.

Sowa, John F. 2000. *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove, CA, Etats-Unis d'Amérique : Brooks/Cole.

Staab, Steffen et Alexander Mädche. 2000. « Axioms are Objects, too - Ontology Engineering beyond the Modeling of Concepts and Relations ». Dans *Proceedings of the 14th European Conference on Artificial Intelligence, Workshop on Applications of Ontologies and Problem-Solving Methods, Berlin, Germany*, 1-16. Berlin, Germany : W. Horn. https://publikationen.bibliothek.kit.edu/1000052359.

Studer, Rudi. 2009. *Handbook on Ontologies*. Édité par Steffen Staab. Berlin, Heidelberg, Allemagne : Springer Berlin Heidelberg : Imprint: Springer.

Studer, Rudi, V. Richard Benjamins et Dieter Fensel. 1998. « Knowledge Engineering: Principles and Methods ». *Data & Knowledge Engineering* 25 (1) : 161-197. https://doi.org/10.1016/S0169-023X(97)00056-6.

Stutzmann, Dominique. 2013. « Ontologie des formes et encodage des textes manuscrits médiévaux ». *Document numerique* Vol. 16 (3) : 81-95.

Suchanek, Fabian M., Gjergji Kasneci et Gerhard Weikum. 2008. « YAGO: A Large Ontology from Wikipedia and WordNet ». *Journal of Web Semantics*, World Wide Web Conference 2007Semantic Web Track, 6 (3) : 203-217. https://doi.org/10.1016/j.websem.2008.06.001.

Sun, J., J. Yue et H. Guan. 2009. « Study on the Application of Ontology in the Coal Mine Water Disaster Control System ». Dans *2009 Fifth International Conference on Semantics, Knowledge and Grid*, 428-429. https://doi.org/10.1109/SKG.2009.79.

Suonuuti, Heidi. 2001. *A Guide to Terminology*. 2ème édition. Finnish Centre for Technical Terminology. Helsinki, Finland : Nordterm. http://www.nordterm.net/filer/publikationer/guider/Guide_to_Terminology.pdf.

Svenonius, Elaine. 1978. « Facet Definition: A Case Study ». *International Classification* 5 (3) : 134-141.

Szabados, Anne-Violaine et Rosemonde Letricot. 2012. « L'ontologie CIDOC CRM appliquée aux objets du patrimoine antique ». Dans *3e Journées d'Informatique et Archéologie de Paris - JIAP 2012*. Paris, France. https://halshs.archives-ouvertes.fr/halshs-00752996.

Szilagyi, Ioan. 2014. « Technologies sémantiques pour un système actif d'apprentissage ». Thèse de doctorat en Sciences de l'information et de la communication, Besançon. http://www.theses.fr/2014BESA1008.

Termium Plus. 1971. « Concept : concept, idée, notion ». *L'Actualité terminologique* 4 (9) : 1.

TICCIH. 2003. « Charte Nizhny Tagil pour Le Patrimoine Industriel. Nizhny Tagil ». Russie : The International Committee for the Conservation of the Industrial Heritage. https://www.icomos.org/18thapril/2006/nizhny-tagil-charter-f.pdf.

Tirole, Delphine. 2016. « Assister les pratiques de lecture savante sur écran à l'aide des outils sémantiques ». Thèse de doctorat en Sciences de l'information et de la communication, Université de Franche-Comté. http://www.theses.fr/2016BESA1009.

Tremblay, Marcel. 1968. *Initiation à la recherche dans les sciences humaines*. Montréal, Canada : McGraw-Hill.

Troncy, Raphaël. 2004. « Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies : application à la description de documents audiovisuels ». Thèse de doctorat en Informatique : Systèmes et Communications, Université Joseph-Fourier - Grenoble I. https://tel.archives-ouvertes.fr/tel-00005263.

—. 2005. « Nouveaux outils et documents audiovisuels : les innovations du web sémantique ». *Documentaliste-Sciences de l'Information* Vol. 42 (6) : 392-404.

Turpin, Béatrice et Jean-Pierre Goudaillier. 2004. *Les mots de la mine*. Paris, France : Maisonneuve et Larose.

Turpin, Michel. 1988. « Le CERCHAR, une quête permanente d'identité ». Édité par Centre de recherche sur la culture technique. *Culture Technique*, n° 18 : 5.

UDC Consortium (The Hague). 2001. *Classification décimale universelle: édition abrégée*. Editions du CEFAL.

UNESCO. 1954. « Convention pour la protection des biens culturels en cas de conflit armé, avec Règlement d'exécution. » La Haye : Centre du patrimoine mondial de l'UNESCO. http://portal.unesco.org/fr/ev.php-URL_ID=13637&URL_DO=DO_TOPIC&URL_SECTION=201.html.

—. 1972. « Convention concernant la protection du patrimoine mondial culturel et naturel. » Paris : Centre du patrimoine mondial de l'UNESCO. https://whc.unesco.org/fr/conventiontexte/.

—. 1982. « Déclaration de Mexico sur les politiques culturelles. » Mexico : Centre du patrimoine mondial de l'UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000052505_fre.

—. 2003a. « Convention pour la sauvegarde du patrimoine culturel immatériel ». Paris : Centre du patrimoine mondial de l'unesco. https://ich.unesco.org/fr/convention.

—. 2003b. « Vitalité et disparition des langues ». UNESCO Publishing. https://ich.unesco.org/doc/src/00120-FR.pdf.

Uschold, Mike et Michael Gruninger. 1996. « Ontologies: principles, methods and applications ». *The Knowledge Engineering Review* 11 (2). Cambridge University Press : 93-136. https://doi.org/10.1017/S0269888900007797.

Van Heijst, Gertjan, Guus Schreiber et Bob Wielinga. 1997. « Using Explicit Ontologies in KBS Development ». *International Journal of Human-Computer Studies* 46 (2) : 183-292. https://doi.org/10.1006/ijhc.1996.0090.

Vandenbussche, Pierre-Yves. 2011. « Définition d'un cadre formel de représentation des Systèmes d'Organisation de la Connaissance ». Thèse de doctorat en Informatique Biomédicale, Université Pierre et Marie Curie - Paris VI. https://tel.archives-ouvertes.fr/tel-00642545.

Verlaet, Lise. 2008. « Modèle communicationnel de balisage générique pour la sémantisation des informations : le cas d'une revue scientifique ». Thèse de doctorat en Sciences de l'information et de la communication, Montpellier 3. http://www.theses.fr/2008MON30034.

Vickery, Brian Campbell. 1963. *La classification à facettes: guide pour la construction et l'utilisation de schémas spéciaux*. Traduit par Paule Salvan. Documentation et information. Paris : Gauthier-Villars.

Vidal, Solveig. 2006. « Visualisation de l'information : un panorama d'outils et de méthodes ». Research Report INIST-V-06-03. Institut de l'Information Scientifique et Technique (INIST-CNRS). https://hal-lara.archives-ouvertes.fr/hal-01456799.

Weissenberger, Roger et Michel Tugler. 1984. *Chronique des mines de potasse d'Alsace*. Bergholtz, France : Ziegler, impr.

Werner, Edeltraud. 2006. Recension de *Review of La prédication verbale et les cadres prédicatifs (Bibliothèque de l'Information Grammaticale, 54)*, par Jacques François. *Zeitschrift für französische Sprache und Literatur* 116 (2). Franz Steiner Verlag : 184-188.

Woods, William Aaron. 1975. « What's in a link: Foundation for Semantic Networks ». Dans *Representation and Understanding; Studies in Cognitive Science*, édité par Daniel G. Brobow et Allan Collins, 35-82. Academic Press.

Yengui, Ameni. 2016. « Système de recherche d'information sémantique pour les bases de visioconférences médicales à travers les graphes conceptuels ». Thèse de doctorat en

Informatique, Faculté des sciences économiques et de gestion Sfax. https://tel.archives-ouvertes.fr/tel-01767195.

Zacklad, Manuel. 2007. « Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI) ». Dans *Actes du 35e Congrès annuel de l'Association canadienne des sciences de l'information. Partage de l'information dans un monde fragmenté : franchir les frontières*, édité par Clément Arsenault et Kimiz Dalkir. Montréal, Canada.

—. 2010. « Évaluation des systèmes d'organisation des connaissances ». *Les Cahiers du numerique* Vol. 6 (3) : 133-166.

—. 2018. « Nouvelles tendances en Organisation des Connaissances ». *Études de communication - langages, information, médiations*, n° 50 (juin) : 91-108. https://doi.org/10.4000/edc.7566.

Zafio, Massiva. 1985. « L'arbre de domaine en terminologie ». *Meta : journal des traducteurs / Meta: Translators' Journal* 30 (2) : 161-168. https://doi.org/10.7202/004635ar.

Zemmouchi-Ghomari, Leila, Faiza Deghmani et Aya Meghnous. 2017. « Génération automatique d'un questionnaire à partir d'une ontologie de domaine ». Dans *28es Journées francophones d'Ingénierie des Connaissances IC 2017*, édité par Catherine Roussey, 205-216. Actes IC 2017 28es Journées francophones d'Ingénierie des Connaissances. Caen, France. https://hal.archives-ouvertes.fr/hal-01570290.

Zeng, Marcia Lei. 2019. « Interoperability ». *Knowledge Organization* 46 (2) : 122-146.

Ressources et corpus utilisés

- ***Dictionnaires et banques terminologiques***

Petit Larousse en couleurs. 1991. Paris, France : Larousse.

TLFi : *Trésor de la langue Française informatisé*, http://www.atilf.fr/tlfi, ATILF - CNRS & Université de Lorraine.

GDT : Québec, Office québécois de la langue française. *Le grand dictionnaire terminologique*, [En ligne], http://www.granddictionnaire.com.

- ***Ressources lexicales sur le vocabulaire minier employé dans les mines du Nord et du Pas-de-Calais et sélectionnées par analyse pour former la base de TerminoMines.***

[memo-mines_RL_19] Association pour la Pérennisation du Patrimoine Historique et Industriel Minier (AAPHIM). 2014. Lexique. En ligne : https://mineurdefond.fr/articles.php?lng=fr&pg=671&mnuid=662&tconfig=0
[memo-mines_RL_39] Bovio, Jean. 1906. *Vocabulaire technique des mineurs du Nord et du Pas-de-Calais*. Douai, France : Quantin-Dégez.
[memo-mines_RL_20] CentreIris. Vocabulaire du mineur. [Site d'information autour des mineurs.] En ligne : http://centriris.fr/vocabulaire/page19.php Lien qui ne fonctionne plus – site totalement réorganisé en 2020/2021, vocabulaire de nouveau accessible à cette adresse **:** https://centriris.fr/les-termes-miniers
[memo-mines_RL_27] Chaîne des terrils. 2007. Lexique. En ligne : https://web.archive.org/web/20070801023056/http://chaine.des.terrils.free.fr/lexique.htm
[memo-mines_RL_31] CHM. Centre Historique Minier de Lewarde. s.d. Lexique. En ligne : http://www.chm-lewarde.com/fr/ressources-et-collections/lexique/
[memo-mines_RL_38] Direction Départementale Des Territoires Et De La Mer Du Nord. Plan de Prévention des Risques Miniers (PPRM) Denain – Haveluy – Lourches. 2017. Pièce n°1 : Note de présentation. En ligne : http://www.hauts-de-france.developpement-durable.gouv.fr/IMG/pdf/1-pprm_note_de_presentation_denaisis_vap_18122017.pdf
[memo-mines_RL_48] Dumont, Gérard. 1999. *Sur les traces d'Antoine Delfosse: mineur de 1768 à 1823 : enquête sur les premiers temps de l'industrie minière dans le nord de la France*. Lewarde, France : Centre historique minier.
[memo-mines_RL_7] Laurent, Lucien. 1994. *Em' vie ed' Galibot : 1932-1969.* Equihen-Plage. 29p
[memo-mines_RL_18] Mongaudon, Jean-Pierre. 1988. *A la lueur de nos lampes*. Noisy le Grand, France : J.-P. Mongaudon.
[memo-mines_RL_42] Lebon, André. 1976. *Les mots de la mine: lexique dialectal de la profession de mineur au début du 20ème siècle dans la région de Denain*. France : s.e.
[memo-mines_RL_30] Mes Ancêtres d'Artois, Généalogie du Pas de Calais et du Nord. s.d. Lexique de la Mine. En ligne : http://ancetresdartois.com/history/mines.html
[memo-mines_RL_21] Nivoit, N. 2009. Gaillette, Lexique. [Billet de blogue]. En ligne : http://www.gaillette.fr/lexique.php
[memo-mines_RL_33] Ohland, B. s.d. « Métiers de la Mine ». Généawiki, Encyclopédie de la Généalogie. En ligne : https://fr.geneawiki.com/index.php?title=M%C3%A9tiers_de_la_Mine&oldid=2007195
[memo-mines_RL_14] Paillart André. 2008. Lexique des termes miniers - 1ère partie : de A à D. [Billet de blogue]. En ligne : http://andredemarles.skyrock.com/3113954741-Lexique-des-termes-miniers-1ere-partie-de-A-a-D.html
[memo-mines_RL_32] Régis F. 2014. Le Bassin Minier Nord Pas-de-Calais. Glossaire de la mine. En ligne : https://bassin-minier-regis.jimdo.com/un-peu-d-histoire/glossaire-de-la-mine/
[memo-mines_RL_35] Régis F. 2014. Le Bassin Minier Nord Pas-de-Calais. Les métiers de la mine. En ligne : https://bassin-minier-regis.jimdo.com/un-peu-d-histoire/les-m%C3%A9tiers-de-la-mine/
[memo-mines_RL_11] Turpin, Béatrice et Jean-Pierre Goudaillier. 2004. *Les mots de la mine*. Paris, France : Maisonneuve et Larose.
[memo-mines_RL_25] Van Rhijn, Lucie. 2006. Exposition virtuelle. Archives nationales du monde du travail (ANMT). Vocabulaire de la mine. En ligne : www.archivesnationales.culture.gouv.fr/camt/fr/memoires/donnees_expositions/06_11_06-07_07_27_mines/expo_virtuelle/html/ressources_complementaires/vocabulaire.php
[memo-mines_RL_16] Viseux, Augustin. 1991. *Mineur de fond: fosses de Lens : soixante ans de combat et de solidarité*. Paris : Plon.
[memo-mines_RL_12] s.n., Fabrice. 2016. Au pays des chtis, Le lexique de la mine. [Billet de blogue]. En ligne : www.aupaysdeschtis.com/le-lexique-de-la-mine/

ANNEXES

Annexe 1 : Extrait de retranscription de l'entretien effectué avec l'ancien mineur Jean à Arenberg

Note : Dans cette retranscription, les éléments en gras correspondent aux instances de classes relevées permettant d'enrichir OntoMines. La suite de trois astérisques *** représente du contenu inaudible.

Ancien mineur: ben voilà je me présente **Jean -anonym**.- d'**origine polonaise** donc mon **père** il est venu de **Pologne** en **1923**, il avait 13 ans quand il est descendu *au **fond*** et moi j'ai commencé avec lui à la sortie de l'**école** parce qu'à la sortie de l'**école** on avait même pas besoin d'aller chercher du travail c'est le ***garde des mines*** qui venait nous chercher il était au courant de nos années et tout donc voilà, j'ai commencé là-bas au **6 de Calonne** avec mon père et on était formés sur le tas par nos vieux comme on dit mon père, les oncles et tout alors venus de **Pologne** en **1923** voilà alors moi j'ai commencé à la sortie d'**école** voilà voilà [silence]

Interviewer: vous saviez ce que vous vouliez faire ?

Ancien mineur: ah travailler à la mine ? ah bah bien sûr c'était une fierté pour nous moi je voulais jamais que mon père il aurait travaillé pour moi vous savez à la sortie de la guerre il fallait du **charbon** et comme c'était la matière première ben nous on savait que si on allait se **marier** jeune, d'abord on s'est **marié** jeune, hein, qu'on aurait possibilité d'avoir une **maison** et être indépendant, donc c'est pour ça qu'on préférait **aller à la mine**, malgré qu'il y avait d'autres travaux aux environs mais voilà et c'est comme ça que j'ai **dévalé** le premier coup **au fond**

Interviewer: vous vous souvenez de votre première descente au fond ?

Ancien mineur: ben oui que je me souviens, on a descendu à 1100 mètres enfin que ça soit 500 mètres, 600 mètres ou 1100 mètres, mais on est tellement compressé dans la **cage** comme des sardines qu'on se n'en rend même pas compte quand on arrive au fond s'il y a la compression on sent quand même quelque chose qui se passe dans le corps mais pff après c'est devenu une habitude

Interviewer: et vous aviez peur la première fois ou pas du tout ?

Ancien mineur: bah il y a quand même une appréhension ben bien sûr il y a une appréhension on se demande qu'est-ce que c'est une **cage** avec un **câble** autant de bonhomme dans la **cage** et l'appréhension elle existe parce que on dit et si le **câble** aurait cassé vous ne voyez pas tout le **personnel** arriver **au fond** ben on était écrabouillé l'appréhension sûr mais après ça devient une habitude

Interviewer: qu'est-ce que vous avez fait comme métier vous au fond ?

Ancien mineur: **au fond** ? ben au début on **transportait les matériaux** pour les vieux c'est-à-dire on leur apportait les **bois**, les **étançons** et après ils nous donnaient les **marteaux-piqueurs** pour nous former sur le tas alors on **tapot dans le charbon** il faut pas croire que parce qu'il est noir qui faut taper n'importe dû, parce qu'il y avait des **clayats**, des **limées** donc c'est eux les vieux nous surveillaient qui disaient voilà il faut **taper** là il faut **taper** là il faut **taper** là donc c'est eux qui nous formaient

Interviewer: donc qu'est-ce que vous avez fait comme métier après quand vous avez été mineur ?

Ancien mineur: ah voilà quand j'ai été **mineur** et que j'ai vu mon père à 4 pattes travailler dans les travaux les plus difficiles mon père ne savait pas s'expliquer et moi à l'école primaire j'étais l'un des meilleurs ouver- ouvr- euh l'un des meilleurs élèves je me suis dit en moi-même pourquoi pas aller plus haut c'est-à-dire aller comme **porion** parce que j'ai vu *** **porion** les **agents de maîtrise** qui commandaient le personnel et tout alors comme je me suis dit pourquoi pas alors après le travail ils ont formé des **écoles** qu'on pouvait devenir **cadre** alors j'ai dit en moi-même et pourquoi pas risquer le coup, alors il fallait faire les 4 ans d'école tous les ans avec examens alors comme je me souviens on était 40 au départ, et à la fin de la 4ème année il y en avait 9 qu'on était reçus pour être admis comme **agent de maîtrise** mais ce qu'il y a, il fallait encore faire un stage de 6 mois et comme les **Houillères** fermaient au **Pas-de-Calais** on nous a proposé deux choses, soit de **quitter la mine**, **aller à l'usine** ou aller **continuer dans les mines** il fallait **partir dans le Nord** alors moi j'étais déjà marié on avait des

enfants et je dis on va risquer on va **partir dans le Nord** et d'un côté on a bien fait, parce que j'ai pu avoir ma **retraite** complète 3 ans avant la **fermeture** c'est-à-dire j'avais l'âge limite **50 ans c'était l'âge limite**

Interviewer: donc vous vous n'avez pas fait la fermeture ?

Ancien mineur: non j'ai eu ma **retraite** en **86** et la **mine elle a fermé en 89** 3 ans avant la fermeture j'ai eu ma retraite ce qui était que vous savez quand on a 50 et qu'on voit un billet à la lampe que à partir de telle date vous ne faites plus partie de l'effectif, qu'on vous remercie pour les services rendus ben écoutez j'étais un peu gêné parce qu'à 50 ans je me trouvais encore jeune malgré les 36 ans passés à la mine je pouvais pas m'imaginer qu'à partir de maintenant je vais être en retraite mais d'un autre côté j'ai remercié le bon dieu j'ai remercié le bon dieu pour la simple raison que j'ai échappé 3 fois et j'ai eu un copain de tué à, de 35 ans qu'on a été le chercher, rechercher qu'il a fallu le laver, le grand-père à ma femme **tué** dans un **éboulement** j'ai eu beaucoup de chance j'ai eu beaucoup de chance, j'étais un peu, comment dire un type plutôt comme on disait trompe la mort j'avais pas peur des **accidents** quand j'étais **agent de maîtrise**, quand je suis arrivé dans le **Nord**, autant que dans le **Pas-de-Calais** on avait plutôt des **européens** c'est-à-dire plus de **français-polonais**, ici (**Arenberg**) j'ai trouvé une **communauté de Nord-Africain** alors comme j'ai vu qu'ils étaient un peu perdus je me suis imaginé que j'ai commencé comme eux à la place voilà alors je les engageais pas dans les grands **travaux à risques** c'est moi qui allait dans les **travaux à risques** et quand c'était garanti parce que le **fond** on peut dire tout ce qu'on veut, le **fond** c'est un autre monde celui qui a pas peur ben je crois qu'il doit pas être normal parce que **au fond** c'est travail, bon il y avait des travaux plus ou moins risqués comme dans les **galeries** comme dans les **bowettes** mais moi pour dire j'ai fait toute ma carrière au **charbon**, dans les **tailles** d'abord je me souviens que il y avait une **taille** elle était tellement dangereuse c'était une **taille** qu'on appelait à **foudroyage retardée** donc c'était une **taille** qu'on disait qu'on va faire beaucoup d'**économies** parce qu'on va mettre un minimum de **bois** alors là croyez-moi quand j'y allais je mangeais même pas parce que j'avais mes boyaux tellement crispés que j'avais même pas faim et puis le bonheur qu'on a eu, tout il s'est écrasé au cours d'un week-end et j'ai fait le dernier **poste** samedi de nuit et quand le lundi je suis arrivé il y avait tout les gens dans le **bureau**, le **directeur**, **ingénieur** il dit bon tout est arrêté dans cette **taille** on abandonne tout et on va plus recommencer ce **système d'exploitation** tellement c'était dangereux

Interviewer: ça c'est la première fois que vous avez réchappé?

Ancien mineur: ouais après j'ai **réchappé** comme mon copain qui a été **tué** à **35 ans** parce qu'il a pris des risques pour une **balle de charbon** qui est partie tisonnée et comme il y avait un grand **pendage** tout il est venu sur lui et il a été bloqué par un **bois** moi j'ai eu la chance que j'étais dans les mêmes conditions mais au lieu de me laisser glisser avec tout le **charbon** qui venait, on aurait dit qu'il y a une main qu'elle m'a reconnu pour me tirer en amont alors là le **charbon** il était presque englouti et après tout doucement le **charbon** s'en allait pour être **chargé** ben on voit la **mort** dans des conditions par là je voyais des, je me voyais déjà de l'autre côté et une fois aussi pris dans les **fumées** dans les **fumées** qu'on avait **tir la dynamite**, dans une **galerie** je me suis précipité pour voir s'il y a pas trop de dégâts et j'étais pris dans les **fumées**, alors là aussi j'ai été presque **asphyxié** par les **fumées**

Interviewer: et après ces 3 coups durs là vous y êtes retournés sans arrières pensées ?

Ancien mineur: ben si les arrière-pensées existent et ont toujours existées jusqu'à la fin ouais non mais vous savez quand on voit un copain qu'on le **remonte** les yeux ouverts, la bouche grande de **charbon** et tout mais c'est pas tout mais après, vu qu'on était amis on se fréquentait en famille et qu'il fallait prévenir sa femme et ses enfants que le père ne va plus revenir ben croyez-moi c'est jusqu'à aujourd'hui je l'ai pas encore oublié et ce qui est chose incroyable que on m'a changé de **poste** comme je suis arrivé lui, on le remontait pour dire il a été **tué** et peu de temps après on s'est rencontré en famille et sa femme Arlette elle a dit, c'est dommage que Jean il sera pas avec moi aujourd'hui comme si elle pressentait ce **danger**, est-ce que j'aurais pu le **sauver** ? je me pose des questions jusqu'à aujourd'hui est-ce que j'aurais pu le **sauver** ? parce qu'il aurait peut-être, enfin c'est vous savez on dit **accident du fond**, ça peut arriver **au jour**, ça peut arriver sur la route, c'est imprévu imprévu non, comme je dis j'ai eu beaucoup de chance, et je remercie le bon Dieu et tous les ans je m'en vais à la **messe de Sainte Barbe** voilà

Interviewer: et le danger vous y pensiez quand vous travailliez dans le fond ?

Ancien mineur: il fallait se garantir vous savez, il fallait se garantir des fois un seul **bois** qui manquait et que le gros **banc** au-dessus de nous où il donnait on pouvait plus se sauver là c'étot cric **l'écrasement** complet ah oui parce que quand on était dans des **terrains** très durs, très lourds donc en plus on disait on mettait des **rallonges articulées** | garantie | chose primordiale au fond garantir celui qui vient derrière nous

Interviewer: et entre les hommes alors ça se passait comment parce que vous pensiez tout le temps aux autres aussi ?

Ancien mineur: toujours, toujours ah oui comme je dis moi j'avais plus peur pour, quand j'étais **agent de maîtrise** j'avais plus peur pour mes **ouvriers** que pour moi-même c'est pour ça que je m'engageais dans les travaux difficiles quand la **chaîne** elle cassait c'est moi qui allait dans le **découvert** et je leur disais vous vous ouvrez l'œil, l'œil et l'oreille [en arabe] /undik laudnik/ je parlais parce que je me suis très bien défendu après vu que j'avais presque tout des **ouvriers nord-africains** avec moi j'ai eu l'idée de faire mon petit **dictionnaire franco-arabe** et mot par mot, mot par mot, ben je me démerdais pas mal je pouvais m'exprimer avec eux et c'était un grand bonheur

Interviewer: vous me regardez là alors ce que vous venez de dire là ça veut dire quoi ?

Ancien mineur: ouvrir l'œil et l'oreille

Interviewer: comment on dit en arabe ?

Ancien mineur: /laidnik udnik/

Interviewer: et en polonais comment on dit ?

Ancien mineur: "oko i uszy"

Interviewer: et il y a d'autres langues que vous parliez ? en patois comment on dit ? en patois *rires*, en patois ah oui il y a un terme aussi en patois pourquoi pas, ben vous ne le disiez pas

Ancien mineur: non

Interviewer: vous le disiez en français, en arabe en polonais

Ancien mineur: en polonais

Interviewer: et trompe la mort c'était votre surnom ?

Ancien mineur: oui, j'avais un **chef porion** d'origine allemande et ben vous savez que les allemands vu qu'ils ont vécu la guerre aussi parce que moi aussi j'ai vécu la guerre jusqu'à une dizaine d'année j'ai ben on a dormi dans des caves dans l'eau et tout, quand les gros bombardiers passaient au-dessus de nous ma grand-mère avec le chapelet à genou parce qu'on est d'origine polonaise et croyants pratiquants donc vous voyez alors en dehors de ça après j'étais comme rassuré, j'avais peur mais rassuré

Interviewer: et donc cette croyance ça vous aidait à aller au danger ?

Ancien mineur: ah oui elle m'a beaucoup aidée ma **croyance**

Interviewer: sur quel point par exemple ?

Ancien mineur: ben comme j'ai dit là j'aurais pu avoir le tour de mon copain qui a été **tué** là on aurait dit vraiment qu'une main, elle m'a tiré vers l'amont parce que si je me relaissais glisser je partais comme lui j'aurais été étouffé parce que là c'est des millions, tout le **charbon** qui vient sur vous, vous êtes complètement **armuché** quoi c'est une tombe, ça devient une tombe

Interviewer: et aujourd'hui avec le recul ça représente quoi la mine dans votre vie ?

Ancien mineur: la mine elle restera toujours **dangereuse** on peut pas éviter mais ce qui est dommage c'est qu'on les ait **fermées** parce que je vois mal maintenant le **charbon** être compensé par le nucléaire et tout ce qui s'ensuit euh le **charbon** nous on vivait avec le **charbon** on **chauffait** la gamelle au **charbon** tout ce qui était c'était **charbon** et après du jour au lendemain quand on voyait disparaître cette matière on était un peu perdu on était un peu perdu

Interviewer: et tous vos copains c'était le charbon aussi ?

Ancien mineur: bah tout le monde ne travaillait pas dans les mêmes **conditions** moi c'était le **charbon** toujours presque dans les **tailles** il y en a qui travaillaient dans les **galeries**, dans les **bowettes** et il y avait d'autres, il y en avait qui **chargeaient** seulement le **charbon** il y en a qui **boisaient** seulement donc c'est pas qu'ils étaient exposés, qu'ils étaient moins exposés au **danger**, il y avait autant de **danger** tout partout mais moi dans les **tailles**, ben croyez-moi euh moi je crois que c'était là le plus grand **danger** les **veines de charbon** parce que les **veines** étaient irrégulières, des fois elles étaient de 40 cm, des fois 1m20 moi j'ai connu aussi des **veines dressants** qu'on appelait, des **veines** qui montaient à 90° bah là les **bois** s'ils étaient mal serrés ou quoi, on montait sur un **bois** ben on dégringolait de ben on passait de l'autre côté de la barrière

Interviewer: et votre famille elle disait quoi de ce métier, votre femme ?

Ancien mineur: ma femme, j'ai jamais discuté, durant mon activité, j'ai jamais parlé à ma femme de la **mine** pour pas la traumatiser vu qu'elle avait son grand-père qui a été **tué** pour pas traumatiser j'en ai jamais parlé à la **mine** je gardais tous pour moi par contre quand après le **service** mil- parce que j'ai fait mon **service militaire** aussi et par contre quand on était en vacances là, pour dire on se défoulait un peu avec d'autres **mineurs** on se défoulait un peu mais sinon à la maison non j'évitais de parler, j'évitais

Interviewer: elle elle sait comment il est mort son grand-père ?

Ancien mineur: euh en voulant **sauver** des copains ouais c'était un **éboulement** euh il voulait **sauver** 2 copains et ils sont restés tous les 3

Interviewer: et vous vous avez connu ça des gens qui ont disparu en voulant en sauver d'autres ?

Ancien mineur: en voulant suivre ?

Interviewer: sauver d'autres

Ancien mineur: en voulant **sauver** d'autres ? dans des **conditions** vous savez même moi ça m'est arrivé que je pense que j'ai sauvé d'autres dans les risques que je prenais, les risques que je prenais, si j'aurais envoyé un autre à ma place, euh voilà la question elle est là, et si il lui arrivait quelque chose ? à ma place et ben aujourd'hui j'aurais souffert ah oui ouais

Interviewer: au final vous êtes fiers d'être mineur ?

Ancien mineur: oui, très fier, très fier d'être, d'avoir été mineur ouais parce qu'il faut dire malgré le **travail** difficile le risque la peur on était quand même privilégié hein parce qu'on a eu quand même une **maison**, on avait le **charbon**, pour dire on avot le nécessaire quoi, le nécessaire parce que après quand on a eu 2 enfants, on avait pas encore eu de voiture mais quand le 3ème il est arrivé on a pu acheter une petite voiture, une SIMCA ***, elle bouffait plus d'essence, plus d'huile que d'essence mais on était fiers parce qu'on avait une voiture

Interviewer: pour vous c'est, vous me regardez ?

Ancien mineur: ouais ouais

Interviewer: euh pour vous c'était une belle vie ?

Ancien mineur: une belle vie ?

Interviewer: hm

Ancien mineur: ben oui, si on pouvait se permettre de faire des sorties parce que dans le temps

Interviewer: problème technique ouais alors votre surnom c'était, c'était quoi ?

Ancien mineur: ch'cahou

Interviewer: et pourquoi ch'cahou ?

Ancien mineur: parce que ben euh quand on **remontait au jour**, la plupart du temps on allait **boire un verre** en face de la mine, il y avait là des **cafés** alors on pouvait aller b**oire un demi de bière** et puis là j'ai raconté une histoire en **patois**, l'histoire d'un chat, je me rappelle plus tellement le thème et puis vu que je disos au lieu de dire le chat j'ai dit ch'cahou alors ce mot il a été enregistré parce que c'était un, un terme nouveau et puis du jour au lendemain dans le **café** la patronne elle m'a surnommé ch'cahou et

après c'est devenu un automatisme que jusqu'à la fin je suis resté ch'cahou c'est un chat en **patois** c'est un chat

Interviewer: d'accord euh je sais plus ce que c'était la question que je posais avant qu'on coupe tu te souviens Adrien ?

Ancien mineur: si j'étais fier du métier

Interviewer: ah oui c'était ça bon on reprend ça alors mais il faut me regardez hein

Ancien mineur: oui

Interviewer: alors vous êtes fier d'avoir été mineur ?

Ancien mineur: oui très fier

Interviewer: pourquoi ?

Ancien mineur: et ben parce que on a gardé quand même pas mal **d'avantages** alors qu'il y a tellement de **chômage** aujourd'hui et ben nous on a quand même le bonheur d'avoir quand même notre **retraite**, malgré qu'elle ait diminué et qu'on a le bonheur de, moi mon bonheur euh c'est de pouvoir donner la petite gringuelle à mes petits-enfants, à mes petits-enfants ben moi je crois que c'est un grand bonheur et nous on a une famille très unie les enfants, les petits-enfants euh on en est même à la 4ème génération et le bonheur c'est ça quand tout le monde se rencontre mais j'aurais pas voulu par exemple avoir des enfants mineurs grâce à Dieu on a eu 4 filles donc le problème ne s'est pas posé mais si j'aurais eu des gamins j'aurais eu peur pour eux

Interviewer: mais pourquoi parce que vous vous avez aimé ce métier

Ancien mineur: mais oui mais non on ne peut pas expliquer c'est un danger continu, un, j'aurais pas voulu que, s'ils auraient insisté on peut pas les empêcher mais s'ils auraient des bonnes têtes ou bien quoi pour pas parler autre que la mine, parce que à l'époque quand on était là ben on avait toute possibilité mais maintenant je vois mal non j'aurais eu trop peur, trop peur de les savoir tant qu'ils seraient pas **remonté au jour** tant que je les aurais pas entendu buquer comme taper à la porte qui est remonté non c'est dur à expliquer mais c'est comme ça

Interviewer: et vous si c'était à refaire vous referiez la même chose ?

Ancien mineur: ben je crois que oui je crois que oui mais j'avais aussi une bonne possibilité en plus de la mine on avait possibilité de faire une carrière au **football** et moi j'avais des qualités de **footballeur** alors quand on travaillait on avait possibilité de rester du **poste du matin** au **charbon** mais **travailler du matin** comme ça les mardi et jeudi on pouvait aller à l'entraînement du **football** donc et comme j'avais pas mal de qualités, j'ai même joué au **stade Bollaert** contre **Lens** au **parc des Princes** contre **Paris** et j'avais des propositions quand j'avais 16 ans d'aller à **Toulouse**, les **dirigeants de Toulouse** ils sont venus parce que j'avais fait un très bon match et ils étaient là comme sur euh, en train de ausculter le match et ils sont venus et ils ont proposé de m'emmener pour être **stagiaire professionnel** mais j'avais encore mes parents, ma grand-mère, pas moyen de locomotion alors quand je les ai vus dans la détresse en train de pleurer qu'ils me reverraient plus, donc je dis bon après tout je reste, je vais rester là et puis je vais rester pour la mine

Interviewer: et vous avez regretté ça ?

Ancien mineur: euh au début oui, mais maintenant quand je revois des copains qui ont tenté la carrière professionnelle, on peut les compter sur les doigts, combien qu'ils ont réussi et quand ils me regardent des fois ben ils me disent tu sais Jean, t'as plus que mi, parce que la carrière n'est pas longue la carrière elle n'est pas longue, par contre je suis en très bonne relation avec un ancien **joueur** qui était considéré comme une vedette de l'équipe de France, un **polonais**, **Maryan Wisniewski** qui est actuellement à **Carpentras** et a joué la **Coupe du monde 58**, c'est l'année où **Juste Fontaine** il a marqué les 13 buts jamais égalés jusqu'à ce jour et de l'autre côté il y avait un autre **joueur Jean Vincent** c'était des **joueurs** qui ont joué ensemble à **Auchel** dans le **Pas-de-Calais** parce qu'il y en avait d'autre il y avait **Maryan Wisniewski**, il y en a eu, il y en avait pas mal de **joueurs** qui avaient des qualités, on avait un bon niveau, on avait un bon niveau de **football** et surtout qu'on avait un **entraîneur** très sérieux qui a fait aussi une bonne carrière **Élie Fruchart**, **Élie Fruchart**, c'était un **entraineur** qui a quand même été pris pour entraîner **l'équipe de France dans les années 58**, mais regretter euh on peut plus retourner en

arrière, après j'avais une proposition d'aller à **Boulogne** quand **Boulogne** a instauré le professionnel mais ma femme elle était enceinte du premier, alors c'est moi qui a eu peur, parce que on avait quand même la **maison**, on avait la **maternité des mines**, on avait tous les **avantages des mines**, donc c'est moi qui a abandonné le projet et puis voilà je suis resté jusqu'à la fin de ma carrière

Interviewer: il y a encore une solidarité polonaise particulière à l'intérieur des mineurs ?

Ancien mineur: toujours, il y a toujours une **solidarité** et ça ne peut pas disparaître

Interviewer: alors expliquez-nous en quoi ça consiste ?

Ancien mineur: ben quand on fait les **fêtes** quand on fait les **fêtes**, si bien euh nous là-bas, pour dire c'est presque tout des **musiciens d'origine polonaise** mais actuellement les **polonais** ils ont le nom **polonais** mais ils ne savent plus le **parler** comme moi, mes enfants déjà euh la plus vieille, ma fille elle comprend, *** elle comprend le **polonais** mais les autres ils ne savent plus un mot mais le plus qui nous reproche c'est nos petits enfants alors ils me disent *dziadzia* grand-père, pourquoi que tu nous as pas appris à parler le polonais parce que moi je sais lire parler écrire couramment le polonais alors euh des fois avec ma femme quand on discute en polonais et tout euh, vous voyez les oreilles des enfants, qu'est-ce que t'as dit ?

Interviewer: et alors pourquoi vous leur avez pas appris ?

Ancien mineur: pourquoi je leur ai pas appris, pour la simple raison que quand les enfants, à la maison moi ma femme on leur parlait en polonais, ils voulaient pas répondre parce qu'ils disaient **à l'école on ne parle pas le polonais** et voilà comment cette langue a été abandonnée par mes enfants parce qu'à l'école on parlait pas le polonais c'est resté comme ça

Interviewer: euh est-ce que c'est important pour vous que la mine soit sauvée aujourd'hui?

Ancien mineur: maintenant qu'elle est devenue **classée monument historique** oh bah oui hein ah oui rien que pour le souvenir, rien que pour ah ben oui et puis ça donne quand même du travail ça donne du travail **classé monument historique** c'est un événement pour dire mondial parce qu'il y a beaucoup de sites qui sont **classés** maintenant **à l'UNESCO** aussi bien là-bas dans le **Pas-de-Calais**, dans toute la **France** euh là-bas hier on a vu une émission euh les racines et les ailes, le **cirque de Gavarnie** et moi j'ai été en haut des montagnes parce que à **l'école des cadres pendant les 4 ans** avec les **Houillères** on était obligé de faire des **stages** dans les montagnes une année on a été dans les **Pyrénées** donc la **brèche de Roland**, la **Gavarnie** et la 2ème année les **Alpes** alors on nous faisait grimper à presque 3800 mètres le point alors hier quand j'ai revu l'émission, que j'ai revu le refuge où on a dormi, le **refuge Felix-Fort**, les glaciers, il y en a presque plus parce que 40 ans en arrière là, il y avait des patinoires hein alors ça reste des souvenirs incroyables mais comme on dit oui grâce à ce système que ici **on a gardé la mine** qu'on peut, tant qu'on est en vie, parce que nous c'est déjà la dernière génération, derrière nous il y en a plus que faire euh, que faire voir aux enfants quand ils viennent, les **écoles** ou quoi ou même des personnes qui viennent de l'extérieur il y en a qui viennent d'autres pays ou quoi, visiter qu'on peut leur expliquer ce que c'était la mine et ben ils sont des fois euh enchantés et oui c'est le souvenir rien qu**'ici** en regardant en haut vous voyez pas à chaque **molette** qu'il y avait là des **vestons**, des **godasses**, des **casques**, les **bancs numérotés** moi quand je viens ici je regarde et je crois qu'il y a toujours les **bancs numérotés** qu'il y a tout le temps les **chaînes** qui sont pendues c'est encore gravé c'est encore dans la mémoire

Interviewer: et pourquoi alors avoir voulu transmettre cette mémoire ?

Ancien mineur: pourquoi avoir voulu transmettre ?

Interviewer: oui

Ancien mineur: ben ce n'est pas venu vraiment de nous-mêmes, c'est pas venu, c'est grâce au film

Interviewer: faudrait que vous regardiez là-bas

Ancien mineur: c'est grâce au **film** nous on peut dire un grand merci à **Claude Berri** parce qu'on, si bien quand on a **arrêté** en **mars 89** j'étais déjà pour dire à la **retraite** mais je dis qu'est-ce qu'il va devenir et après il y a eu ce **Claude Berri** qui cherchait l'environnement pour un **film Germinal** parce que **Germinal** c'est quand même un **livre** qui a été écrit par **Emile Zola** en **1800 et quelques** et quand **Claude Berri** il a vu les vieux **corons** parce que les **corons** ils étaient encore à l'ancien style et tout

parce qu'il aurait pu partir en **Pologne**, ça aurait revenu moins cher mais comme ici c'étaient presque des mineurs du coin alors lui il voulait des gens vrais comme nous qu'on connaissait le travail qu'on pouvait donner même des conseils alors de là c'est parti alors quand on a **tourné le film** croyez-moi on était enchanté hein trouver des **Depardieu** des **Renaud** des **Miumiu** enfin il y en a au moins 7, 8 qui sont déjà décédés mais la mémoire elle reste parce que maintenant on peut dire un type comme **Renaud** c'était le gars euh un cœur d'or il mangeait avec nous il **chantait** avec nous, c'est d'ailleurs de là que c'est parti les **chansons en patois** qu'il chante ***tout en haut de ch'terril***, ***viens boire une goutte ed jus***, ch'est à table qu'on chantait tous

Interviewer: vous pouvez les chanter ces chansons à Renaud en patois ?

Ancien mineur: ***tout en haut de ch'terril*** ?

Interviewer: ouais

Ancien mineur: oh j'saurais pas

Interviewer: non c'est là qu'il faut regarder

Ancien mineur: non, non

Interviewer: et une histoire en patois non plus ?

Ancien mineur: une histoire là vous me prenez vraiment au mot une histoire ça va revenir allez voir, il y en a tellement ah ouais ouais ouais une histoire en **patois** c'est un gamin qui est revenu de **l'école**, il est revenu de **l'école** et il demande à son père : papa, est-ce que c'est loin l'**Italie** ? et son père il a réfléchit, ben il dit, ben non, c'est pas loin l'**Italie** il dit, je travaille avec un **italien** et ben tous les jours il vient travailler à **vélo** ah ouais après il y avait un chtiot aussi, il est venu au catéchisme et le curé lui demande, alors Joseph, saurais-tu me dire quand est-ce que Jésus Christ est mort ? oh Monsieur le Curé vous savez nous on reste dans un petit village, on n'a même pas de télévision, on a même pas de poste de télévision, on n'a même pas de poste TSF ben on savot même pas qu'il était malade

Interviewer: *rires* bon est-ce que vous avez d'autres questions ?

Ancien mineur: non

Interviewer: vous avez une petite anecdote en plus à nous raconter pour finir ?

Ancien mineur: une anecdote, et je dis pourvu que ça reste, pourvu que tout il reste tel que et qu'on améliore de mieux en mieux une chose qui est dommage c'est qu'**on a pas gardé le terril** parce qu'on aurait pu aller ***tout en haut de ch'terril*** hein ah ouais malheureusement il a été **exploité** pour **remblayer les puits** il a été, une partie est parti vers la **centrale électrique**, une partie les **concassages** vers les autoroutes et voilà sans quoi on avait un très beau **terril**, **110 m de haut**, l'un des plus beaux on aurait coupé la tête du terril, on aurait fait une plateforme et entre le grand **chevalet** un **téléphérique** ça aurait encore donné du boulot et les gens pourraient voir aux alentours ohlala à partir du haut des **ch'terrils**

Interviewer: ah oui à l'estaminet, non là, à l'estaminet euh vous avez connu des gens qui maquillaient leur feuille de paie pour que la femme elle voit pas qu'il manquait de l'argent ?

Ancien mineur: ah parce qu'à l'époque les carnets de chèque n'existaient pas donc il y avait la **quinzaine** et on était **payé en espèces** et ben oui, il y avait des trucs à faire hein parce que il faut dire les **femmes** à l'époque elles savaient pas compter comme vous ajd même les euros, vous les comptez bien mais nous de notre temps on donnait l'argent comme ça, la femme c'est tout juste qu'elle contrôlait *** assez pour aller chercher la farine le sucre et le pain hein donc on prenait un petit peu pour aller à l'**estaminet** pour aller boire un coup avec les copains ben les **ducasses** vous savez c'était des grands évènements à l'époque, une **ducasse** c'était phénoménal, les **bals volants**, les orchestres c'était si bien autant qu'on était vraiment pour dire crevé toute la semaine à travailler, ben quand ça arrivait déjà le dimanche, et le samedi et le dimanche c'était la fête la fête ah oui, on était pour dire ce qu'on gagnait ben on était tellement contents qu'on pouvait donner ben les **cafés** à l'époque les **bistrots** les **estaminets** mais ils étaient remplis, les **ducasses**, les gens se bousculaient c'était la foire presque à chaque fois, vous savez quand il y a de l'argent il y a du bonheur il faut pas une grosse somme comme aujourd'hui les **footballeurs** qui gagnent des millions ou les chanteurs ou bien quoi non, nous avec le peu qu'on avait on pouvait encore s'amuser et boire un coup, voir les copains, aller au **bal**, aller au **cinéma** parce qu'au **cinéma** à l'époque il fallait réserver le vendredi pour avoir une place le dimanche les strapontins,

les films dans le temps c'est, ben il y avait pas de télé il faut dire, il fallait quand même passer son dimanche quelque part mais on n'était pas malheureux parce qu'il faut dire nos **vacances** vu qu'on élevait nos bêtes, les lapins, les cochons, les oies, les poules on les élevait nous-mêmes, après quand on tuait les bêtes surtout les cochons quand on tuait le cochon on découpait les lards on le faire frire sur les payelles et puis on faisait notre **briquet** au **fond** alors on mettait les tranches de lard avec un petit peu de sel comme cha cha, nos **vacances**, même nos **vacances** c'était dans les champs, on allait travailler dans les champs pour les fermiers ouais et puis on ne se plaignait pas pour nous c'était naturel c'était naturel, on **vacances**, mooon qu'est-ce que c'est, on ne savait pas, non mais une fois qu'on ne sait pas ben après bon ben quand on a eu possibilité le premier coup d'aller en **vacances** à **la Napoule**, à **la Napoule**, en **autobus** [siffle] oh la la la fiesta ah ah ah ben on s'imaginait pas là-bas quand on voyait les tables garnies avec la bidoche, les fromages et tout du matin au soir hein et après on a eu l'occasion d'aller à **la Napoule** avec l'avion *** avec l'avion bah le premier coup quand on a monté dans l'avion, écoutez bien, on était là à **Lesquin** vlà les autobus qui s'amènent, premier, deuxième, troisième, dixième autobus, 10 autobus de 40 personnes 400 personnes ? L'animal il va pas décoller, mais quand il a décollé, j'étos avec m'femme main dans la main ben je transpirais c'est comme le premier coup quand j'ai dévallé au **fond**, des gouttes de peur, l'appréhension alors les enfants ils disaient viens vir papa ils étaient à côté de l'hublot oh la laisse tranquille, mais ouf hein quand il a atterri ouf

Interviewer: donc vous aviez plus peur en avion qu'à la mine ?

Ancien mineur: la même appréhension, le premier coup et après c'est mi qui faisot peur aux autres et le 2ème coup, ch'étot mieux et quand j'ai vu un gros **polonais** qui s'amenait là il faisait au moins 120 kilos je ne le connaissais même pas, c'est le premier coup *** alors je lui ai dit [polonais] : mets-toi pas du côté de l'aile parce que l'avion il va baller quand j'ai vu le gros transpirer des gouttes plus grosses que mi non non je dis je l'ai rassuré l'avion il balle pas va je l'ai rassuré

Interviewer: et donc ça veut dire qu'après votre poste du matin en fait vous continuiez à travailler après au jardin

Ancien mineur: ah oui on allot glaner pour les bêtes après notre **poste** c'était pas fini on remontait alors mardi jeudi j'allais à l'entraînement mais après il y avait tout le temps, quand on a des bêtes il faut aller chercher l'herbe, les orties pour les oies, qu'il fallait couper ah oui, on n'était jamais au chômage non au soir, on dormait bien au soir, quand on avot fini le double poste ouais pensait même pas à aller branquiller comment faire du grappillage ou quoi ah oui pendant la **guerre** on a eu des misères, manque de **chauffage** donc il fallait aller sur les **terrils** grappiller du **charbon** et du schlamm pour chauffer la gamelle et on était poursuivi par les **gardes des mines** et si il nous attrapait on était sanctionné ça et dans les **corons** il fallait tout le temps le **garde des mines**, il avait une autorité incroyable il fallait **balayer les ruisseaux** il fallait que ça soit nickel, ça et puis comme aujourd'hui regardez un peu, quand on faisait les cultures de pommes de terre, on faisait des allées, et ben les engrais c'était quoi ? c'était nos excréments parce que il y avait pas de water à l'époque, les waters ils étaient au bout du... alors il fallait chercher du water et tout et on mettait ça et il y avait personne de malade, il y avait personne, il y avait pas de virus rien et tout était naturel ah oui alors aujourd'hui on se demande de quoi que ça vient quand on entend parler des maladies là ben on se demande de quoi que ça vient même l'eau, même l'eau on allait ben quand on jouait au **football** dans la rue eh ben on allait boire, il y avait des pompes sur le trottoir pour une dizaine de maison ben on allait glousser là, on allait glousser l'eau on se rafraichissait jamais entendu *** une chose aussi, je vais vous parler des **cafus** il y avait des **cafus** que j'ai connues ben croyez-moi ben il y en a encore 2 qu'elles sont en vie beh elles ont des rudes mérites hein, de notre époque c'était comme ça quand on allait au **bal** et que moi j'ai beaucoup de, j'ai 2 cousins et des qui sont pas **mariés**, qui n'ont jamais été **mariés** pour la simple raison que ils étaient **silicosés** de bonne heure alors quand on venait au **bal** et tout les jeunes filles entre elles parce que les jeunes filles elles étaient tout le temps à une table à part, elles discutaient entre elles, et ben quand un qui était **silicosé** qui allait la chercher à **danser** elle lui refusait la **danse** parce qu'elle confondait tuberculose avec **silicose** vous voyez ? Elles ont pas pu faire la comparaison alors elles disaient je veux pas **danser** avec lui, il a la tuberculose alors c'est pour ça moi j'ai eu beaucoup de copains qui n'ont pas pu se **marier** parce qu'ils étaient reconnus **silicosés** ouais les **cafus** avant elles étaient faut dire elles étaient aussi jolies que vous et puis pas de cosmétiques hein, nous on les reconnaissait par les yeux, elles avaient les yeux noircis par les **poussières** ça moi je les ai connues dans la rue elles passaient bras dessus, bras dessous avec les foulards, les coterons et tout ah ouais

Interviewer: et elles, elles se mariaient les cafus ?

Ancien mineur: les **cafus**, elles se **mariaient** ah bah oui hein mais nous on les draguait les **cafus**, nous, les mineurs ben la simple mais les **cafus** elles avaient la **retraite** comme un mineur, elles **travaillaient aux mines** mais quel travail, quel travail vous vous rendez compte ? parce que au **fond** il y a pas de water alors on faisait nos excréments sur les **convoyeurs** alors les excréments ils passaient avec le **charbon** et elles avaient pas de gants à l'époque, elles poussaient les excréments les, tout /pwa/ c'était des travail d'homme (sic), c'était jamais des travaux de femme et puis il y avait tout le temps des **surveillants** qui étaient là derrière eux en train /ahu ahu/ elles ont souffert les **cafus** hein, elles ont souffert moi j'en ai connu ben croyez-moi hein, c'était terrible pour elles

Interviewer: et au triage aussi c'étaient des femmes qui trouvaient les excréments ?

Ancien mineur: ah bah c'est au **triage** que je parle des tr- oui parce que le **charbon** brut il **remontait** au **jour** et était **transporté** sur des **convoyeurs** et elles elles **triaient**, elles mettaient les gros blocs de caillou d'un côté et le **charbon** de l'autre elles **triaient**, voilà on les appelait les **cafus**, un surnom donné, terme minier hein, on disait **trieuse** mais ça après ça s'est amélioré parce qu'ils ont pu aller dans les usines, dans les filatures, il y en a beaucoup qui sont partis qui ont quitté la mine qui sont partis dans les filatures à **Roubaix**

Interviewer: d'accord on coupe ? euh non, est-ce qu'on peut parler encore des cafus je ne sais pas comme c'était des femmes [inaudible]

Ancien mineur: ah au fond non moi j'ai pas connu, il faut dire qu'il faut pas exagérer la chose hein, non de notre temps euh j'ai pas connu de **cafus** au **fond** mais peut-être qu'il y en avait avant parce qu'ils **transportaient** le **charbon** sur le dos dans, comme on voit dans le film **Germinal** hein qu'elles se découvrent, c'est un **film**, c'est un **film**, il faut pas comparer un **film** comme **Germinal** avec la vérité c'est comme dans le **film** on voit euh comment qu'elle l'ampute de, elle lui coupait le zizi ben dans le **livre** de **Emile Zola** il y pas ce terme là émmaculé, ça n'existe pas, c'est encore un machin inventé enfin il faut pas en vouloir à **Claude Berri** hein parce que **Claude Berri** c'était un d'origine juive-polonaise vous le savez, c'était **Langman** et pourquoi il a changé de nom parce que son père a été emmené à **Auschwitz**, chez nous il y avait beaucoup de noms **Hauffman**, **Lichman** tout, d'origine juive-polonaise ah ouais ils étaient sous la peur hein, parce que quand ils passaient dans les rues avec *siffle*

Interviewer: et de votre temps les cafus elles étaient seulement en surface ?

Ancien mineur: au jour, elles **triaient** le **charbon** au **jour** et elles venaient se laver à la maison, bras dessus, bras dessus comme ça, ben il n'y avait pas de voitures hein, elles passaient 3, 4 comme ça, bras dessus, et elles chantaient même, quand elles avaient fini leur journée elles étaient contentes, elles chantaient un air euh voilà

Interviewer: et après quand elles étaient mariées elles s'arrêtaient elles s'occupaient des enfants

Ancien mineur: ah oui, ah oui, quand elles se **mariaient** après elle arrêtaient d'être **cafus**, elles arrêtaient oh oui parce qu'elles avaient déjà pas mal d'années de, il faut dire que nous les mineurs on se **mariait** de bonne heure pour la simple raison que vu le travail difficile, les risques donc et avoir la possibilité d'avoir une maison surtout, parce que nous là-bas où on restait au **Pas-de-Calais** on avait /fu/ qu'une piaule dans une **maison des mines** il y avait grand-mère, ma famille et une piaule alors on a posé une condition, d'accord, on veut être muté, on accepte d'être muté dans le **Nord** à une condition d'avoir une **maison** et ben croyez-moi quand on arrive de là-bas et qu'on a une **maison** à nous des **Houillères**, hein on dit une **maison** qu'elle est louée mais c'est un grand bonheur qu'on est indépendants, la vie elle change complètement

Interviewer: et il y avait beaucoup de, entre les jeunes filles et les jeunes hommes mineurs, ça se draguait beaucoup ?

Ancien mineur: le dragage existait toujours ah bah oui hein ça n'empêchait pas, la drague elle a toujours existé depuis que le monde est monde hein, on draguot les **cafus** et elles avaient des beaux yeux comme vous pas de cosmétique, pas de rouge à lèvres, ben il y en avait pas c'était comme ça

Interviewer: mais elles elles fréquentaient pas les cafus, les cafus

Ancien mineur: qui?

Interviewer: les cafus

Ancien mineur: ah non elle allait pas, non, au **bal** on se rencontrait seulement au **bal** et séparément parce que à l'époque c'était comme ça, les filles quand il y avait les **bals** volants avec la toile, les filles elles étaient tous sur un banc avec leurs parents la plupart du temps et les garçons de l'autre côté, ah attention pour aller danser il fallait demander l'autorisation au vieux vous permettez Monsieur, c'était comme ça, ah ouais, méfiant comme tout hein mais après quand c'étot la **ducasse** dans leur patelin ou quoi ils nous invitaient pour la tarte c'était la fête à la maison ah ouais complètement une autre vie mais je regrette pas, je préfère ce que j'avais avant que maintenant

Interviewer: pourquoi qu'est-ce qu'il y a maintenant

Ancien mineur: pourquoi, c'était plus humain, c'était, il y avait moins de jalousie, moins de jalousie, maintenant quand il y en a un qui a une brouette ben l'autre il voudrot avoir une auto euh tout de suite, il y a des esprits je sais pas, le monde, beaucoup de changements moi je dis euh la télévision elle a fait beaucoup de mal, la télévision qu'est-ce qu'elle a fait de mal et elle continue à faire mal surtout avec leurs appareils là qu'on voit que *** des bagarres là et tout et de notre temps non attention hein même au **cinéma**, c'était propre, très propre, les films avec **Gabin**, euh tout et après quand ça mis à l'érotisme les st- ça a démoli tout, ça démolit les ménages, ça démolit ah, c'est plus la même vie hein moi ben regardez-moi, on a eu nos noces d'or il y a 2 ans de ça, 52 ans, 52 ans de mariage […]

Annexe 2 : Extrait de retranscription de l'entretien effectué avec l'ancien mineur Martial à Givenchy-en-gohelle

Note : La transcription suivante contient le symbole # qui indique des superpositions dans les prises de parole.

Martial: pour extraire une gaillette il faut louquer un carbon, faut louquer

Amélie: louquer

Martial: louquer c'est frapper au marteau-piqueur, on louque, on frappe, louquer in patos chest frapper, donc le marteau-piqueur, il frappe, on louque

Amélie: et vous pouvez me parler un peu plus de

Henri: alors on va commencer ben ça c'était le coron ch'coron et puis ça ça c'était la maison d'ech l'ingénieur

Martial: ch'étot

Henri: ch'étot la maison de l'ingénieur donc il y avait dans chaque maison, il y avait des gardins, c'était des petits potagers, on appelait ça des gardins, bon qu'est-ce qu'il y avait euh bon ici on va arriver sur ch'carreau de fosse, le carreau de fosse donc ça ça représentait la maison de ch'garde et puis ici c'était les bureaux, les bureaux de la mine quoi, c'est là où est-ce qu'il y avait chl'ingénieur tous ces gens-là qui participaient

Amélie: il y avait pas un nom en particulier en ?

Henri: pff

Henri: ici

Martial: ah non ch'est les bureaux

Amélie: les bureaux

Henri: ch'bureaux bon là là là c'était ch'l'artour d'air comme on les appelait, là où ce que l'air remontait cha ch'étaient les puits, les différents types de puits ça c'était ch'lavabo et non la salle des pendus ils disaient, soit les mineurs ne disaient jamais ça soit qu'ils disaient ch'lavabo ou bien les bains douche

Amélie: d'accord

Henri: alors là c'était un atelier là c'était le transformateur, ch'transfo où est-ce qu'il y avait toute l'électricité là c'était la chaufferie el chaufferie voilà et puis après il y avait les différents hangars et là-bas il y avait ch'l'atelier avec le truc il y avait un truc, un vieux qui soudait à l'intérieur, ch'étot l'soudeux

Amélie: soudeux

Henri: donc là c'est les ingénieurs, fin pas les ingénieurs, les mineurs s'en vont ils s'en vont prindre el cache

Amélie: d'accord et dans la hiérarchie des mineurs donc il y avait

Henri: ben il y avait

Martial: surveillant surveillant porion chef porion ingénieur

Amélie: d'accord

Martial: hein et puis il y avait des petits chefs d'équipe aussi hein alors là vous le voyez partir al cache ch'est parce qu'ils vont déchindre, ils vont partir au front, parce que il y a l'déchinte puis après il y a l'armonte, quand c'est fini quoi

Amélie: et ensuite ?

Martial: ben là-bas derrière ben il y a

Amélie: après c'est à l'intérieur de la mine

Martial: ouais ben ça là c'est pareil c'est à part que ça, là c'est le puits de retour d'air ça c'est le puits d'entrée d'air parce que au fond del fosse pour avoir de l'oxygène il fallot qu'il y ait une circulation d'air donc euh et ch'trou d'intrée d'air il étot tout le temps plus profond parce il y avait des machins et li el retour d'air il étot moins profond donc l'air i descindot, il se réchauffait au fond et naturellement il repartait pour remonter mais il était aidé par de gros gros ventilateurs donc ce qui permettait d'avoir un cycle de renouvelé de l'air au fond et on voit d'ailleurs ch'est matérialisé par droite qui est là qui l'arloque quoi

Amélie: et au niveau des transports, donc il y avait les berlines et tout ça

Martial: des berlines des locos au début ch'étot des bidets, hein

Amélie: des bidets

Martial: après il y a des des structures c'est comme des berlines mais il y a pas de plaques c'est des "trucks" on appelle ça des trucks euh bon ch'est locos on avot des vélos des vélos aussi pour se déplacer dans certains cas et voilà quoi

Amélie: d'accord

Elise: est-ce que les trucs, ça vient des trucks en anglais ?

Martial: de quoi ?

Amélie: les trucs est-ce que ça vient de truck en anglais ?

Martial: ouais

Henri: ouais c'est ça

Elise: d'accord

Martial: alors là vous me posez des pièges parce que mi je parle pas du tout inglais hein

Elise: non c'est euhm pour savoir

Henri: ouais c'était anglais

Martial: ouais, ch'est comme briquet, soit disant je vous l'ai expliqué l'autre coup briquet ?

Virginie: vous pouvez le redire

Amélie: ah oui je veux bien !

Martial: ah ouais alors el briquet du mineur c'est le casse-croûte hein donc souvent ch'étot des tartines avec del graisse et l'graisse c'est du saindoux avec un' pomme ou un' orange alors on mettot cha dans un' bannette et on avait un bidon, un' gourde, un bidon donc le mien il est là je vais vous le faire voir ah oui de l'eau et du café voilà alors pourquoi briquet alors là ch'est il y a un grand, un respon- un classique chez Monsieur Guy Dubois, ch'est un' machin qui fait référence, il expliquait que briquet ça venot, si vous voulez en Angleterre trois cents ans avant nous, ils ont déjà exploité du carbon, cent ans après ils sont partis en Wallonie, deux cents ans après en 1700 et quelques ici, 1721 chez nous il y a eu la première fosse, ch'est des maîtres mineurs de Wallonie qui sont venus chez nous mais qu'ils avaient appris des mots anglais alors d'après Guy Dubois "break", on pause, on fait la pause et par déformation, à force de le répéter "break" breaker, on a breaké, ben on a briqué, alors les mineurs au départ c'était pas simplement la pause, ils ont dit ben on va en profiter pour manger un morceau donc "break", briquet, déformation voilà et alors beaucoup de gens confrontent cha parce qu'ils disent que briquet ça vient de Monsieur Raoul Briquet, un député socialiste de la région, en fait non, parce que si on prend Germinal qui est sorti en 1804 la première fois, Monsieur Briquet n'avait que 10 ans, or Zola emploie déjà ce ce mot hein d'accord ce mot de briquet mais la confusion vient du fait que Monsieur Briquet par la suite a été élu député et au député il a réussi à faire passer une loi parce que le temps de pause, du casse-croûte n'était pas compté dans les huit heures, c'était huit heures de travail plus vingt-cinq minutes ou une demi-heure de briquet et grâce à ce monsieur Briquet cette demi-heure, elle a été intégrée dans les huit heures de travail, alors de là la confusion, euh alors maintenant euh hein

Amélie: Et vous mettiez quoi dans le briquet alors ?

Martial: ben del graisse du saindoux, une pomme, des fois un bout de saucisson, du café et j'vais aller vous querre min bout'lot ah ouais

Amélie: le bout'lot

Martial: min bout'lot, bout'lot bien

Amélie: d'accord, d'accord

Martial: ch'est le mien il est bien bochelé hein ? boches

Amélie: #1 pourquoi il est bochelé alors #

Martial: #2 les boches ch'est cha # ch'est des cailloux cha, des cailloux

Amélie: #1 ils sont tombés ? #

Martial: #2 hein si vous avez # par exemple vous tenez ichi et ben nous on appelle ça un boche, vous avez un' boche donc euh il est bochelé hein voilà voilà et je l'ai jamais nettoyé et il sera jamais nettoyé

Amélie: d'accord et vous mettiez de l'eau dedans

Martial: de l'eau et du café ouais

Amélie: #1 du café #

Martial: #2 ah ouais oui oui # ouais ouais ouais, bon certains mettaient un peu autre-chose mais de toute façon c'était pas, ch'était pas tellement bon hein, le vin était, avait un bon goût mais fallait le supporter au fond

Amélie: et vous- et vous chiquiez euh ?

Martial: ouais alors j'vais vous faire vire "quelque chose" alors

Amélie: oui

Virginie: vous pouvez revenir par là, non mais là c'est bien mais juste pas trop loin

Martial: cha ch'est du roll

Amélie: c'est?

Martial: Ch'est du roll, du tabac chiqué, ch'est des feuilles ed toubac, nous on dit du toubac hein roulé, bon alors ici il est fort sec bien sûr hein et cha in coup' un morceau et in l'mettot dans chette blague à toubac, qu'on appelle nous les mineurs on appelle cha un /satio/, alors au départ c'était des vessies de porc, alors l'avantage d'un /satio/ ch'est que quand cha ch'est coupé ch'est d'dins voyez si je l'mets dans l'poche pis que ch'travaille que je me blesse pas par rapport à si j'avais une boîte, en plus cha maintient frais pi ravisé, quand vous prenez un' chique au fond il y a toujours des poussières, j'prinds un chique hein, clac il s'est déjà refermé tout seul voilà alors nous

Virginie: alors faudrait revenir devant

Martial: et euh in l'mettot al cave, al cave au fond, à la maison, au frais avec du rhum, on mettait du rhum dessus, oh ça, un bon' chique oh +[pron=yaya]

Virginie: vous pouvez refaire la démo avec le

Martial: ah ouais ? alors ch'prends un' chique hein hop non c'est brulant

Elise: et ça vous évoque quoi quand vous sortez la la chique

Martial: cha ch'roll là, ch'est un bâton de roll cha

Elise: les autres les autres éléments du briquet ça vous ça vous fait #1 penser à #

Martial: #2 alors cha cha # euh avant de faire briquet bon on buvait un coup pour se rincer hein, on se rinçait tout de même parce que mais voilà bon tout le monde ne chiquait pas hein alors la chique il y en a ils disent que c'est pour rimplacer el cigarette, y en a qui disent que lorsqu'on chique ben on respire par le nez donc les poils du nez retiennent un peu de poussière enfin il y a toute sorte euh mais vous dire ce qui est vrai ou ce qui est faux là-dedans, je sais pas, je vous dis ce que hein voilà

Virginie: et ce que disait Elise quand vous utilisez ces objets ça vous replonge dans

Martial: comment ?

Virginie: quand vous u-, quand vous avez ça dans la main ça vous replonge

Martial: ah ben je me vois au fond en train de chiquer hein

Virginie: vous vous voyez au

Martial: #1 fond ouais #

Virginie: #2 ah oui je me vois #

Martial: quand je parlais de qué qu'cosse du fond, ichi quel que soit le sujet, mi je suis au fond, mi j'suis dans min monde ah ch'est pour ça que des fois j'interviens ailleurs et je précise toujours au gens n'hésitez pas à m'arrêter

Virginie: oui

Martial: parce que j'vas commencer, j'vas partir et je suis parti, je suis dans min film et puis bah et je parle, je parle #1 sans m'rindre compte que ils comprenaient pas donc il faut #

Virginie: #2 vous êtes au fond de la mine, ouais # stop

Martial: stop

Virginie: #1 vous êtes où là au fond #

Martial: #2 bah ouais non mais # ch'est ch'est chacun sin truc hein

Virginie: et il faisait, il faisait vraiment fin la la lumière, il faisait noir

Martial: #1 mais il faisait sombre, il faisait ? #

Virginie: #2 ah ouais il faisait noir hein ah non il faisot fort noir #

Martial: noir noir noir noir noir d'ailleurs on avot un' lampe électrique hein ouais ouais

Virginie: #1 on n'y voyait pas grand chose #

Amélie: #2 comment comment vous appeliez c'était la lampe à euh le #

Martial: la lampe-chapeau

Amélie: la lampe-cha-

Martial: #1 la lampe-chapeau #

Amélie: #2 et en # en ch'ti?

Martial: eh ben c'est nous in disot l'lampe, l'lampe tout court on disot pas l'lampe au chapeau hein parce que en fait le chapeau ch'est une barrette

Amélie: et vous disiez pas, vous disiez pas capiau ?

Martial: hein ?

Amélie: capiau

Martial: ah capiau c'est un chapeau aussi #1 ch'est dehors, ch'est pas al fosse hein #

Virginie: #2 c'est pas la lampe ça #

Amélie: oui oui non le

Martial: #1 le chapeau #

Amélie: #2 un haut de forme #

Martial: ch'est un capiau hein d'accord ?

Virginie: mais votre casque alors ? #1 vous disiez quoi ? #

Martial: #2 j'ai pas de casque une barrette #

Virginie: une barrette

Martial: une barrette et pourquoi une barrette, parce que les première barrettes qu'ils ont eu, c'était des barrettes en cuir

Virginie: #1 oui #

Martial: #2 hein # donc ça ça faisait penser au curés italiens et en italien ils appelaient ça barretta donc les premiers mineurs ont eu des barrettes en cuir, ben barretta c'est devenu barrette et après c'est le fameux casque, c'est leur vrai nom casque hein

Virginie: #1 mais c'est resté barrette #

Martial: #2 mo- moderne avec la # lampe électrique et mais ch'est resté barrette dans notre langage à nous c'était barrette

Virginie: #1 d'accord #

Martial: #2 parce que # euh j'ai pour moi aller au fond, arriver au fond puis dire à un cop- à un mineur mets bien ton casque l'autre il va dire art beaux-arts hein cha ira mieux

Amélie: et vous connaissiez beaucoup de d'italiens #1 de marocains, des polonais ? #

Martial: #2 ah ouais ouais bien sûr j'ai des copains italiens ben à Saint en Gohelle il y en a, polonais ben je connais polonaise #

Amélie: et vous parliez euh tous en #1 chti ? #

Martial: #2 ben et vous savez # on on parlait tous in chti mais euh il y a un langage, enfin il y a des mots spécifiques qui me reviennent pas euh bon mi j'ai été en Pologne il y a trois quatre jours, j'sais pas parler polonais mais au fond del fosse avec les mineurs polonais, on a appris des mots, ben là-bas, on disot un mot par exemple euh pour dire merci *dziękuję ci* bah les gens ah, *do widzenia* au revoir oui mais je sais que des mots et je sais pas l'intégrer dans une phrase complète mais je sais quelques mots comme cha

Amélie: d'accord et quand elles sont arrivées en fait euh ces vagues d'immigration #1 étrangères, elles ont appris, d'accord #

Martial: #2 mes beaux-parents sont arrivés dans les années 1930, ch'étot la 2ème vague # d'accord et bon ils étaient parqués dans des camps au départ et tout ça a pas été facile pour eux hein déjà le pays, la langue

Amélie: donc ils ont appris le vocabulaire minier euh #1 euh sur, dans la fosse ? #

Martial: #2 ah ouais ouais vous inquiètez pas ouais ouais cha cha # cha va vite les échanges ça va vite, ça va plus vite qu'à l'école croyez-moi

Amélie: d'accord

Martial: hein par exemple ils entendaient des mineurs français dire *casser ch'gueule li le polonais*, ça veut dire attention je vais te casser la figure quoi mais bon ils arrivaient à se faire comprendre

Amélie: #1 d'accord et vous aviez des lexiques, quelque chose, des supports pour traduire ? #

Martial: #2 mais, ah non non non non ch'était # comme cha, ch'était à la bonne franquette entre nous #1 oh non non #

Virginie: #2 à l'oral #

Amélie: #1 c'était c'était de l'oral #

Martial: #2 ouais ouais #

Amélie: #1 c'était de l'oral, oui c'est beaucoup d'oral #

Martial: #2 beaucoup d'oral donc c'est pour ça que # nous des fois on prononce des mots quand on était du rendement, si le gars là-bas le prononçait ben tant pis on répète mal hein euh

Amélie: d'accord #1 d'accord donc il y avait beaucoup de #

Martial: #2 ouais ouais ouais ouais ouais puis après # à l'école, les échanges, tout cha bon puis mi d'min temps déjà il y avait déjà beaucoup plus d'échanges

Amélie: et au niveau du français du coup vous parliez pas du tout français au fond ou il y avait un mélange euh ?

Martial: Oh non on parle patois hein, on parlot on parlot patois parce que cha faisot cha faisot "cucul" de parler français hein au fond

Virginie: #1 Et le garde il parlait français, patois ? #

Martial: #2 ah non, ch'garde il parlait franç- euh # patois

Virginie: #1 ah, et par contre le directeur de la mine il parlait français lui ? #

Amélie: #2 tout le monde parlait patois #

Martial: euh il parlait français mais quand il était au fond avec certaines équipes ch'était un peu du paternalisme il faisait exprès de parler patois

Virginie: ouais c'était par euh, vous voyez je suis parmi vous quoi, je suis avec vous hein. Et les ingénieurs ?

Martial: bah il y en avait des bons mais

Virginie: il y en a qui parlaient un peu patois aussi quand même ?

Martial: ouais ouais il y en avait des bons, il y avait des crapules qui bon souvent ils, c'est arrivé qu'ils changeaient un carreau, bon ça c'était de sa faute hein

Virginie: et je vous ai, j'ai complétement oublié de vous parler c'est le médecin du travail

Martial: ah le médecin des Houillères

Virginie: #1 le médecin des Houillères #

Martial: #2 ah ben chaque fosse # chaque cité comme nous à Hersin Coupigny il y avait deux médecins des Houillères, chaque cité on devait aller à tel médecin pas à un autre

Virginie: ah oui mais ça c'est en dehors mais mais le médecin sur la mine, c'est-à-dire pour les ouvriers dans la mine, le médecin du travail #1 ça vous l'- #

Martial: #2 ah non ça # fallait euh, on passait à notre médecin traitant qui nous envoyait à l'hôpital ou à la clinique passer des radios #1 tout cha ch'était là-bas #

Virginie: #2 d'accord #

Martial: ch'était là-bas, ch'était pas ouais ouais, un peu comme le médecin conseil de sécurité sociale quoi

Virginie: ouais d'accord

Eva: et euhm pour la ducasse, vous avez parlé des tartes au sucre et d'une tarte à /bobords/

Martial: ah la tarte à gros bords, ah à gros bords ah parce qu'elle a des

Eva: #1 gros bords #

Martial: #2 ouais ben ch'est parce que # si vous voulez el tarte à gros bords c'est une tarte al crème

Eva: #1 ah d'accord #

Martial: #2 de crème donc elle # #1 donc elle est creuse et puis il y a des gros bords au bord pour dire que l'crème #

Eva: #2 ah oui donc elle #

Virginie: #1 pour retenir la crème #

Martial: #2 #

Eva: #1 retenir d'accord #

Martial: #2 hein voilà alors nous après #

Eva: ça donne faim

Virginie: ben ouais c'est les tartes au libouli ça #1 non #

Martial: #2 ah ben ouais ch'est cha, ch'est libouli #

Eva: les tartes au libouli #1 #

Virginie: #2 moi aussi je connais #

Eva: #1 la tarte à gros bords, moi je disais pas gros bords nous on dit libouli #

Virginie: #2 on m'en avait parlé #

Martial: ouais la tarte au libouli, tarte au ouais c'est la tarte avec, c'est vrai que

Virginie: bah voilà, faut que ça remonte bien #1 parce que c'est liquide avant que ça aille dans le four #

Eva: #2 pour que ce soit une belle, avant que ça, bah oui #

Virginie: #1 pour que ça prenne un peu #

Elise: #2 moi j'avais une autre # question, vous avez parlé un peu des polonais

Martial: #1 comment? #

Elise: #2 mais vous avez pas # vous avez parlé un petit peu des polonais

Martial: ouais

Elise: mais pas beaucoup des autres vagues et #1 des marocains des #

Martial: #2 ben disons des grosses vagues chez nous # euh au fond des mines, il y a eu 29 nationalités différentes mais les grosses nationalités c'est les polonais, les italiens, et puis les maghrébins

Virginie: vous avez travaillé vous même avec des maghrébins ?

Martial: #1 oui bien sûr, ouais #

Virginie: #2 ouais et puis # #1 des ouais il y avait des polonais des italiens #

Martial: #2 ouais bien sûr #

Elise: est-ce qu'il y avait des différences ?

Martial: des ?

Elise: des différences

Virginie: #1 cultu- des différence culturelles ouais racontez #

Martial: #2 au départ ouais mais après c'était euh au départ ouais mais # mais euh au fond un mineur ch'est un mineur, on voit pas qui il est, on voit le mineur, s'il y avait un problème, s'il était blessé si, on partait le chercher et eux pareil, par contre au jour des fois al ducasse ben si il est trop près d'une copine et comme chez nous il y a des cités polonaises, des cités italiennes, ben fallot pas trop fallait le savoir quoi hein ben c'est normal hein, chacun mais après c'étot fini c'est pas bon après

Virginie: #1 mais #

Martial: #2 #

Virginie: et disons que alors au fond on a compris la solidarité et au jour dans le coron c'était plus par fin, par respect de chaque origine ou c'était presque du racisme, c'était comment #1 en fait? #

Martial: #2 non il y avot # pas de racisme

Virginie: #1 c'était pas du racisme, non #

Martial: #2 non il y avot pas de racisme comme on le # connaît maintenant il y avait pas de racisme non non, c'était le, le mineur il reste le mineur c'est tout, bon bien sûr euh les polonais ils avaient leurs

chorales, leurs équipes de foot, leurs églises polonaises, leurs chapelles polonaises, cha ch'est vrai mais enfin bon après tout hein c'est pas, un peu comme les musulmans ils ont un

Virginie: #1 une mosquée #

Martial: #2 un # un machin, bon ben c'est tout voilà

Virginie: ouais, c'était dans le respect

Martial: voilà, on respecte l'autre c'est tout et puis c'est

Elise: les enfants ils jouaient ensemble, ils allaient à l'école ensemble ?

Martial: ah ouais ouais ouais bien sûr, d'ailleurs il y a beaucoup de Polonais qui ont eu des gros diplômes et puis tout et qui s'en sont sortis que ce soit par le sport ou par les études hein ouais

Virginie: et est-ce qu'on vous a interdit un moment de jouer par exemple s'il y avait des familles je sais pas qu'est-ce que vous avez vu arriver vous des maghrébins, des italiens ?

Martial: euh surtout des ma- des maghrébins

Virginie: voilà alors par exemple est-ce qu'il y a des familles de maghrébins qui arrivaient, est-ce que vous avez eu int-, au début interdiction de jouer avec les enfants maghrébins ou vous avez pu tout de suite vous mêler à eux

Martial: non, non non non non chez nous non

Virginie: chez vous non

Martial: ben ils sont arrivés et je travaillos déjà al fosse à 15, 16 ans, ils sont arrivés al fosse c'est vrai qu'au départ ch'est un peu drôle quoi, bon ils avaient pas l'habitude

Virginie: dépaysement bien sûr

Virginie: oui il y avait pas ouhla attention

Eva: cloisonnement

Virginie: #1 cloi- voilà #

Martial: #2 de toute façon # euh pfff dans la cité minière ça n'existait pas, s'il y avait un problème il était réglé tout de suite, viens derrière, on va s'expliquer, c'était comme ça et puis c'était fini, d'ailleurs combien de fois l'épouse elle me dit parce que j'avais une belle renommée hein elle me dit tes meilleurs copains c'est parce que tu t'es bagarré avec eux, on finissait toujours par aller boire un coup, c'était

Virginie: #1 d'abord faut et après glouglou #

Martial: #2 ah ouais ouais ouais ouais j'avais # j'avais de la réputation de ce côté-là ouais ohlala

Virginie: bah heureusement que vous nous tapez pas hein

Martial: non, non, non j'ai jamais été comme cha quoi mais je veux dire, euh dans un bal je vois un copain qui était avec s'copine tout ça et moi j'étos avec un' femme bien-sûr, fin copine de l'époque on dansait tout ça puis je voyais le copain qui été embêté, je passe à côté je dis quo qu'il y a là, ben il dit c'l'autre là-bas, je dis qu'est-ce qu'il y a, l'autre là-bas, qu'est-ce qu'il y a, il m'emmerde il dit j'ose pas, bon d'accord, on danse, ben quand il dansait, clac coup d'épaule, se retournait boum, c'est fin, on discutait pas, on tapait, ah ouais ouais ouais

Virginie: #1 réglement de compte #

Martial: #2 j'ai # j'ai été mis dehors, j'avais plus le droit de rentrer à la salle des fêtes alors les vieux quand ils ont, parce que les vieux ils me disaient toujours ti tu vas aller au 100 mille briques, les 100 mille briques c'était le surnom de la prison de Béthune

Virginie: ahlala

Martial: ti te va y aller et quand ils ont su que j'étais en prison,

Virginie: #1 ils ont cru non mais ouais mais non non gardien #

Martial: #2 ah ils ont cru que vous étiez vraiment derrière les barreaux ? #

Virginie: ah oui après #1 ah bon #

Martial: #2 comment que t'as fait comment qu't'as fait ? #

Martial: ils prennent n'importe qui

Valérie: ben oui

Martial: ah ouais bon ben cha cha fait partie de l'ensemble si j'avais pas été mineur, peut-être que j'aurais pas été comme ça quoi euh je ne sais pas

Virginie: ah ouais

Amélie: #1 alors j'ai juste une petite question sur la maquette, alors ça c'est quoi ? #

Virginie: #2 vas-y, vas-y #

Martial: cha ch'est chcatiau d'ieaux le château d'eau

Amélie: le château d'eau et là aussi ?

Martial: euh ouais, là ch'étot l'puits de fosse, là ch'est pour la gare quoi hein pour les locos à vapeur

Amélie: d'accord et là le terril comment vous avez #1 appelé #

Martial: #2 là ch'est un terril plat cha # #1 hein parce que #

Amélie: #2 d'accord et en #

Martial: j'ai pas pu faire un conique parce que pour la maquette je l'ai prêtée

Amélie: et en patois il y a un nom #1 spécial ? #

Martial: #2 non # #1 un ch'terril mais nous in met pas in met pas de L hein on dit /terri/ on dit pas /terril/ pas /terril/ #

Amélie: #2 mais quand il est, d'accord #

Martial: faut pas le dire hein /terri/

Amélie: et vu qu'il est plat, il y a pas un nom spécial

Martial: non non il y a soit soit soit soit terril plat soit terril conique

Amélie: d'accord, ok et au niveau du charbon, la qualité du charbon il y avait des termes spécifiques?

Martial: euh mais là ch'étot des termes de géologues, nous c'est du carbon soit du carbon par exemple il y avot du Saint-Edouard, du Sainte-Catherine, du Saint Edouard Cochon, du Saint Edouard Levant, du Bienvenue inférieur, du Bienvenue supérieur mais pour nous ch'étot du carbon, la différence qu'il y avait entre les deux on savot que Bienvenue il prenot bien, il prenot bien feu avec du bos, avec du bois on mettait, que par rapport à du Saint Leonard cha on le savait parce que mais après vous dire les différences de formation tout non non on connaît pas hein

Virginie: #1 qu'il soit gras ou maigre ? #

Martial: #2 ch'est les géologues # ch'est les géologues là, ouais il y a du gras, il y a du maigre tout cha mais là faut voir des spécialistes hein

Amélie: et les berlines qui étaient vides ou pleines, il y avait des termes spécifiques #1 aussi non, les ? #

Martial: #2 alors euh un' berline vide c'est un barrou #

Amélie: barrou

Martial: pleine ch'est une balle

Amélie: #1 ok il y avait que ces #

Martial: #2 alors la balle # soit disant que cha vient des premiers mineurs qui étaient dans l'coin ichi ch'étaient des fermiers et dans les fermes il y avaient des balles de paille, des gros de balle qu'ils appelaient des balles les fermiers

Amélie: #1 d’accord #

Martial: #2 donc quand ils ont vu les berlines comme ça # c’était ils ont appelé cha l’balle

Amélie: d’accord ça je peux les garder ?

Martial: ben ch’est pour vous, on va aller là-bas peut-être il y a encore des autres choses à vir

Annexe 3 : Hiérarchie des classes issue de la version 5.0.4 du CIDOC CRM, instances relevées du patrimoine minier et note sur les classes selectionnées dans le projet TECTONIQ

Les classes du modèle sont identifiées par un identifiant Id (première colonne) et un nom de classe (3ème colonne). La deuxième colonne du tableau correspond aux classes sélectionnées pour décrire le patrimoine textile du projet TECTONIQ (Text. : ok), puis celles sélectionnées pour décrire le patrimoine minier dans le cadre du projet MémoMines (Mines : ok). La quatrième colonne correspond à une explication courte de la classe et fournit des exemples d'instances issues des corpus du patrimoine minier. Les classes polyhiérarchiques ne sont décrites qu'une fois (mention « déjà décrit » dans la quatrième colonne).

Id	Text. /Mines	Classes							Instances issues du patrimoine minier
E1	ok/ok	CRM Entity							Regroupe toutes les classes et instances
E2	ok/ok	-	Temporal Entity						Cette classe est trop générique
E3	ok/ok	-	-	Condition State					Cette classe comprend les états des objets caractérisés par une certaine condition sur une période de temps. Ex : L'état fosse n°4 des mines de Lens « en ruine » pendant la guerre 1914-1918
E4	ok/ok	-	-	Period					Ex : Période d'activité de la fosse Ledoux de 1901 à 1988.
E5	ok/ok	-	-	-	Event				Ex : La Seconde Guerre mondiale, catastrophe de Courrières ou la Grève des mineurs du Nord–Pas-de-Calais (1941)
E7	ok/ok	-	-	-	-	Activity			Ex : Construction des corons
E8	-/-	-	-	-	-	-	Acquisition Event		Ex : Le don d'un objet au musée de Lewarde (mais nous avons peu de renseignements là-dessus)
E9	-/ok	-	-	-	-	-	Move		Ex : Déplacement d'un objet (mais nous avons peu de renseignements là-dessus)
E10	-/-	-	-	-	-	-	Transfer of Custody		Ex : Changement de détenteur d'un objet (mais nous avons peu de renseignements là-dessus)
E11	-/ok	-	-	-	-	-	Modification		Ex : Modification d'un objet matériel. Réhabilitation des cités minières, rénovation de 23 000 logements, de Béthune à Valenciennes à partir de 2013
E12	-/-	-	-	-	-	-	-	Production	Production d'un nouvel item avec succès Ex : chevalement du puits Nord (n°2) de la Fosse du Sarteau à Fresnes-sur-Escaut Ex : Acte de peindre
E79	-/-	-	-	-	-	-	-	Part Addition	Ajout de quelque chose sur une production… Ex : ajout d'une pièce sur une machine dans le but de la faire remarcher (mais nous avons peu de renseignements là-dessus)
E80	-/-	-	-	-	-	-	-	Part Removal	Contraire de E79 Ex : suppression d'un descripteur dans ThesoMines

E13	-/-	-	-	-	-	-	Attribute Assignment			Mesure d'un objet de musée – affectation d'un attribut à un objet par quelqu'un Ex : (mais nous avons peu de renseignements là-dessus)
E14	-/-	-	-	-	-	-	-	Condition Assessment		Cette classe décrit l'action consistant à formuler une expertise de l'état de conservation d'un objet sur une période donnée, soit par examen direct, soit par mesurage, soit par étude de sources historiques. Ex : pas d'exemple
E15	-/-	-	-	-	-	-	-	Identifier Assignment		Attribution d'un identifiant à un objet lors d'un inventaire Ex : attribution d'un DOI à un article scientifique sur le bassin minier
E16	-/-	-	-	-	-	-	-	Measurement		Mesure d'un objet – date Ex : mesure de la teneur de grisou dans les tailles avec un grisoumètre
E17	-/-	-	-	-	-	-	-	Type Assignment		Attribution d'un type. Ex : indexation d'une œuvre
E65	ok/ok	-	-	-	-	-	Creation			Création d'un poème, d'une œuvre, une peinture Ex : écriture de Germinal par Emile Zola, écriture du poème *L'Caillau à fleur* en 1907 par Jules Mousseron
E83	ok/ok	-	-	-	-	-	-	Type Creation		Création d'un type (d'une classe) Ex : création du descripteur « matériel-aérage » dans ThesoMines le 9 février 2020
E66	-/-	-	-	-	-	-	Formation			Ex : Formation de l'association des amis de la mine
E85	-/-	-	-	-	-	-	Joining			Personne devenant membre d'une association Ex : René Lukasiewicz, Président de l'association des Amis de Germinal Nouvelle directrice-conservatrice : Amy Benadiba à Lewarde le 1er mars 2016
E86	-/-	-	-	-	-	-	Leaving			Départ d'un acteur d'un groupe Ex : départ de Jean-Pierre Kucheida de l'EPINORPA
E87	-/-	-	-	-	-	-	Curation Activity			Activité de conservation d'une collection (peu de renseignements là-dessus) Ex : conservation de la collection des Archives départementales du Nord
E63	ok/ok	-	-	-	-	Beginning of Existence				Début d'existence Ex : début des HBNPC
E67	-/-	-	-	-	-	-	Birth			Naissance d'une personne Ex : naissance d'un ancien mineur
E81	-/-	-	-	-	-	-	Transformation			Ex : Transformation d'une fosse en un musée
E12	-/-	-	-	-	-	-	Production			Déjà décrit
E65		-	-	-	-	-	Creation			Déjà décrit
E83		-	-	-	-	-	-	Type Creation		Déjà décrit
E66		-	-	-	-	-	Formation			Déjà décrit
E64	-/ok	-	-	-	-	End of Existence				Fin d'item physique, d'un groupe, d'un être vivant Ex : un ancien mineur, Emile Basly en 1928 (plus général pour la mort d'un être E69 Death)

E6	-/ok	-	-	-	-	-	Destruction				Destruction Ex : destruction d'une ville, d'une fosse, de bâtiments lors de la deuxième guerre mondiale (intentionnel donc aussi *E7 Activity*) Non intentionnel Ex : catastrophe de Courrières
E68	-/-	-	-	-	-	-	Dissolution				Fin d'une association, d'un groupe de personne Ex : fin du CERCHAR
E69	-/ok	-	-	-	-	-	Death				Personne tuée – aussi *E7 Activity* / seulement la mort des humains Ex : Les 42 personnes mortes lors de la catastrophe de Liévin du 27 décembre 1974.
E81		-	-	-	-	-	Transformation				Déjà décrit
E77	ok/ok	-	Persistent Item								Quelque chose qui existe (soit un E39 actor, soit E70 une chose) Ex : Jean-François Caron, le musée de Lewarde, un terril
E70	ok/ok	-	-	Thing							Une chose, classe très générique Ex : un cheval, un outil, une personne, une église, une notice,
E72	ok/-	-	-	-	Legal Object						Item matériels ou immatériels sur lesquels un E30 peut être appliqué Ex : définitions du dictionnaire de Turpin 2004
E18	ok/ok	-	-	-	-	Physical Thing					Entité matérielle fabriquée ou naturelle avec ses caractéristiques Ex : E19, E20, E21, E22, E84, E24…
E19	ok/ok	-	-	-	-	-	Physical Object				Objet matériel, un élément complet plutôt que des parties Ex : un outil (un pic)
E20	-/ok	-	-	-	-	-	-	Biological Object			Objet biologique : un être vivant ou des excréments… Ex : un cheval à la mine
E21	-/ok	-	-	-	-	-	-	-	Person		Un ancien mineur, un acteur, un artiste…une personne Ex :
E22	ok/ok	-	-	-	-	-	-	Man-Made Object			Un objet fabriqué résultat d'actions d'ordre technique, fabriqué volontairement par un humain Ex : la machine d'extraction d'Oignies, le boutelot de Martial, la maquette exposée à Arenberg « puit minier le Voreux du film Germinal de C. Berri »
E84	ok/ok	-	-	-	-	-	-	-	Information Carrier		Un objet qui contient une information Ex : une affiche, une pancarte, un livre, la notice d'utilisation d'une machine Un objet E22 qui est utilisé pour supporter cette information (un CD vierge est un E22 sans être un E84) Ex : le journal de la Voix du Nord du 9 janvier 2001
E24	ok/ok	-	-	-	-	-	Physical Man-Made Thing				Des objets qui sont fabriqués pour une utilité Ex : un outil, une passerelle (peut être un E22, un E25 ou E78), une collection muséale, un terril
E22		-	-	-	-	-	-	Man-Made Object			Déjà décrit
E84		-	-	-	-	-	-	-	Information Carrier		Déjà décrit

E25	-/ok	-	-	-	-	-	-	Man-Made Feature	Les caractéristiques fabriquées sont les caractéristiques matérielles qui résultent de l'intervention humaine. Ex : Une galerie de mine, un terril (il y a une différence avec E26 Physical Feature, c'est l'intervention humaine) La différence avec E26 Physical Feature est que E25 est le résultat d'une intervention humaine. Ex : Les galeries souterraines du site de Blegny-Mines, le terril sabatier de Raismes.
E78	-/-	-	-	-	-	-	-	Collection	Collection de musée Ex : la collection de fossiles houillers du Centre Historique Minier de Lewarde
E26	ok/ok	-	-	-	-	-	Physical Feature		Cette classe comprend les caractéristiques identifiables qui sont physiquement attachées de manière intégrale à des objets physiques particuliers. Ex : Les galeries souterraines du site de Blegny-Mines, le puits n°9 d'Oignies, (avec ou non l'intervention humaine)
E27	-/ok	-	-	-	-	-	-	Site	Un site est un lieu distinct qu'il est possible de représenter au moyen d'un objet iconographique, tel qu'une photographie, un tableau ou une carte. Ex : les sites inscrits à l'UNESCO, le site minier du 9-9BIS à Oignies.
E25		-	-	-	-	-	-	Man-Made Feature	Déjà décrit
E90	ok/ok	-	-	-	-	Symbolic Object			Objet symbolique : Cette classe comprend les symboles identifiables et toute agrégation de symboles, tels que les caractères, les identificateurs, les panneaux de signalisation, les emblèmes, les textes, les ensembles de données, les images, les partitions musicales, les objets multimédias, les codes de programmes informatiques ou les formules mathématiques qui ont une structure objectivement reconnaissable et qui sont documentés comme des unités uniques. Ex : les paroles de *tout en haut de ch'terril* de Renaud, le titre de cette même chanson, une adresse avec des coordonnées spatiales
E73	ok/ok	-	-	-	-	-	Information Object		Cette classe comprend les articles immatériels identifiables, tels que les poèmes, les blagues, les ensembles de données, les images, les textes, les objets multimédias, les prescriptions procédurales, les codes de programmes informatiques, les algorithmes ou les formules mathématiques, qui ont une structure objectivement reconnaissable et qui sont documentés comme unités individuelles. Ex : les poèmes de Jules Mousseron
E29	ok/-	-	-	-	-	-	-	Design or Procedure	Des plans, des schémas... des procédures. Ex : le contenu du Plan

											de Prévention des Risques Miniers du « Lensois »
E31	ok/ok	-	-	-	-	-	-	Document			Cette classe englobe tout ce qui apporte un témoignage sur la réalité, que ce soit intentionnellement ou de manière fortuite. Il peut s'agir d'un texte, de graphique, d'images, de son, de vidéo. Ex : des livres d'histoire, des cartes géographiques du bassin minier établie par l'UNESCO, une photographie de l'harmonie des mineurs de Wingles en 1938
E32	ok/-	-	-	-	-	-	-	-	Authority Document		Encyclopédies, thésaurus, listes d'autorité qui définissent une terminologie ou des systèmes conceptuels pour une utilisation cohérente Ex : ThesoMines, TerminoMines, OntoMines
E33	ok/ok	-	-	-	-	-	-	Linguistic Object			Du texte. Des textes écrits ou oraux. Ex : une retranscription d'un entretien avec un ancien mineur ou un article de presse de la Voix du Nord
E34	-/-	-	-	-	-	-	-	-	Inscription		Une inscription sur un *E24 Physical Man-Made Thing*. Une légende sous une photographie, un petit mot, une signature, une marque… Ex : « casque obligatoire » sur un panneau signalétique à la fosse 2 d'Oignies
E35	ok/ok	-	-	-	-	-	-	-	Title		Titre d'une œuvre Ex : *Autour des terrils* (titre d'un recueil de Jules Mousseron), *tout en haut de ch'terril* (titre d'une chanson de Renaud)
E36	ok/-	-	-	-	-	-	-	Visual Item			Un item visuel, le symbole d'une marque, quelque chose qui se reconnait facilement Ex : le logo de la Mission Bassin minier, du projet MémoMines
E37	-/-	-	-	-	-	-	-	-	Mark		Une signature, un symbole, des signes ou de courts textes qui annonce le créateur, le propriétaire, des dédicaces, le but Ex : une signature sur une vieille carte postale
E34	-/-	-	-	-	-	-	-	-	-	Inscription	Une inscription sur E24. Une légende, un petit mot, une signature, une marque… Ex :
E38	-/-	-	-	-	-	-	-	-	Image		L'image reconnaissable sur une multitude de surface Ex : les terrils jumeaux de Loos-en-Gohelle sur une carte postale
E41	ok/ok	-	-	-	-	-	Appellation				Les différentes dénominations linguistiques des choses matérielles ou immatérielles (concepts) Ex : la « cité des électriciens » pour cité minière, « musée de Lewarde » pour musée ou lieu, « garde des mines » pour surveillant, « jupon d'fosse » pour vêtement.
E42	ok/-	-	-	-	-	-	-	Identifier			Identifiant des objets (suite de chiffres et lettres…)

										Ex : ISBN 978-2-915398-20-5 du livre Les Gueules Noires, 2020 d'Emile Morel.
E44	ok/-	-	-	-	-	-	-	Place Appellation		Une adresse, des coordonnées géographiques, des noms de lieu etc. Ex :
E45	-/ok	-	-	-	-	-	-	-	Address	Adresse Ex : Fosse Delloye, Rue d'Erchin, 59287 Lewarde
E46	-/-	-	-	-	-	-	-	-	Section Definition	Désignation de parties d'un objet matériel Ex : le puits de mine fait partie du carreau de fosse
E47	-/ok	-	-	-	-	-	-	-	Spatial Coordinates	Coordonnées spatiales Ex : 50.42676231211688, 2.779834024319626 coordonnées du chevalement de La fosse N3bis à Liévin
E48	ok/ok	-	-	-	-	-	-	-	Place Name	Nom d'un lieu : ville, département Ex : Arenberg, Nord–Pas-de-Calais, bassin minier du NPC
E49	ok/-	-	-	-	-	-	-	Time Appellation		Appellation d'une période Ex : première guerre mondiale
E50	ok/ok	-	-	-	-	-	-	-	Date	Une date Ex : 05-08-1990, 1992
E75	ok/-	-	-	-	-	-	-	Conceptual Object Appellation		Appellation d'un objet selon : toutes les appellations spécifiques aux produits intellectuels ou aux modèles standardisés. ISBN, ISO… Ex : ISBN 978-2-915398-20-5 du livre Les Gueules Noires, 2020 d'Emile Morel.
E82	-/-	-	-	-	-	-	-	Actor Appellation		Nom d'acteur (organisme, personne) Ex : Mission Bassin minier, association Bassin Minier uni, association Germinal
E51	ok/ok	-	-	-	-	-	-	Contact Point		Adresse, mail etc, numéro de tel Ex : Les Amis de la mine, Mairie de Oignies, Hôtel de ville Place de la IV^{e} République 62590 Oignies Tél. 03 21 74 80 50 (la mairie) / 03 21 37 74 80 (le musée) Mail richard.szymczak@numericable.fr
E45		-	-	-	-	-	-	-	Address	Déjà décrit
E35		-	-	-	-	-	-	Title		Déjà décrit
E71	ok/ok	-	-	-	Man-Made Thing					Cette classe comprend les articles fabriqués par l'homme, distincts et identifiables, qui sont documentés en tant qu'unités individuelles. Ex : -un poème de Jules Mousseron - peintures de Jonas au musée de la mine à Anzin -la chanson *Les corons* de Pierre Bachelet
E24	ok/ok	-	-	-	-	Physical Man-Made Thing				Des objets qui sont fabriqués pour une utilité Ex : un outil, une passerelle (peut être un E22, un E25 ou E78)
E22		-	-	-	-	-	Man-Made Object			Déjà décrit
E84		-	-	-	-	-	-	Information Carrier		Déjà décrit
E25		-	-	-	-	-	Man-Made Feature			Déjà décrit

E78		-	-	-	-	-	Collection					Déjà décrit
E28	ok/ok	-	-	-	-	Conceptual Object						Un objet conceptuel : un fait connu et transmis comme la mémoire de quelque chose, une définition d'un concept… Ex : la définition du briquet dans TerminoMines
E90		-	-	-	-	-	Symbolic Object					Déjà décrit
E73		-	-	-	-	-	-	Information Object				Déjà décrit
E29		-	-	-	-	-	-	-	Design or Procedure			Déjà décrit
E31		-	-	-	-	-	-	-	Document			Déjà décrit
E32		-	-	-	-	-	-	-	-	Authority Document		Déjà décrit
E33		-	-	-	-	-	-	-	Linguistic Object			Déjà décrit
E34		-	-	-	-	-	-	-	-	Inscription		Déjà décrit
E35		-	-	-	-	-	-	-	-	Title		Déjà décrit
E36		-	-	-	-	-	-	-	Visual Item			Déjà décrit
E37		-	-	-	-		-	-	-	Mark		Déjà décrit
E34		-	-	-	-	-	-	-	-	-	Inscription	Déjà décrit
E38		-	-	-	-	-	-	-	-	Image		Déjà décrit
E41		-	-	-	-	-	-	Appellation				Déjà décrit
E42		-	-	-	-	-	-	-	Identifier			Déjà décrit
E44		-	-	-	-	-	-	-	Place Appellation			Déjà décrit
E45		-	-	-	-	-	-	-	-	Address		Déjà décrit
E46		-	-	-	-	-	-	-	-	Section Definition		Déjà décrit
E47		-	-	-	-	-	-	-	-	Spatial Coordinates		Déjà décrit
E48		-	-	-	-	-	-	-	-	Place Name		Déjà décrit
E49		-	-	-	-	-	-	Time Appellation				Déjà décrit
E50		-	-	-	-	-	-	-	-	Date		Déjà décrit
E75		-	-	-	-	-	-	-	Conceptual Object Appellation			Déjà décrit
E82		-	-	-	-	-	-	-	Actor Appellation			Déjà décrit
E51		-	-	-	-	-	-	-	Contact Point			Déjà décrit
E45		-	-	-	-	-	-	-	-	Address		Déjà décrit
E35		-	-	-	-	-	-	-	Title			Déjà décrit
E89		-	-	-	-	-	Propositional Object					
E73		-	-	-	-	-	-	Information Object				Déjà décrit
E29		-	-	-	-	-	-	-	Design or Procedure			Déjà décrit
E31		-	-	-	-	-	-	-	Document			Déjà décrit
E32		-	-	-	-	-	-	-	-	Authority Document		Déjà décrit
E33		-	-	-	-	-	-	-	Linguistic Object			Déjà décrit
E34		-	-	-	-	-	-	-	-	Inscription		Déjà décrit
E35		-	-	-	-	-	-	-	-	Title		Déjà décrit
E36		-	-	-	-	-	-	-	Visual Item			Déjà décrit
E37		-	-	-	-	-	-	-	-	Mark		Déjà décrit
E34		-	-	-	-	-	-	-	-	-	Inscription	Déjà décrit
E38		-	-	-	-	-	-	-	-	Image		Déjà décrit

E30	ok/ok	-	-	-	-	-	-	Right	Droits qu'une personne ou un groupe a sur quelque chose, sur une œuvre, une propriété Ex : tel propriétaire à les droits sur telle habitation
E55	ok/ok	-	-	-	-	-	Type		Poids, longueur, profondeur constituent des types de la classe *E54 Dimension*. Portrait, croquis, animation pourraient constituer des types de E38 Image. Écrit, oral pourraient constituer des types de la classe *E56 Language*. Excellent, bon, mauvais pourraient constituer des types de la classe *E22 Man-Made Object*. La classe permet de pointer vers un descripteur de thésaurus (ici ThesoMines). Ex : bâti civil dans ThesoMines
E56	-/ok	-	-	-	-	-	-	Language	Langue Ex : picard, français, italien, polonais (la norme en code ISO est recommandée)
E57	ok/ok	-	-	-	-	-	-	Material	Matériau des objets Ex : soutènement en bois ou en fer
E58	-/-	-	-	-	-	-	-	Measurement Unit	Unité de mesure Ex : poignie d'pouche (une quinzaine de centimètre), mètre cube, degré
E39	ok/ok	-	-	Actor					Agent (organisme, personne…) qui joue un rôle dans une activité. Ex : Association des mineurs marocains du Nord
E74	ok/ok	-	-	-	Group				Groupe de personnes Ex : les anciens mineurs du Bassin minier du Nord–Pas-de-Calais
E40	ok/-	-	-	-	-	Legal Body			Collectivité (institution qui a du pouvoir pour faire des choses) Ex : musée de Lewarde, Mission Bassin minier, BRGM
E21	-/ok	-	-	-	Person				Une personne, un individu Ex : un ancien mineur, le/la maire de Lens, le fils d'un ingénieur
E52	ok/ok	-	Time-Span						Tranche chronologique Ex : -307 et -299 millions d'années : période pendant laquelle le bassin houiller commence sa formation avec des sédiments fluvio-lacustres.
E53	ok/ok	-	Place						Lieu Ex : Mazingarde, Département du Pas-de-Calais
E54	ok/ok	-	Dimension						Dimension Ex : la teneur en grisou de 10m^3 dans une taille
E59	-/-	Primitive Value							Valeur primitive Ex : 3, 1994-1997, « don du monsieur qui est passé hier après-midi au musée »
E60	-/-	-	Number						Nombre Ex : 3
E61	-/-	-	Time Primitive						Primitive temporelle Ex : 1994 – 1997
E62	-/-	-	String						Chaîne de texte libre, bitmaps, graphiques vectoriels… Ex : une citation issue d'un article accompagnée de sa note

Annexe 4 : Classes et propriétés sélectionnées pour représenter le patrimoine minier et liens vers schémas d'aide à l'instanciation

E1_CRM Entity

Cette classe est très générique et est la base de tous les schémas.

https://cmapscloud.ihmc.us:443/rid=1WL1Q2XVV-XMXZMG-252LPSV
Liste des descripteurs histoire sociale, histoire politique

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*

E7_Activity

Pour tous les schémas *E7 Activity*, une propriété *P15 was influenced by* a été ajoutée suite à la première phase d'instanciation pour noter l'influence d'un *E7 Activity* sur un autre *E7 Activity*, ou noter la cause de l'un sur l'autre.

E7_Activity_Schéma 11
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs OPERATION abattage, chassage, découronnement, forage, minage, rauchage, exhaure, dénoyage, drainage, pompage, aérage, ventilation, traitement-minerai, broyage, concassage, criblage, lavage, opération-triage, transport[171], chargement, déblayage, déblocage, desserte, marinage, raclage, reprise, ripage, soutirage, opération-stockage, prospection, fonçage, sondage, traçage *Descripteurs ajoutés* chronométrage, pointage

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P16 used specific object* : *E70 Thing*
E7 Activity : *P19 was intented use of* : *E71 Man-Made Thing*
E7 Activity : *P32 used general technique* : *E55 Type*
E7 Activity : *P117 occurs during* : *E2 Temporal Entity*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

[171] *Transport* est une instance de *E9 Move*.

E7_Activity_Schéma 12
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs ACTIVITE métier[172], métier-abattage, abatteur, boutefeu, piqueur, métier-administratif, métier-aérage, métier-cadre[173], chef-siège, chef-taille, chef-trait, directeur, géologue, géomètre, ingénieur, médecin, porion, métier-électricité, métier-exhaure, métier-prospection, bouveleur, foreur, prospecteur, traceur, métier-sécurité, gazier, lampiste, chef lampiste, about, pompier-mineur, sauveteur[174], métier, boiseur, foudroyeur, remblayeur, métier-transport, accrocheur, aiguilleur, chargeur, conducteur, conducteur-cheval, dispatcheur, encageur, freinteur, herscheur, machiniste, métier, stockage, moulineur, poseur-voie, reculeur, métier-triage, métier-soutènement, métier-administration, métier-stockage, accident-travail, accident-transport, blessure, chute *Descripteurs ajoutés* emploi, licenciement, recrutement, électricien, électromécanicien, mécanicien, chef-porion, chef-équipe, surveillant, métier-creusement, métier-entretien, laveur, technicien, contrôleur-empoussiérage, sanction

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P11 had participant* : *E39 Actor*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P16 used specific object* : *E70 Thing*
E7 Activity : *P19 was intented use of* : *E71 Man-Made Thing*
E7 Activity : *P32 used general technique* : *E55 Type*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

E7_Activity_Schéma 13
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs Temps-travail, poste, poste-après-midi, poste-matin, poste-nuit, repos, temps-repas, maladie professionnelle, silicose[175] *Descripteurs ajoutés* heure supplémentaire, briquet, vacances

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P16 used specific object* : *E70 Thing*
E7 Activity : *P19 was intented use of* : *E71 Man-Made Thing*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

[172] Attention, les EP de métier « personnel » et « main d'œuvre » sont considérés comme des *E74 Group*
[173] Nous avons ajouté la relation *P11 had participant (participated in)* : *E39 Actor* pour lier l'activité de commande ou encadrement avec l'agent de maîtrise/porion et ses employés (personnel).
[174] Si l'activité de sauvetage est rencontrée dans un corpus, noter l'instance *E7 sauvetage* et non *E7 sauveteur*.
[175] À considérer plutôt comme des *E3 Condition State*

E7_Activity_Schéma 14
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs conflit militaire, bataille-Denain, deuxième guerre mondiale, guerre franco-allemande, première guerre mondiale, histoire-exploitation, après-mine, arrêt-exploitation, découverte-charbon, histoire-technique, modernisation, nationalisation, élection, équilibre écologique, faune, flore, revégétalisation (supprimé de ThesoMines), scarification-sol (supprimé de ThesoMines), risque naturel, accident-terrain, affaissement, étang-affaissement, effondrement, effondrement brutal, effondrement localisé, tassement, explosion, coup-poussière, incendie, inondation, fuite, pollution, pollution-eau, pollution-sol, pollution-sonore *Descripteurs ajoutés* bataille-charbon, plan social, coup-grisou

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P11 had participant* : *E39 Actor*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P21 had general purpose* : *E55 Type*
E7 Activity : *P117 occurs during* : *E2 Temporal Entity*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

E7_Activity_Schéma 15
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs crise économique, chômage, productivité, coût, rendement, tourisme, base-loisirs, musée, arrêt-production, reconversion, foncier, reconversion-lieu, reconversion-mineur, reconversion-site restauration

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P11 had participant* : *E39 Actor*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P16 used specific object* : *E70 Thing*
E7 Activity : *P21 had general purpose* : *E55 Type*
E7 Activity : *P117 occurs during* : *E2 Temporal Entity*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

E7_Activity_Schéma 16
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs activité domestique, jardinage, nettoyage, art, art graphique, art vivant, exposition, littérature, coutume et tradition, commémoration, danse traditionnelle, fête, bal, banquet, ducasse, nouvel an, musique, chorale, fanfare, harmonie, éducation, enseignement primaire, enseignement secondaire, enseignement supérieur, formation professionnelle, religion, événement religieux, Noël, Sainte-Barbe, Sainte-Catherine, sport et loisir, colombophilie, combat-coq, cyclisme, football, javelot, jeu, billon, bourle, jeu-cartes, fléchette, vie associative *Descripteurs ajoutés* mariage, baptême, communion, messe, Pâques, boxe, catch, boules

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P11 had participant* : *E39 Actor*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P16 used specific object* : *E70 Thing*
E7 Activity : *P21 had general purpose* : *E55 Type*
E7 Activity : *P117 occurs during* : *E2 Temporal Entity*

E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*
E7 Activity : *P161 has spatial projection* : *E53 Place* (pour décrire par exemple un bal Polonais)

E7_Activity_Schéma 17
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs avantage social, bourse scolaire, bourse-mines, caisse-retraite, pension, œuvre sociale, conflit social, conflit linguistique, grève, revendication, immigration[176], immigration belge, immigration italienne, immigration marocaine, immigration polonaise *Descripteurs ajoutés* chauffage, logement[177] , congé payé,

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P11 had participant* : *E39 Actor*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P21 had general purpose* : *E55 Type*
E7 Activity : *P117 occurs during* : *E2 Temporal Entity*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

E7_Activity_Schéma 18
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs condition-vie, condition-enfant, galibot, condition-femme, condition-retraité, paternalisme

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P10 falls within* : *E92 Spacetime Volume*
E7 Activity : *P11 had participant* : *E39 Actor*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P21 had general purpose* : *E55 Type*
E7 Activity : *P117 occurs during* : *E2 Temporal Entity*
E7 Activity : *P134 continued* : *E7 Activity*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

E7_Activity_Schéma 33
https://cmapscloud.ihmc.us:443/rid=1VPSJBDPR-1HZNHGC-7DMNTB
Liste des descripteurs risque sanitaire

E7 Activity : *P1 is identified by* : *E41 Appellation*
E7 Activity : *P2 has type* : *E55 Type*
E7 Activity : *P7 took place at* : *E53 Place*
E7 Activity : *P8 took place on or within* : *E18 Physical Thing*
E7 Activity : *P14 carried out by* : *E39 Actor*
E7 Activity : *P15 was influenced by* : *E1 CRM Entity*
E7 Activity : *P16 used specific object* : *E70 Thing*
E7 Activity : *P19 was intended use of* : *E71 Man-Made Thing*
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*

[176] Le concept d'immigration et ses descripteurs spécifiques peuvent être décrit plus spécifiquement par un *E9 Move*
[177] Fait d'être logé

E9_Move

https://cmapscloud.ihmc.us:443/rid=1WJ2RXS5R-1JQ1J14-20W5NRX
Liste des descripteurs descente, remonte, immigration, transport *Liste des descripteurs ajoutés* immigration portugaise, immigration belge, immigration italienne, immigration marocaine, immigration polonaise Note : Pour la remonte du charbon utiliser la propriété *P16 used specific object E70 Thing* car *E19 Physical Object* ne peut pas représenter E18 charbon La *P25* ne peut déplacer que des *E19 Physical Object* (dont *E21 Person*), et ne peut déplacer des *E74 Group*. Pour cela utiliser la *P11 had participant* (participated in) : *E39 Actor* Pour mentionner le transport par un animal, il faudrait ajouter *E9 Move* : *P12 occurred in the presence of (was present at)* : *E77 Persistent Item*

E9 Move : *P1 is identified by* : *E41 Appellation*
E9 Move : *P2 has type* : *E55 Type*
E9 Move : *P16 used specific object* : *E70 Thing*
E9 Move : *P11 had participant* : *E39 Actor*
E9 Move : *P25 moved* : *E19 Physical Object*
E9 Move : *P26 moved to* : *E53 Place*
E9 Move : *P160 has temporal projection* : *E52 Time-Span*

E11_Modification

E11_Modification_Schéma 19
https://cmapscloud.ihmc.us:443/rid=1VPSMGJ2X-1GVKQQW-7DXQLY Pour un but qui ne pointe pas vers un type (P21), la P20_had_specific_purpose : *E5 Event, E7 Activity*... a été ajoutée
Liste des descripteurs soutènement, ancrage, boisage, cadrage, foudroyage, remblayage, soutènement marchant, réhabilitation

E11 Modification : *P1 is identified by* : *E41 Appellation*
E11 Modification : *P2 has type* : *E55 Type*
E11 Modification : *P8 took place on or within* : *E18 Physical Thing*
E11 Modification : *P14 carried out by* : *E39 Actor*
E11 Modification : *P16 used specific object* : *E70 Thing*
E11 Modification : *P19 was intended use of* : *E71 Man-Made Thing*
E11 Modification : *P20 had specific purpose* : *E5 Event*
E11 Modification : *P31 has modified* : *E24 Physical Man-Made Thing*
E11 Modification : *P32 used general technique* : *E55 Type*
E11 Modification : *P117 occurs during* : *E2 Temporal Entity*
E11 Modification : *P126 employed* : *E57 Material*
E11 Modification : *P134 continued* : *E7 Activity*
E11 Modification : *P160 has temporal projection* : *E52 Time-Span*

E11_Modification_Schéma 20
https://cmapscloud.ihmc.us:443/rid=1VPSMGJ2X-1GVKQQW-7DXQLY
Liste des descripteurs restauration, décontamination, dégazage

E11 Modification : *P1 is identified by* : *E41 Appellation*
E11 Modification : *P2 has type* : *E55 Type*
E11 Modification : *P7 took place at* : *E53 Place*
E11 Modification : *P8 took place on or within* : *E18 Physical Thing*
E11 Modification : *P11 had participant* : *E39 Actor*
E11 Modification : *P14 carried out by* : *E39 Actor*
E11 Modification : *P16 used specific object* : *E70 Thing*
E11 Modification : *P21 had general purpose* : *E55 Type*
E11 Modification : *P31 has modified* : *E24 Physical Man-Made Thing*
E11 Modification : *P117 occurs during* : *E2 Temporal Entity*
E11 Modification : *P126 employed* : *E57 Material*
E11 Modification : *P134 continued* : *E7 Activity*
E11 Modification : *P160 has temporal projection* : *E52 Time-Span*

E13_Attribute Assignement

E13_Attribute Assignement_Schéma 24
https://cmapscloud.ihmc.us:443/rid=1VKB3K989-1RLGY2T-6TVVR4
Liste des descripteurs classement, Inventaire, Monument historique, patrimoine mondial

E13 Attribute Assignement : *P1 is identified by* : *E41 Appellation*
E13 Attribute Assignement : *P2 has type* : *E55 Type*
E13 Attribute Assignement : *P7 took place at* : *E53 Place*
E13 Attribute Assignement : *P11 had participant* : *E39 Actor*
E13 Attribute Assignement : *P14 carried out by* : *E39 Actor*
E13 Attribute Assignement : *P21 had general purpose* : *E55 Type*
E13 Attribute Assignement : P140 assigned attribute to : *E1 CRM Entity*
E11 Modification : *P160 has temporal projection* : *E52 Time-Span*

E65 Creation

https://cmapscloud.ihmc.us:443/rid=1VRC74JPW-2CD16CN-5W0KJ4
Liste des descripteurs *à ajouter*

E65 Creation : *P1 is identified by* : E50 Date
E65 Creation : *P14 carried out by* : *E39 Actor*
E65 Creation : *P94 has created* : *E73 Information Object*

E83_Type Creation

E83_Type Creation_Schéma 21
https://cmapscloud.ihmc.us:443/rid=1VRC74JPW-2CD16CN-5W0KJ4
Liste des descripteurs longue-taille, chambre-magasin, chambre-pilier *Descripteurs ajoutés* méthode-exploitation

E83 Type Creation : *P1 is identified by* : *E41 Appellation*
E83 Type Creation : *P2 has type* : *E55 Type*
E83 Type Creation : *P7 took place at* : *E53 Place*
E83 Type Creation : *P8 took place on or within* : *E18 Physical Thing*
E83 Type Creation : *P14 carried out by* : *E39 Actor*
E83 Type Creation : *P16 used specific object* : *E70 Thing*
E83 Type Creation : *P19 was intended use of* : *E71 Man-Made Thing*
E83 Type Creation : *P20 had specific purpose* : *E5 Event*
E83 Type Creation : *P117 occurs during* : *E2 Temporal Entity*
E83 Type Creation : *P160 has temporal projection* : *E52 Time-Span*

E83_Type Creation_Schéma 22
https://cmapscloud.ihmc.us:443/rid=1VRC74JPW-2CD16CN-5W0KJ4
Liste des descripteurs nuisance sanitaire, étroitesse, nuisance-sonore, nuisance thermique, obscurité

E83 Type Creation : *P1 is identified by* : *E41 Appellation*
E83 Type Creation : *P2 has type* : *E55 Type*
E83 Type Creation : *P7 took place at* : *E53 Place*
E83 Type Creation : *P8 took place on or within* : *E18 Physical Thing*
E83 Type Creation : *P14 carried out by* : *E39 Actor*
E83 Type Creation : *P16 used specific object* : *E70 Thing*
E83 Type Creation : *P19 was intended use of* : *E71 Man-Made Thing*
E83 Type Creation : *P160 has temporal projection* : *E52 Time-Span*

Schéma 23
https://cmapscloud.ihmc.us:443/rid=1VRC74JPW-2CD16CN-5W0KJ4
Liste des descripteurs énergie, consommation-énergie, économie-énergie, pénurie-énergie, type-énergie, électricité, énergie fossile[178], énergie hydraulique, énergie mécanique, énergie pneumatique, traction animale

E83 Type Creation : *P1 is identified by* : *E41 Appellation*
E83 Type Creation : *P2 has type* : *E55 Type*
E83 Type Creation : *P7 took place at* : *E53 Place*
E83 Type Creation : *P8 took place on or within* : *E18 Physical Thing*
E83 Type Creation : *P11 had participant* : *E39 Actor*
E83 Type Creation : *P14 carried out by* : *E39 Actor*
E83 Type Creation : *P16 used specific object* : *E70 Thing*
E83 Type Creation : *P19 was intended use of* : *E71 Man-Made Thing*
E83 Type Creation : *P21 had general purpose* : *E55 Type*
E83 Type Creation : *P117 occurs during* : *E2 Temporal Entity*
E83 Type Creation : *P160 has temporal projection* : *E52 Time-Span*

E64_End of existence, E69_Death et E6_Destruction

https://cmapscloud.ihmc.us:443/rid=1WJ6DJGSF-1FK540Q-210Y977 Concernant la *E6 Destruction*, si l'instance est intentionnelle, la noter également comme un *E7 Activity*
Liste des descripteurs E69 Death mort *E6 Destruction* démantèlement

Classes et propriétés	**Description des instances**
E69_Death : P1 is identified by (identifies) : *E41 Appellation*	Dénomination de l'instance en contexte
E69_Death : P2 has type (is type of) : *E55 Type*	Lien vers un descripteur de ThesoMines
E69_Death : P7 took place at (witnessed) : *E53 Place*	Lieu précis de la mort
E69_Death : P8 took place on or within (witnessed) : *E18 Physical Thing*	Si la mort se déroule à l'intérieur ou sur un *E25 Man-Made Feature*
E69_Death : P11 had participant (participated in) : *E39 Actor*	Participants
E69_Death : P93 took out of existence (was taken out of existence by) : *E77 Persistent Item*	Mort d'un groupe ou d'un animal
E69_Death : P117 occurs during (includes) : *E2 Temporal Entity*	Si la mort arrive pendant une autre activité
E69_Death : P160 has temporal projection (is temporal projection of) : *E52 Time-Span*	Période de la mort

E6 Destruction : *P13 destroyed* : *E18 Physical Thing*

[178] Attention l'employé pour « charbon » (désignant le produit) du descripteur énergie fossile se rattache au schéma 25, *E18 Physical Thing*.

E39_Actor, E21_Person et E74 Group

Note : ajout de la propriété *P44* pour décrire l'état de santé d'une personne (un mineur a la silicose) Note : la propriété *P1 is identified by* est utilisée au lieu de *P131 is identified by* : *E82 Actor Appellation* Note : la propriété *P107.1* est une sous-propriété de *P107*. **https://cmapscloud.ihmc.us:443/rid=1VRMTHDYK-Z7ZRG7-70RXWS**
Ajout de descripteurs : *E74 Group* (EP de métier : main d'œuvre, personnel) *E21 Person* (descripteurs spécifiques de « métier » et leur EP, à lister) *descripteurs ajoutés* acteur social comité-entreprise, comité-hygiène et sécurité, compagnie, patronat, régime minier, société secours minière, armée, police, service-sécurité, syndicat, parti politique, livreur Dans les textes, ceci correspond à tous les adjectifs décrivant des communautés : polonais, arabes, nord-africains, français, mineurs, famille, italiens etc.

E21 Person : *P1 is identified by* : *E41 Appellation*
E21 Person : *P2 has type* : *E55 Type*
E21 Person : *P44 has condition* : *E3 Condition State*
E21 Person : *P74 has current or former residence* : *E53 Place*
E21 Person : *P152 had parent* : *E21 Person*

E74 Group : *P1 is identified by* : *E41 Appellation*
E74 Group : *P2 has type* : *E55 Type*
E74 Group : *P74 has current or former residence* : *E53 Place*
E74 Group : *P107 has current or former member*
E74 Group : *P107.1 kind of member*

E18_Physical Thing

E18_Physical Thing_Schéma 25
https://cmapscloud.ihmc.us:443/rid=1VRC8YCCJ-23SRV21-5WBYLM
La *P44 has condition* (is condition of) : *E3 Condition State* a été ajoutée pour pouvoir noter par exemple l'état de l'inclinaison d'une veine (dressant, plateure etc.) La *P10 falls within* a également été ajoutée pour pouvoir situer les E18 dans un espace plus large comme un E25 ou un E27, par exemple un gisement (E18) dans un chantier (E25). *Liste des descripteurs* substance dangereuse, grisou, puteux *Descripteurs ajoutés* gisement, amiante, poussière, charbon (EP d'énergie fossile)

E18 Physical Thing : *P1 is identified by* : *E41 Appellation*
E18 Physical Thing : *P2 has type* : *E55 Type*
E18 Physical Thing : *P10 falls within* : *E92 Spacetime Volume*
E18 Physical Thing : *P46 is composed of* : *E18 Physical Thing*
E18 Physical Thing : *P44 has condition* : *E3 Condition State*
E18 Physical Thing : *P101 had as general use* : *E55 Type*
E18 Physical Thing : *P156 occupies* : *E53 Place*

E18_Physical Thing_Schéma 32
https://cmapscloud.ihmc.us:443/rid=1VRC8YCCJ-23SRV21-5WBYLM
Liste des descripteurs édifice religieux équipement-sport et loisir

E18 Physical Thing : *P1 is identified by* : *E41 Appellation*
E18 Physical Thing : *P2 has type* : *E55 Type*
E18 Physical Thing : *P10 falls within* : *E92 Spacetime Volume*
E18 Physical Thing : *P44 has condition* : *E3 Condition State*
E18 Physical Thing : *P46 is composed of* : *E18 Physical Thing*
E18 Physical Thing : *P101 had as general use* : *E55 Type*
E18 Physical Thing : *P156 occupies* : *E53 Place*

E24_Physical_Man-Made_Thing

E24_Physical_Man-Made_Thing_Schéma 34
https://cmapscloud.ihmc.us:443/rid=1VRN2HN3C-12KJKF2-71N0LV
Liste des descripteurs carreau-fond[179]

E24 Physical Man-Made Thing : *P1 is identified by* : *E41 Appellation*
E24 Physical Man-Made Thing : *P2 has type* : *E55 Type*
E24 Physical Man-Made Thing : *P43 has dimension* : *E54 Dimension* : *P90 has value* : *E60 Number* : *P91 has unit* : *E58 Measurement Unit*
E24 Physical Man-Made Thing : *P44 has condition* : *E3 Condition State* : *P4 has time-span* : *E52 Time-Span*
E24 Physical Man-Made Thing : *P53 has former or current location* : *E53 Place*
E24 Physical Man-Made Thing : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object

E22_Man-Made_Object_Schéma 1
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs bâti civil

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P51 has former or current owner* : *E39 Actor*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 2
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs bâti-transport, transport ferroviaire, cavalier, gare, transport fluvial, rivage, transport maritime

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P48 has preferred identifier* : *E42 Identifier*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P51 has former or current owner* : *E39 Actor*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P101 had as general use*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 3
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs chapelle, église, mosquée, presbytère

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*

[179] Attention, l'EP « fosse » de carreau-fond est noté E27 Site.

E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P48 has preferred identifier* : *E42 Identifier*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P51 has former or current owner* : *E39 Actor*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P101 had as general use*
E22 Man-Made Object : *P128 carries* : *E90 Symbolic Object*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 4
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs salle-fête, équipement sanitaire, centre-rééducation, dispensaire, équipement municipal, épicerie, maison syndicale, monument commémoratif, monument-morts *Descripteurs ajoutés* équipement commercial, halles, hôpital, maternité, café

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P48 has preferred identifier* : *E42 Identifier*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P101 had as general use*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 5
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs accrochage, dispatch, point-chargement

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P43 has dimension* : *E54 Dimension*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P59 has section* : *E53 Place*
E22 Man-Made Object : *P101 had as general use* : *E55 Type*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 5ter
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs cuvelage

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P43 has dimension* : *E54 Dimension*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P59 has section* : *E53 Place*
E22 Man-Made Object : *P101 had as general use* : *E55 Type*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 5quar
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs

carreau-surface[180], ateliers centraux, bâti-administration, bâti-machine, bâti-recette, bâti-triage, centrale électrique, château-eau, chaufferie, chevalement, douche-vestiaire, infirmerie, lavoir, local technique, local-stockage, lampisterie, local benzine, salle-compresseur
Descripteurs ajoutés :
poste-commande, poste-secours, salle-pointage, carin

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P43 has dimension* : *E54 Dimension*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P101 had as general use* : *E55 Type*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 6
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs

MATERIEL
tenue, casque, chaussures, gamelle, gourde, sac, vêtement, matériel-abattage, matériel-abattage-explosif, détonateur, explosif, dynamite, nitrate fuel, matériel-abattage manuel, marteau, pic, matériel-abattage mécanique, haveuse, mineur continu, rabot, tunnelier, matériel-forage, barre-mine, chariot-foration, jumbo, foreuse, foreuse à couronne, marteau-perforateur, marteau-piqueur, matériel-aérage, buse, porte-aérage, ventilateur, ventilateur primaire, ventilateur secondaire, matériel-exhaure, crépine-aspiration, pompe, matériel-prospection, matériel-sondage, sondeuse, trépan, matériel-soutènement, matériel-boisage, matériel-foudroyage, effondreur, matériel-remblayage, matériel-soutènement marchant, vérin hydraulique, scie, matériel-sécurité, antidéflagrant, arrêt-barrage, barrière-sûreté, détecteur-gaz, jeton, matériel-éclairage, accroche-lampe, lampe, cuirasse, lumignon, mouchette, matériel-étanchéité, matériel-incendie, matériel-secourisme, matériel-signalisation, sirène, matériel-extraction, machine-extraction, tour d'extraction, câble, cage, taquet, molette, treuil, matériel-communication, généphone, interphone, matériel-pointage, cahier-pointage, matériel-sauvetage, équipement respiratoire, matériel-transport, berline, basculeur, culbuteur, ravanceur, camion, tombereau, chargeuse, chargeuse transporteuse, pelleteuse, convoyeur, couloir, couloir oscillant, couloir roulant, matériel-manutention et stockage, contrepoids, pelle, matériel-manœuvre, aiguillage, matériel-calage et fixation, matériel-bureau, concasseur, matériel-triage, matériel-énergie, truck

Descripteurs ajoutés
foret, matériel-chauffage, modem, compresseur, convertisseur, transformateur, matériel-entretien, burin, cintreuse, hache, masse, câble, tambour, matériel-pointage, pulvérisateur, télévigile, bus, cric, palan, treuil, matériel-mesure, ampèremètre, détecteur-gaz, enregistreur-cordée, nacelle, rail, télésiège, train, vélo, distinction, médaille, tabac, matériel-administration, matériel-traitement-minerai

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P43 has dimension* : *E54 Dimension*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P51 has former or current owner* : *E39 Actor*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P55 has current location* : *E53 Place*
E22 Man-Made Object : *P58 has section definition* : *E46 Section Definition*
E22 Man-Made Object : *P59 has section* : *E53 Place*
E22 Man-Made Object : *P101 had as general use* : *E55 Type*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 7
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs
costume traditionnel

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*

[180] Si l'employé pour « jour » est identifié dans le texte, noter l'instance comme un lieu E53 Place « jour »

E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P51 has former or current owner* : *E39 Actor*
E22 Man-Made Object : *P101 had as general use* : *E55 Type*
E22 Man-Made Object : *P128 carries* : *E90 Symbolic Object*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 8
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs repas, aliment, platsek, pain d'alouette, boisson, briquet

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E22_Man-Made_Object_Schéma 9
https://cmapscloud.ihmc.us:443/rid=1VQS4D5WN-1HW19BS-33ZP1P
Liste des descripteurs activité domestique, jardinage, nettoyage, art, art graphique, art vivant, exposition, littérature, musique, religion, événement religieux, Noël, Sainte-Barbe, Sainte-Catherine, sport et loisir, colombophilie, combat-coq, cyclisme, football, javelot, jeu, billon, bourle, jeu-cartes, fléchette

E22 Man-Made Object : *P1 is identified by* : *E41 Appellation*
E22 Man-Made Object : *P2 has type* : *E55 Type*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P49 has former or current keeper* : *E39 Actor*
E22 Man-Made Object : *P51 has former or current owner* : *E39 Actor*
E22 Man-Made Object : *P53 has former or current location* : *E53 Place*
E22 Man-Made Object : *P101 had as general use* : *E55 Type*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*

E25_Man-Made_Feature

E25_Man-Made_Feature_Schéma 5bis
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs galerie principale, maillage, voie-base, voie-tête *Descripteurs ajoutés* albraque, étage

E25 Man-Made Feature : *P1 is identified by* : *E41 Appellation*
E25 Man-Made Feature : *P2 has type* : *E55 Type*
E25 Man-Made Feature : *P10 falls within* : *E92 Spacetime Volume*
E25 Man-Made Feature : *P43 has dimension* : *E54 Dimension*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P59 has section* : *E53 Place*
E25 Man-Made Feature : *P101 had as general use* : *E55 Type*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*

E25_Man-Made_Feature_Schéma 26
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs paroi, refuge, sol, toit, avaleresse

E25 Man-Made Feature : *P1 is identified by* : *E41 Appellation*
E25 Man-Made Feature : *P2 has type* : *E55 Type*
E25 Man-Made Feature : *P10 falls within* : *E92 Spacetime Volume*

E25 Man-Made Feature : *P43 has dimension* : *E54 Dimension*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State*
E25 Man-Made Feature : *P45 consists of* : *E57 Material*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P101 had as general use* : *E55 Type*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*

E25_Man-Made_Feature_Schéma 26bis
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs galerie, plan incliné, descenderie, montage, puits, bouniou, bure, goyot, puits-aérage, puits-exhaure, puits-extraction, raval *Descripteurs ajoutés* garage, écurie

E25 Man-Made Feature : *P1 is identified by* : *E41 Appellation*
E25 Man-Made Feature : *P2 has type* : *E55 Type*
E25 Man-Made Feature : *P10 falls within* : *E92 Spacetime Volume*
E25 Man-Made Feature : *P43 has dimension* : *E54 Dimension*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State*
E25 Man-Made Feature : *P45 consists of* : *E57 Material*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P101 had as general use* : *E55 Type*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*

E25_Man-Made_Feature_Schéma 27
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs chantier, cloche Pour noter qu'un chantier se situe dans un lieu précis utiliser la P161 : *E25 Man-Made Feature* (chantier) : *P161 has spatial projection* (is spatial projection of) : *E53 Place* (Arenberg)

E25 Man-Made Feature : *P1 is identified by* : *E41 Appellation*
E25 Man-Made Feature : *P2 has type* : *E55 Type*
E25 Man-Made Feature : *P10 falls within* : *E92 Spacetime Volume*
E25 Man-Made Feature : *P43 has dimension* : *E54 Dimension*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State*
E25 Man-Made Feature : *P45 consists of* : *E57 Material*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*
E25 Man-Made Feature : *P161 has spatial projection* : *E53 Place*

E25_Man-Made_Feature_Schéma 28
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs déblai, front, havée

E25 Man-Made Feature : *P1 is identified by* : *E41 Appellation*
E25 Man-Made Feature : *P2 has type* : *E55 Type*
E25 Man-Made Feature : *P43 has dimension* : *E54 Dimension*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State*
E25 Man-Made Feature : *P45 consists of* : *E57 Material*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*

E25_Man-Made_Feature_Schéma 29
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs ouverture, quartier

E25 Man-Made Feature : *P1 is identified by* : *E41 Appellation*
E25 Man-Made Feature : *P2 has type* : *E55 Type*
E25 Man-Made Feature : *P43 has dimension* : *E54 Dimension*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State*
E25 Man-Made Feature : *P53 has former or current location* : *E53 Place*

E25 Man-Made Feature : *P101 had as general use* : *E55 Type*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*

E25_Man-Made_Feature_Schéma 30
https://cmapscloud.ihmc.us:443/rid=1VRC2K7FC-40HSJB-5V9VXY
Liste des descripteurs passe (supprimé de ThesoMines) – schéma à supprimer

E73_Information Object, E31_Document et E33_Linguistic Object

https://cmapscloud.ihmc.us:443/rid=1WJX0M6QQ-1HR2RC2-2278WC3 **Schéma validé le 12/02/21** *Descripteurs ajoutés* **Pour E56 Language** langue **pour E31 Document** littérature, art, art graphique, art vivant, musique

E31 Document : *P1 is identified by* : *E41 Appellation*
E31 Document : *P2 has type* : *E55 Type*
E31 Document : *P102 has title : E35 Title*
E31 Document : *P70 documents : E1 CRM Entity*

E33 Linguistic Object : *P1 is identified by* : *E41 Appellation*
E33 Linguistic Object : *P2 has type* : *E55 Type*
E33 Linguistic Object : *P73 has translation* : *E33 Linguistic Object*
E33 Linguistic Object : *P72 has language* : *E56 Language*

E27_Site

E27_Site_Schéma 10
https://cmapscloud.ihmc.us:443/rid=1VRC430CB-2955BXB-5VJBNF
Liste des descripteurs cimetière, prieuré, centre-vacances, équipement scolaire, centre ménager, école, habitat minier, cité minière, cité moderne, cité pavillonnaire, cité-jardin, coron, habitat-ingénieur, habitat-patron, friche industrielle, terril, zone inondable, zone tampon Proposition : ajouter la propriété *P46 is composed of* (forms part of) : *E18 Physical Thing* pour faire le lien entre le terril et ses composants (schistes…) *Descripteurs ajoutés :* marché, jardin, bassin-décantation, fosse (EP de carreau-fond qui est un E24) **Note :** Concernant les fosses, la question de leur situation spatiale et de leur évolution dans le temps se pose. Celles-ci ont une spatialité différente en souterrain par rapport à la surface. Sur Wikipédia, chaque fosse possède un jeu de coordonnées spatiales issues d'une base BRGM mais ne correspondant qu'à un seul point (et non une plage de points), par exemple la Fosse n° 6 - 6 bis - 6 ter des mines de Bruay a pour coordonnées 50,464819, 2,566567. Si ces coordonnées sont utilisées pour décrire les fosses, toutes les instances liées à ces lieux seront représentées sur un point unique. **Note :** Toujours concernant les fosses sur Wikipédia, les différents étages sont notés (correspondant à la profondeur en mètres des différents niveaux souterrains). Il serait intéressant de pouvoir faire le lien entre ces étages et les fosses (propriété à identifier). **En tant que fils de *E18 Physical Thing*, E27_Site permet de décrire :** **Le site sur lequel ou à l'intérieur duquel se situe :** *E7 Activity* : *P8 took place on or within* : *E18 Physical Thing* *E11 Modification* : *P8 took place on or within* : *E18 Physical Thing* *E11 Modification* : *P8 took place on or within* : *E18 Physical Thing* *E83 Type Creation* : *P8 took place on or within* : *E18 Physical Thing* *E4 Period*. *P8 took place on or within* (witnessed) : *E18 Physical Thing* **Le site détruit par :** *E6 Destruction* : *P13 destroyed* : *E18 Physical Thing*

La composition de :
E18 Physical Thing : *P46 is composed of* : *E18 Physical Thing*
E22 Man-Made Object : *P46 is composed of* : *E18 Physical Thing*

Le constituant de :
E22 Man-Made Object : *P45 consists of* : *E18 Physical Thing*

En tant que fils de E92 Spacetime Volume, E27 Site est utilisée pour inscrire les classes suivantes dans un espace temps :
E7 Activity : *P10 falls within* : *E92 Spacetime Volume*
E18 Physical Thing : *P10 falls within* : *E92 Spacetime Volume*
E22 Man-Made Object : *P10 falls within* : *E92 Spacetime Volume*
E25 Man-Made Feature : *P10 falls within* : *E92 Spacetime Volume*
E27 Site : *P10 falls within* : *E92 Spacetime Volume*

E27 Site : *P1 is identified by* : *E41 Appellation*
E27 Site : *P2 has type* : *E55 Type*
E27 Site : *P10 falls within* : *E92 Spacetime Volume*
E27 Site : *P44 has condition* : *E3 Condition State*
E27 Site : *P48 has preferred identifier* : *E42 Identifier*
E27 Site : *P49 has former or current keeper* : *E39 Actor*
E27 Site : *P51 has former or current owner* : *E39 Actor*
E27 Site : *P53 has former or current location* : *E53 Place*
E27 Site : *P101 had as general use* : *E55 Type*
E27 Site : *P160 has temporal projection* : *E52 Time-Span*

E92_Spacetime_Volume

E92_Spacetime_Volume_Schéma 35
https://cmapscloud.ihmc.us:443/rid=1VT2V8CKB-26MR2P1-G44KVP
Liste des descripteurs site et paysage **E92 Spacetime Volume est utilisée pour inscrire les classes suivantes dans un espace temps :** *E7 Activity* : *P10 falls within* : *E92 Spacetime Volume* *E18 Physical Thing* : *P10 falls within* : *E92 Spacetime Volume* *E22 Man-Made Object* : *P10 falls within* : *E92 Spacetime Volume* *E25 Man-Made Feature* : *P10 falls within* : *E92 Spacetime Volume* *E27 Site* : *P10 falls within* : *E92 Spacetime Volume* *E92 Spacetime Volume : P161 has spatial projection : E53 Place*

E92 Spacetime Volume : *P1 is identified by* : *E41 Appellation*
E92 Spacetime Volume : *P2 has type* : *E55 Type*
E92 Spacetime Volume : *P160 has temporal projection* : *E52 Time-Span*

E52_Time-Span et E50 Date

https://cmapscloud.ihmc.us:443/rid=1WJ6KLXRR-XC05SF-2118LYX
Schéma acté
Note : pour décrire qu'une activité dure 6 mois entre 1987 et 1990, noter
E52 Time-Span (6 mois) : P81 ongoing throughout : E61 Time Primitive (1987-1990)

E52 Time-Span est utilisée pour inscrire les classes suivantes dans une période temporelle :
E7 Activity : *P160 has temporal projection* : *E52 Time-Span*
E9 Move : *P160 has temporal projection* : *E52 Time-Span*
E11 Modification : *P160 has temporal projection* : *E52 Time-Span*
E83 Type Creation : *P160 has temporal projection* : *E52 Time-Span*
E92 Spacetime Volume. *P160 has temporal projection* (is temporal projection of) : *E52 Time-Span* (noter la période)
E24 Physical Man-Made Thing : *P160 has temporal projection* : *E52 Time-Span*
E3 Condition State : *P4 has time-span* : *E52 Time-Span*
E22 Man-Made Object : *P160 has temporal projection* : *E52 Time-Span*
E25 Man-Made Feature : *P160 has temporal projection* : *E52 Time-Span*
E27 Site : *P160 has temporal projection* : *E52 Time-Span*
E92 Spacetime Volume : *P160 has temporal projection* : *E52 Time-Span*

E50 Date est utilisée pour spécifier une date
E65 Creation : *P1 is identified by* : E50 Date
E52 Time-Span : P78 is identified by : E50 Date

E52 Time-Span : P78 is identified by : E50 Date
E52 Time-Span : *P2 has type* : *E55 Type*
E52 Time-Span : P81 ongoing throughout : E61 Time Primitive

E53_Place

https://cmapscloud.ihmc.us:443/rid=1VRNGGBY6-MHVS56-7419H7
Cette classe instancie tous les lieux et noms de lieux qui peuvent se situer sur une carte grâce à des coordonnées GPS tels que les villes, villages, régions.
Note : utiliser Place Name pour noter les noms de lieu et non *E41 Appellation*
Note : *au jour*, *au fond*, *jour* et *fond* peuvent être des instances de *E53 Place* même si celles-ci ne sont pas précises, c'est-à-dire même si elles n'ont pas de coordonnées GPS (elles pourraient être localisées sur un schéma).

E53 Place est utilisée dans les cas suivants :

Pour indiquer la localisation actuelle ou ancienne :
- E27 Site : *P53 has former or current location* : *E53 Place (terril de Loos-en-gohelle)*
- *E25 Man-Made Feature* : *P53 has former or current location* : *E53 Place (puits n°2 de la fosse Sabatier à Raismes)*
- *E22 Man-Made Object* : *P53 has former or current location* : *E53 Place (église Sainte-Barbe d'Arenberg)*
- *E24 Physical Man-Made Thing* : *P53 has former or current location* : *E53 Place* (souterrain de la fosse 9-9bis d'Oignies)

Pour indiquer la résidence courante ou ancienne :
- *E21 Person* : *P74 has current or former residence* : *E53 Place (JM Minot à Waziers)*
- *E74 Group* : *P74 has current or former residence* : *E53 Place (Association Germinal à Anzin)*

Pour indiquer la localisation actuelle :
- *E22 Man-Made Object* : *P55 has current location* : *E53 Place (pelle de mineur du père de Martial à Givenchy-en-Gohelle)*

Pour indiquer la localisation sur une surface :
- *E22 Man-Made Object* : *P59 has section* : *E53 Place*
- *E25 Man-Made Feature* : *P59 has section* : *E53 Place*

Pour inscrire quelque chose dans une spatialité :
- *E25 Man-Made Feature* : *P161 has spatial projection* : *E53 Place*
- *E7 Activity* : *P161 has spatial projection* : *E53 Place* (pour décrire par exemple un bal Polonais)

Pour indiquer la localisation d'un événement :
- *E7 Activity* : *P7 took place at* : *E53 Place*
- *E11 Modification* : *P7 took place at* : *E53 Place*
- *E13 Attribute Assignement* : *P7 took place at* : *E53 Place*
- *E13 Attribute Assignement* : *P7 took place at* : *E53 Place*
- *E4 Period*. *P7 took place at* (witnessed) : *E53 Place* (noter le lieu précis de la mort)

- *E13 Attribute Assignement* : *P7 took place at* : *E53 Place*

Pour localiser un déplacement :
- *E9 Move* : *P26 moved to* : *E53 Place*
- *E9 Move* : *P26 moved to* : *E53 Place*

Pour indiquer le lieu dans lequel se trouve un objet :
- *E18 Physical Thing* : *P156 occupies* : *E53 Place*

Pour indiquer le lieu dans lequel se trouve un lieu :
E53 Place : P89 falls within : *E53 Place*

E53 Place : *P87 is identified by* : E44 Place Appellation
E53 Place : *P87 is identified by* : E45 Address
E53 Place : *P87 is identified by* : E47 Spatial Coordinates
E53 Place : *P87 is identified by* : E48 Place Name
E53 Place : P89 falls within : *E53 Place*

E3_Condition State

https://cmapscloud.ihmc.us:443/rid=1WR090TKB-29P08FF-D8JBZZ
Propriétés sélectionnées :
Lien vers thésaurus (P2), dénomination (P1), durée de l'état (P4), et précision de l'état par un état (P5). La durée semble primordiale d'après le CIDOC CRM.

Note pour plus tard :
P117 occurs during (includes) : *E2 Temporal Entity* (pour faire le lien entre l'état et un *E7 Activity* ex : asphyxie pdt un coup de grisou)
Descripteurs ajoutés
inclinaison, dressant, plateure, semi-dressant, asphyxie, maladie professionnelle, silicose

E3 Condition State permet de définir l'état des classes suivantes :
E21 Person : *P44 has condition* : *E3 Condition State (personne asphyxiée, malade)*
E18 Physical Thing : *P44 has condition* : *E3 Condition State (gisement dressant)*
E24 Physical Man-Made Thing : *P44 has condition* : *E3 Condition State (souterrain remblayé)*
E22 Man-Made Object : *P44 has condition* : *E3 Condition State (gare en cours de transformation)*
E25 Man-Made Feature : *P44 has condition* : *E3 Condition State (chantier en cours de creusement)*
E27 Site : *P44 has condition* : *E3 Condition State (fosse inutilisée)*

E72_Legal Object et E30_Right

https://cmapscloud.ihmc.us:443/rid=1WR09MCZW-6CM6YY-D8KMGL
Réfléchir aux éventuelles propriétés
La propriété P104 permet de relier le régime minier et chacun des droits qui le constitue (système-soins etc.) D'après le CIDOC CRM : « This property links a particular E72 Legal Object to the instances of E30 Right to which it is subject. »
Descripteurs ajoutés
E72 Legal Object
régime minier, statut-mineur (cadre général qui s'applique à toutes personnes travaillant à la mine)

E30 Right : (droits et devoirs)
chauffage, logement, congé payé
règlement + instances du règlement
système-soins
sanction

E54_Dimension

https://cmapscloud.ihmc.us:443/rid=1WR0RSJGV-3S5KFZ-DBLV0G
Note : Quelques objets physiques (E24, E22, E25) peuvent être décrits par une valeur numérique et une unité.
Les instances numériques de salaire en tant que quantité d'argent précise pourraient être décrites grâce à cette classe et ses propriétés.

E28_Conceptual Object

https://cmapscloud.ihmc.us:443/rid=1WR0MKCBZ-2B7H9VR-D9SBVK
Descripteurs ajoutés
solidarité, salaire (encore en discussion)

Annexe 5 : Liste des vedettes matières RAMEAU identifiées permettant d'indexer des documents traitant d'aspects du domaine minier et alignement avec les descripteurs de ThesoMines

Les vedettes matières ci-dessous possèdent en lien hypertexte leur identifiant ARK dans le catalogue de la BnF.

La mention « -trop générique » signifie que la vedette dépasse le cadre du domaine, « absent » signifie qu'elle est absente de ThesoMines v1.0 (et peut être un candidat terme) et la mention « -voir OntoMines » signifie que le concept porté par la vedette est contenu dans OntoMines.

RAMEAU	ThesoMines v1.0
Abattage à l'explosif	abattage ; explosif
Abattage hydraulique	abattage ; énergie hydraulique
Boutefeux (ouvriers)	boutefeu
Boisage	boisage
Carrières (exploitations minières)	- trop générique
Carrières (exploitations minières) -- Droit	- trop générique
Carrières souterraines (exploitations minières) -- Accidents	accident-travail ou accident-terrain
Charbon -- Industrie et commerce	économie
Charbon -- Mines et extraction -- Accidents	accident-travail ou accident-terrain
Charbon -- Mines et extraction -- Aspect environnemental	environnement naturel
Charbon -- Mines et extraction -- Droit	règlement
Charbon -- Mines et extraction remplacée par Mines de charbon	- trop générique
Charbon -- Réserves	- absent
Conservation des ressources minières	- trop générique
Construction -- Industrie	bâti
Constructions souterraines	carreau-fond
Contrôle des pressions de terrain (mines)	accident-terrain
Coups de poussières	coup-poussière
Coups de toit (mines)	effondrement
Courrières, Catastrophe de (1906)	risque naturel
Cours d'eau souterrains	voir exhaure
Électricité dans l'exploitation minière	énergie électrique
Excavation	déblayage ; déblai
Exhaure	exhaure
Femmes mineurs	condition-femme
Forage	forage
Gaz dans les mines	substance dangereuse
Géologie minière	- voir géologue
Industries minières	- trop générique
Industries minières -- Personnel	métier
Ingénieurs des mines	ingénieur
Liévin, Catastrophe de (1974)	risque naturel ; voir OntoMines pour lieu
Longue taille (méthode d'exploitation)	longue-taille
Machines minières	matériel

RAMEAU	ThesoMines v1.0
Marteaux perforateurs	marteau-perforateur
Mines (sites d'extraction)	- voir OntoMines (*E27_Site* et *E53 Place*)
Mines (sites d'extraction) -- [Localisations géographiques]	- voir OntoMines (*E53 Place*)
Mines (sites d'extraction) -- Accidents	accident-travail ou accident-terrain
Mines (sites d'extraction) -- Appareils et matériel	matériel
Mines (sites d'extraction) -- Constructions	bâti
Mines (sites d'extraction) -- Fermeture	arrêt-production
Mines (sites d'extraction) -- Mesures de sécurité	matériel-sécurité ; service-sécurité ; métier-sécurité ; comité-hygiène et sécurité
Mines (sites d'extraction) -- Remblayage	remblayage
Mines (sites d'extraction) -- Ventilation	aérage
Mines abandonnées	friche industrielle
Mines de charbon	- trop générique
Mines de charbon -- Droit	règlement
Mines et extraction	- trop générique
Mineurs (ouvriers)	métier
Mineurs de charbon	métier
Montceau-les-Mines, Catastrophe de (1958)	risque naturel
Politique minière	histoire politique
Prospection	prospection ; métier-prospection ; matériel-prospection
Puits (excavations) -- Fonçage	creusement
Radio dans l'exploitation minière	radio
Régions minières	- voir OntoMines (*E53 Place*)
Ressources minérales	- absent
Sauvetage minier	sauveteur ; matériel-sauvetage
Sécurité sociale -- Mineurs (ouvriers)	avantage social
Sondages (mines)	sondage
Technique minière	histoire technique ; opération ; méthode-exploitation
Topographie minière	- absent
Villes minières	- voir OntoMines (*E53 Place*)
Voies de roulage (mines)	roulage

Annexe 6 : Liste des indexations thématiques de la terminologie et alignement avec les descripteurs de ThesoMines

La mention « - *trop spécifique* » signifie que le ou les termes indexés par la thématique est trop spécifique pour être inséré(s) dans ThesoMines. Le travail d'alignement est en cours.

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
ABATTAGE	abattage	explosif	action			*minage*
			action interdite			*minage*
			appareil	air comprimé		*matériel-aérage, énergie pneumatique*
			élément de bourrage			*matériel-abattage-explosif*
			explosif			*explosif*
			lieux	trou		*-trop spécifique voir minage*
			outil			*matériel-abattage-explosif*
			textile de protection	eau		*- trop spécifique voir minage*
			tir des explosifs			*minage*
ACCIDENT	accident	rail				*accident-transport*
ACTIONS	actions	abattage				*abattage*
		boisage				*boisage*
		agrandissement galerie				*creusement*
		bâti				*-trop spécifique voir bâti*
		boisage				*-doublon, boisage*
		cacher ses outils				*-trop spécifique voir opération*
		caler				*-trop spécifique voir opération*
		conduite				*transport*
		construction				*-trop spécifique voir carreau-fond*
		coupe				*abattage*
		creusement				*creusement*
		déblayage				*déblayage*
		déboisage				*boisage*
		décadrage				*cadrage*
		défaire				*-trop spécifique, absent*
		dégagement				*-trop spécifique voir carreau-fond*
		descendre				*descente*
		descente				*descente*
		eau				*métier-exhaure*
		éclairage				*-trop spécifique voir matériel-éclairage*
		économie				*-trop spécifique voir salaire*
		effectuer un ouvrage				*-trop spécifique, absent*
		évacuation de l'eau				*métier-exhaure*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		exhaure				*métier-exhaure*
		exploitation				*creusement traitement-minerai abattage*
		expression				*-trop spécifique, absent*
		forage				*forage*
		havée				*havée*
		lampiste	nettoyage de la lampe			*lampiste*
		manœuvre				*-trop spécifique, absent*
		manutention				*-trop spécifique, absent*
		nettoyage				*métier-entretien*
		organisation du travail				*-trop spécifique, absent*
		ouvrage	agrandissement galerie			*creusement*
			infrastructure de transport			*-trop spécifique voir poseur-voie*
			voie			*-trop spécifique voir poseur-voie*
		pelleter				*-trop spécifique voir remblayage et pelle*
		percer une galerie				*creusement*
		prévention	poussières			*-trop spécifique voir antidéflagrant*
		remblayage				*remblayage*
		remonter au jour				*remonte*
		risque				*-trop spécifique, absent*
		sécurité	eau			*-trop spécifique voir métier-sécurité et cuvelage*
		sondage				*sondage*
		soutènement				*soutènement*
		technique	boisage			*boisage*
			exploitation			*-trop spécifique, absent*
			manutention			*-trop spécifique, voir matériel-manutention et stockage*
			sécurité			*métier-sécurité*
			soutènement			*soutènement et métier-soutènement*
		technique minière	abattage			*abattage et métier-abattage*
		tir				*minage et boutefeu*
		transport				*métier-transport*
		transport et manutention				*métier-transport*
		travail				*productivité*
		travail préparatoire				*métier-creusement*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		tri				*métier-triage*
		triage				*métier-triage*
		utilisation d'une pièce				*-trop spécifique, absent*
		manutention et stockage				*ripage*
	résultat d'une action	boisage				*boisage*
		caler				*-trop spécifique, absent*
AIR	air	action	aérage			*aérage*
AIR&EAU	air&eau	aérage				*aérage et cheminée*
		air	partie d'appareil			*-trop spécifique, voir matériel-aérage*
			transport			*matériel-aérage*
		air comprimé	aérage secondaire			*aérage et ventilateur secondaire*
			moyen de transport (partie de)			*-trop spécifique, voir matériel-aérage ou matériel-exhaure*
			transport			*-trop spécifique, voir matériel-aérage ou matériel-exhaure*
		air ou eau	moyen de transport (partie de)			*-trop spécifique, voir matériel-aérage ou matériel-exhaure*
		eau	appareil	poussières		*pulvérisateur*
			protection			*matériel-étanchéité*
ANIMAL	animal	transport				*traction animale*
APPAREIL	appareil	air comprime				*-trop spécifique, voir énergie pneumatique*
		communication				*matériel-communication*
		mesure	inclinaison			*matériel-mesure*
			inclinaison de galerie			*matériel-mesure*
		sécurité	grisou			*matériel-sécurité, détecteur-gaz, arrêt-barrage, antidéflagrant*
APPELLATION	appellation	mineur	blessé			*Voir blessure*
			creusement			*métier-creusement*
			délégué			*Voir syndicat*
			galibot			*galibot*
			péjoratif			*-trop spécifique, absent*
			rescapé			*-trop spécifique, absent*
			syndicat			*syndicat*
			tâche			*métier-creusement*
		personne	culture			*Voir colombophilie*
		personnes				*-trop spécifique, absent*
	appellation de lieu					*-trop spécifique, absent*
BATI	bâti minier					*ateliers centraux*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
	bâti minier (partie de)					*Voir chevalement*
BOISAGE&SOUTENEMENT	boisage&soutènement	pièces				*matériel-soutènement*
		techniques				*soutènement*
CHARBON	charbon	résultat de combustion				*-absent*
	la mine l'or noir	charbon				*-absent, voir énergie fossile*
CHIMIE	chimie					*-absent*
COMMUNICATION	communication					*-trop spécifique, absent*
CULTURE	culture	fête				*ducasse*
		fumer				*Voir tabac*
		jeu				*jeu*
		langue				*langue*
		personnage				*-trop spécifique, voir littérature*
		pratique				*colombophilie*
DANGER	danger	affaissement				*affaissement*
		éboulement	personne			*-trop spécifique, voir éboulement*
		feu				*incendie*
		fuites air comprimé				*aérage et fuite*
		gaz				*substance dangereuse*
		grisou	appareil	sécurité		*détecteur-gaz*
		inondation				*inondation*
		poussières				*poussière*
DOCUMENT	document	objet du mineur				*-absent*
DROIT	droit					*compagnie*
EAU	eau	évacuation	matériel			*matériel-exhaure*
ECLAIRAGE&ACCESSOIRES	éclairage&accessoires	action	éclairage			*-trop spécifique, voir lampiste*
			éteindre			*-trop spécifique, voir lampiste*
			relevage de mèche			*-trop spécifique, voir lampiste*
		appareil	rallumage des lampes			*matériel-éclairage*
		chandelier				*lampe*
		débris de mèche				*matériel-éclairage*
		équipement	sécurité	grisou		*arrêt-barrage*
		essence				*matériel-éclairage*
		jeton				*jeton*
		lampe				*lampe*
		lampe (partie de)				*cuirasse lumignon mouchette*
		outil	allumage			*-trop spécifique, voir lampe*
		pièce	accrochage			*-trop spécifique, voir lampe*
			accrochage ou lampe			*-trop spécifique, voir lampe*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
			lampe (partie de)			*-trop spécifique, voir lampe*
		textile	accrochage de lampe			*-trop spécifique, voir lampe*
			protection	lampe		*-trop spécifique, voir lampe*
ENERGIE	énergie	coffret				*matériel-énergie*
ENVIRONNEMENT	environnement	eau				*inondation*
EQUIPEMENT	équipement	sécurité	grisou			*arrêt-barrage*
ESPACE	espace					*-trop spécifique, absent*
EXPRESSION	expression	actions	manoeuvre			*-trop spécifique, voir berline*
			travail			*-trop spécifique, voir productivité*
			vêtement			*vêtement*
		arrêt du travail				*-trop spécifique, absent*
		communication				*-trop spécifique, absent*
		difficulté du travail				*-trop spécifique, absent*
		interjection	difficulté travail			*-trop spécifique, absent*
		signal depart travail				*-trop spécifique, absent*
		tomber				*-trop spécifique, absent*
		travail				*-trop spécifique, absent*
GEOLOGIE	géologie	charbon	combustible			*-domaine de la géologie absent du thésaurus*
			gaz			*-domaine de la géologie absent du thésaurus*
			particules			*-domaine de la géologie absent du thésaurus*
			terre			*-domaine de la géologie absent du thésaurus*
		charbon ou schiste				*-domaine de la géologie absent du thésaurus*
		exploitation	terrain			*-domaine de la géologie absent du thésaurus*
		galerie				*-domaine de la géologie absent du thésaurus*
		gisement (partie de)				*-domaine de la géologie absent du thésaurus*
		période				*-domaine de la géologie absent du thésaurus*
		poussière	charbon			*poussières*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		poussières	charbon	pierre		*-domaine de la géologie absent du thésaurus*
		produit	résultat	lavage		*lavage*
		pyrite de fer				*-domaine de la géologie absent du thésaurus*
		roc&veine	accident	veine		*gisement*
			arbre ?			*-domaine de la géologie absent du thésaurus*
			carbonate de fer			*-domaine de la géologie absent du thésaurus*
			charbon	pierre	danger	*-domaine de la géologie absent du thésaurus*
			craie			*-domaine de la géologie absent du thésaurus*
			danger			*-domaine de la géologie absent du thésaurus*
			excavation	danger	grisou	*cloche*
			fossile			*-domaine de la géologie absent du thésaurus*
			grès			*-domaine de la géologie absent du thésaurus*
			houille			*énergie fossile*
			marne	argile		*-domaine de la géologie absent du thésaurus*
				sable		*-domaine de la géologie absent du thésaurus*
			mesure			*ouverture*
			partie			*-domaine de la géologie absent du thésaurus*
			protection	grisou		*-trop spécifique, voir arrêt-barrage*
			pyrite			*-domaine de la géologie absent du thésaurus*
			rectitude			*-trop spécifique, voir dressant*
			risque	éboulement		*effondrement*
			rocher			*-domaine de la géologie absent du thésaurus*
			rupture			*-domaine de la géologie absent du thésaurus*
			schiste			*terril*
			terre ou charbon			*-domaine de la géologie absent du thésaurus*
		schiste				*terril*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		terre				*-trop spécifique, voir matériel-remblayage*
	qualification	terrain (type de)				*-domaine de la géologie absent du thésaurus*
HABITAT MINIER	habitat minier					*habitat minier, coron*
INSTALLATIONS	installations					*-trop générique, absent*
LAMPE	lampe					*lampe*
LIEUX	lieux	obstacle				*-trop spécifique*
		atelier	outil			*local stockage*
		bâtiment	nettoyage/triage			*bâti-triage*
			surface	tri		*bâti-triage*
		cage				*cage*
		carreau	jour			*carreau-surface*
		cavité				*chambre-magasin et chambre-pilier*
		chantier	voie	mesure		*havée*
		chantier (partie de)				*front*
		chantier souterrain				*chantier*
		chargement des berlines				*point-chargement*
		chargement des péniches				*rivage*
		chargement des produits				*point-chargement*
		décrochage wagonnets				*accrochage*
		dépilage				*absent voir soutènement*
		dépôt de charbon				*absent voir chantier*
		douches				*douche-vestiaire*
		espace	descente			*puits*
		front de taille				*front*
		lavoir				*lavoir*
		magasin	stockage			*local-stockage*
		manutention				*absent voir chantier*
		mine	jour	galerie		*formation professionnelle*
		mine (partie de)	chantier			*havée*
			cul de sac			*absent voir carreau-fond*
			fond			*carreau-fond*
			jour			*carreau-surface*
			plafond			*toit*
			sol			*sol*
		niveau				*étage*
		niveau d'exploitation				*étage*
		partie à remblayer				*voir remblayage*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		partie du carreau				*carreau-surface et matériel-soutènement*
		partie du puits	niveau	haut		*bâti-recette*
		plan incliné				*plan incliné*
		puits	aérage			*puits-aérage*
			exhaure			*puits-exhaure*
			retour d'air			*puits-aérage*
		puits (partie de)				*cuvelage*
		puits ou surface, puits, galeries, chantiers				*carreau-fond*
		recette	plan incliné			*plan incliné et accrochage*
		recette des eaux				*accrochage*
		recette du puits				*accrochage*
		refuge				*refuge*
		répartition du matériel				*local-stockage*
		réservoir	eau			*bouniou*
		salle d'attente				*local-stockage*
		secteur de travail				*quartier*
		situation de l'exploitation				*absent*
		stockage	éclairage			*lampisterie*
		taille	formation			*formation professionnelle*
		terrain				*gisement*
		transport				*Voir accrochage*
		trou pour le boisage				*Trop spécifique, voir soutènement*
		veine (partie de)				*gisement*
		vestiaire				*douche-vestiaire*
		voie	bout			*galerie*
			croisement			*accrochage*
			eau			*albraque*
			front de taille			*front*
			impasse	stockage		*local-stockage*
			inclinée			*descenderie*
			moyen de transport			*descenderie*
			refuge			*refuge*
			remblayée			*remblayage*
		voie (aérage)				*puits-aérage*
		voie (partie de)	porte			*porte-aérage*
		voie (partie de)	fond			*galerie*
			paroi			*paroi*
			renflement			*voir paroi*
MANUTENTION& STOCKAGE	manutention et stockage > appareil	manutention horizontale continue vrac				*ripage*
MATERIAU	matériau	bois				*absent, voir matériel-boisage*
		moyen de transport				*descenderie*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
MATERIEL	matériel	prévention	coup de poussière			
	materiel	contrepoids				
	matériel	aérage				
		appareil	manutention			
		bois				
		boisage				
		communication				
		contrepoids				
		exhaure				
		forage				
		jeton				
		machine	extraction			
			vapeur			
		manoeuvre				
		manutention				
		masse				
		matériel (partie de)				
		moyen de transport				
		nettoyage				
		outil	creusement			
		prévention	coup de poussière			
		réservoir	eau			
		risque				
		soutènement				
		transport				
		treuil	transport			
		tuyau	alimentation en air comprimé			
	matériel&accessoires	abattage	engin manuel			
MESURE	mesure	marque				
		orientation du gisement				
		taux				
		unité				
METIER	métier	abattage	explosif			
		apprenti				
		cadre	inspection			
		creusement				
		déblayage				
		déboisage				
		décadrage				
		dégagement	charbon			
		éclairage				
		encagement	berlines			
		entretien				
		galibot				

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		géomètre				
		grade				
		lampiste				
		ouvrier				
			abattage			
			boisage			
			conducteur de machines	abattage		
			creusement			
			distribution	charbon		
			entretien	puits de mine		
			grade			
			machines			
			mesure	charbon		
				gaz		
			mineur			
			outils			
			remblayage			
			sécurité			
			soutènement			
			surface			
			tir			
			traçage			
			transport			
			travail	galerie		
			tri			
		recette du puits				
		santé				
		sécurité				
		surveillant				
		transport				
		tri	surface			
		vente	charbon			
METIER&TACHE	métier&tâche					
		approfondissement du puits				
		galibot				
		lampiste				
		manutention&transport				
		maréchal-ferrand				
		outils				
		porion	aérage			
		remblayage				
		tri				
		abattage				
		accrochage				
		agent de maîtrise	mesurer le travail			
		apprenti	transport berlines			

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		assemblage des tuyaux				
		bordure de chantier				
		boutefeu	tir			
		conducteur				
		creusement				
		déblayage				
		déboisage				
		entretien des galeries				
		garde				
		hercheur				
		ingénieur				
		lampiste				
		manoeuvre				
		manutention				
		manutention&transport				
		outils				
		porion	aérage			
		réception wagonnets				
		suppléant du porion				
		surveillant	propreté du charbon			
		transport				
		tri				
		vérification	direction des galeries			
MINE	mine					
OUTIL	outil (partie de)					
OUTIL&PIECES	outils&pièces	air comprimé				
		appareil d'abattage	air comprimé			
		arracher le schiste				
		bois				
		burin	forage des trous de mines			
		câble				
		contre-poids	manutention			
		élément de repérage de trous de mine				
		immobiliser les câbles				
		levier				
		manutention				
		marteau				
		marteau-piqueur				
		montre				
		outil	boucher fuites cuvelage			
			marteau			

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
			pic			
		outil (partie de)				
		outil d'abattage				
		outil de coupage				
		outil de creusement				
		outil de rognage				
		outil de serrage				
		outil pour percer des trous				
		partie d'appareil				
		pelle				
		pic				
		pièce	forage			
			manoeuvre			
			manutention			
		pièce en fer				
		pierre d'aiguisage				
		pioche				
		planche				
		scie				
		serrer les étançons				
		treuil				
		ustensile	nettoyage trou de mine			
OUVRAGES	les ouvrages	agrandissement galerie				
		bâti				
		dispositif	mesure	gaz		
		étanchéité				
		prévention	eau			
		soutènement				
		voie				
OUVRIER	ouvrier	abattage				
		mineur				
PAIE	paie					
PERIODE	période	géologique				
		travail				
POSTE	poste (type de)					
PREVENTION	prévention	opération				
PROCESSUS	processus	risque	éboulement			
PRODUCTIVITE& APPOINTEMENTS	productivité&appointements	action	temps de travail			
		arrêt de travail				
		blocage				
		boisage				
		changement travail				
		distance de travail				
		expression	compétition			
			dérision	travail		
		fin journée				

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
		gain				
		qualité du travail				
		quantité				
		retraite				
		temps de travail	retard			
		vitesse de travail ou qualité de travail				
PROSPECTION MINIERE	prospection minière					
REPAS&REPOS	repas&repos	aliment				
		boisson				
		contenant				
		expression	repos			
		gourde				
		gourde 2				
		position				
		repas				
		repos				
		sac				
RISQUE	risque	adj				
		fuite				
		matériel				
		tomber				
		verbe				
RISQUE NATUREL	risque naturel					
SANTE	santé	maladie				
SECURITE	sécurité	appareil	prévention			
		dispositif				
		matière	poussières			
		soutènement				
SOCIOLOGIE	sociologie	écologie	économie			
SOUTENEMENT	industrie minière > soutènement					
	soutènement	cadres&piles				
		élément				
TECHNIQUE MINIERE	technique minière	abattage				
TEMPS	temps					
TRANSPORT	les transports					
	transport	homme&matériel	accessoire	appareil	déplacement	
					manutention	
				appareil d'évacuation des charbons		
				câble		
				câble (partie de)		
				écartement des rails		
				engin de sûreté		

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
				immatriculation		
				infrastructure de transport		
				manoeuvre		
				moyen de transport		
				panier		
				plaque de chargement		
				sécurité		
				support des molettes		
				tuyau		
				r berline		
			action			
			chargement			
			convoyeur (partie de)			
			dispositif	échange de wagonnet		*matériel-transport*
				manoeuvre		*matériel-manœuvre*
				roulage		*roulage*
			dispositif accrochage des wagonnets			*matériel-transport et accrochage*
			entonnoir	stockage		*matériel-transport*
			infrastructure	rail		*rail*
			manoeuvre			*matériel-manœuvre*
			matériel d'évacuation			*matériel-transport*
			moyen			*matériel-transport*
			moyen de transport	cage		*cage*
				caractéristique		*-trop spécifique voir matériel-transport*
				changement de direction		*-trop spécifique voir matériel-transport*
				échelle		*-trop spécifique voir matériel-transport*
				locomotive		*locomotive*
				manoeuvre		*matériel-manœuvre*
				remonte pente		*télésiège*
				tonneau		*cage*
			moyen de transport (partie de)	barreau d'échelle		*-trop spécifique voir matériel-transport*
				sécurité		*barrière-sûreté*
			partie de rail			*rail*
			rails (partie de)			*rail*
			réservoir	stockage		*-trop spécifique, absent*
			système			*matériel-transport*
			système évacuation			*ventilateur ou treuil*

INDEXATION	Déclinaison de l'indexation (séparation par le signe _)					***Descripteur ThesoMines***
	1er niveau	**2ème niveau**	**3ème niveau**	**4ème niveau**	**5ème niveau**	
			voie			*matériel-transport*
			voie ferrée			*cavalier*
		minerai	moyen de transport			*cage*
	transport air comprimé					*matériel-aérage*
TRI	tri					*triage*
VETEMENT	vêtement	protection	eau			*vêtement*
		travail				*vêtement*
	vêtements	expression				*-trop spécifique, voir vêtement*
		vêtement de travail				*vêtement*

Liste des sigles et acronymes utilisés

ANR	Agence Nationale de la Recherche
ANMT	Archives Nationales du Monde du Travail
ARK	*Archival Resource Key*
BRGM	Bureau de Recherches Géologiques et Minières
BMU	Bassin Minier Uni
CAUE	Conseil d'Architecture, d'Urbanisme et de l'Environnement
CC	*Colon Classification*
CDD	Classification Décimale de Dewey
CDU	Classification Décimale Universelle
CERCHAR	Centre d'Études et Recherches des Charbonnages de France
CHM	Centre Historique Minier
CIDOC CRM	*CIDOC object-oriented Conceptual Reference Model*
CLADIC	Centre Liégeois d'Archives et de Documentation de l'Industrie Charbonnière
DC	*Dublin Core*
DOLCE	*Descriptive Ontology for Linguistic and Cognitive Engineering*
EDM	*Europeana Data Model*
EM	Employer
EP	Employé(s) pour
FRBR	*Functional Requirements for Bibliographic Records*
FRBRoo	*Functional Requirements for Bibliographic Records objet-oriented*
HBNPC	Houillères du Bassin du Nord et du Pas-de-Calais
IA	Intelligence Artificielle
INICHAR	Institut national de l'industrie charbonnière
KOS	*Knowledge Organizing System* ou *Knowledge Organization System*
KOP	*Knowledge Organizing Processes*
KO	*Knowledge Organization*
LCC	*Library of Congress Classification*
LOD	*Linked Open Data*
LIS	*Library and Information Science*
MT	Micro-Thésaurus
MBM	Mission Bassin Minier
OCR	*Optical Character Recognition*
OWL	*Ontology Web Language*
POC	Processus d'organisation des connaissances
RAMEAU	Répertoire d'Autorité Matière Encyclopédique Et Alphabétique Unifié
RC	Représentation des connaissances
RDF	*Resource Description Framework*
RI	Recherche d'Information
RTO	Ressources Terminologiques ou Ontologiques ; Ressources Termino-Ontologiques
RVM	Répertoire de Vedettes Matières
SEO	*Search Engine Optimization*
SI	Système d'Information
SKOS	*Simple Knowledge Organisation System*
SIC	Sciences de l'Information et de la Communication
SOC	Système d'Organisation des Connaissances
SFSIC	Société Française des Sciences de l'Information et de la Communication
TA	Terme Associé
TAL	Traitement Automatique des Langues
TALN	Traitement Automatique du Langage Naturel
TGIR	Très Grande Infrastructure de Recherche
TG	Terme Générique
TS	Terme Spécifique
UNESCO	*United Nations Educational, Scientific and Cultural Organization*
URI	*Uniform Resource Identifier*
W3C	*World Wide Web Consortium*
XML	*Extensible Markup Language*

Table des figures

Figure 1 : Représentation des thèmes périphériques et centraux de la science de l'information par (Le Coadic 1984) 10
Figure 2 : Schéma représentant la définition de l'OC issue de (Hjørland 2008) du point de vue particulier de la LIS disponible à l'adresse : https://cmapscloud.ihmc.us:443/rid=1WV9S58GL-1HMJC51-VV987J 18
Figure 3 : Schéma représentant la différence entre information et knowledge @gapinvoid 19
Figure 4 : Exemple de terminogrammes dans Motbis pour le descripteur *exploitation minière* 33
Figure 5: Représentation de l'arbre de Porphyre issue de (Delengaigne et Garcia 2018) 34
Figure 6 : Carte mentale Patrimoine minier en Europe issue de l'Encyclopédie Universalis © *Encyclopædia Universalis* 35
Figure 7 : Représentation heuristique des *Linked Open Data* à l'adresse https://lod-cloud.net/clouds/lod-cloud.svg 36
Figure 8 : Carte conceptuelle réalisée pour représenter la réduction du CIDOC CRM pour la classe *E11 Modification* (Cf. Représentation des modèles obtenus) 36
Figure 9 : Exemple d'un réseau sémantique (Ghio 1997, 47) 37
Figure 10 : Index sous forme de graphe conceptuel relatif à la visioconférence « Vidéochirurgie à Strasbourg » issu de (Yengui 2016, 159) 39
Figure 11 : « *Proposed taxonomy of KOS dimensions* » dans (Souza, Tudhope, et Almeida 2012, 189) 50
Figure 12 : Taxonomie des animaux présents dans l'univers minier, réalisée par l'auteure 56
Figure 13 : Nuage de mots-clés réalisé grâce à l'outil Voyant-Tool à partir de l'article de (Francis et Quesnel 2007) 60
Figure 14 : "*Small section of the CIDOC CRM class hierarchy*" (version 4.2.1) par Bernhard Haslhofer 68
Figure 15 : Diagramme représentant les enjeux des ontologies dans les thèses françaises en Information-Communication. Réalisation par l'auteure. 71
Figure 16 : Pyramide de la famille de modèle du CIDOC CRM (extensions) http://www.cidoc-crm.org/node/191 75
Figure 17 : Zone tampon des biens inscrits au patrimoine mondial de l'UNESCO disponible en ligne à https://whc.unesco.org/fr/list/1360/multiple=1&unique_number=1776 89
Figure 18 : Photographie d'une berline marquant l'activité minière à la fosse 9 d'Oignies de 1930 au 21 décembre 1990, fosse 2 d'Oignies 90
Figure 19 : Maquette de la mine réalisée par Martial Ansart, ancien mineur interrogé à Givenchy-en-Gohelle en juin 2018 94
Figure 20 : Schéma représentant une partie du patrimoine minier en surface et en souterrain lors de la première temporalité 94
Figure 21 : Schéma représentant la visibilité actuelle, avec un souterrain remblayé 95
Figure 22 : Illustration de la variété des objets du patrimoine, photographies issues de visites de sites miniers, du fond photographique du CHM de Lewarde et de la documentation Mission Bassin minier Patrimoine mondial. 96
Figure 23 : Diagrammes de quelques acteurs du patrimoine minier 100
Figure 24 : Capture de la fresque chronologique sur le site : *https://fresques.ina.fr/memoires-de-mines/fresque* 100
Figure 25 : Utilisation du moteur de recherche pour identifier des termes dans les notices descriptives des vidéos de la fresque Mémoires de Mines, ici requête du terme « lavage » présent dans quatre médias. 101
Figure 26 : Organisation thématique des documents vidéo de la fresque https://fresques.ina.fr/memoires-de-mines/pie 102
Figure 27 : Première page du plan de recherche des archives du CHM 103
Figure 28 : À gauche, carte de vœux de Sainte-Barbe. Image CHM. À gauche, Photographie d'identité d'ouvriers marocains arrivés au centre de Noyelles sous Lens, 12/01/1976, Druelle. Image CHM. 105
Figure 29 : Exemples de termes contenu dans Chorosz et Société de l'industrie minérale (1987) 107
Figure 30 : Locaux de la Mission Bassin minier sur le site du 9-9bis à Oignies 107
Figure 31 : Frise temporelle des documents correspondant au sujet « Mines » dans Nordnum 110
Figure 32 : classement conceptuel du thésaurus d'Auclaire (Auclaire 1990) 117
Figure 33 : Extrait du domaine sécurité du thésaurus du CLADIC en octobre 2018 118
Figure 34 : Fréquence des termes indexés par le vocabulaire de (B. Turpin et Goudaillier 2004) 124
Figure 35 : Nombre de termes indexés par le vocabulaire de (B. Turpin et Goudaillier 2004) par sous-thématique : le travail du mineur 125

Figure 36 : Structure des descripteurs sous « métier » dans ThesoMines sur Opentheso 127
Figure 37 : Alignement des concepts et construction des descripteurs du thésaurus 127
Figure 38 : Capture des candidats termes pour l'enrichissement du thésaurus sur Opentheso 130
Figure 39 : Représentation en arbre des classes de haut-niveau du CIDOC CRM (v.6.2.1) 133
Figure 40 : Représentation conceptuelle des classes et propriétés sélectionnées pour décrire le concept de commémoration 139
Figure 41 : Capture d'écran des propriétés de la classe *E5 Event* 139
Figure 42 : Nombre de cartes conceptuelles réalisée par classe sélectionnée 140
Figure 43 : Indexation de 29 tours de parole du corpus audiovisuel avec le terme *abattage* (EP *coupe*) 142
Figure 44 : Forme « abattage » dans un tour de parole transcrit d'Edmond 143
Figure 45 : Nombre de termes par ressources lexicales 145
Figure 46 : Page d'accueil de ThesoMines 148
Figure 47 : Les niveaux de spécificité de ThesoMines 149
Figure 48 : Classes du CIDOC CRM sélectionnées pour représenter et décrire le patrimoine minier. 150
Figure 49 : Instances de classes relevées dans un tour de parole de Jean (ancien mineur) 151
Figure 50 : Illustration d'une instanciation 152
Figure 51 : Extrait du graphe représentant l'instanciation du segment « Présentation de Jean » 153
Figure 52 : Instance ajoutée dans les connaissances partagées sur OKAPI et ajout d'un lien au descripteur « musique » de ThesoMines 153
Figure 53 : Graphe représentant l'instanciation de l'entretien avec l'ancien mineur Jean, en noir au centre l'instance E21 Person, Jean 154
Figure 54 : Zoom sur l'instanciation, Jean est au centre du graphe 155
Figure 55 : Extrait de représentation de l'instanciation des *E65 Creation* et *E7 Acivity* à partir d'un extrait d'entretien 156
Figure 56 : Représentation du recouvrement des SOC 160
Figure 57 : Recouvrement de TerminoMines et ThesoMines 161
Figure 58 : Pourcentage des descripteurs issus de TerminoMines dans chacune des branches de ThesoMines. 161
Figure 59 : OntoMines saturé 163
Figure 60 : Cinq tours de parole indexés par le descripteur *accrochage* et ses équivalents sur OKAPI 166
Figure 61 : Ambiguïté du descripteur *café* 166
Figure 62 : Appareil de sauvetage Fenzy – Photographies prises par l'auteure à la fosse 2 d'Oignies 173
Figure 63 : Descripteurs (surlignés en jaune) faisant référence à des acteurs dans ThesoMines 174
Figure 64 : Représentation d'un énoncé avec surlignage en jaune des références à ThesoMines 175
Figure 65 : Instances de *E21 Person* et *E74 Group* issues du témoignage de Jean 176
Figure 66 : Illustration de l'environnement des instances (losange violet) en lien avec l'instance nommée *Racing Club de Lens*. Schéma réalisé avec OntoGraph sur Protégé 178
Figure 67 : Cavalier inscrit à l'UNESCO. Extrait de la brochure Index éléments par commune à https://bassinminier-patrimoinemondial.org/wp-content/uploads/2017/05/Porter-a-connaissance-Interactif-1.pdf 181
Figure 68 : Connaissances « sites classés à l'UNESCO dans le Pas-de-Calais » et « Jean Vincent a joué au football à Auchel, ville du Pas-de-Calais » et « Influence de la fermeture des mines dans le Pas-de-Calais sur le départ de Jean vers le Nord » décrites par OntoMines. En orange : liens avec ThesoMines, en bleu : termes utilisés par Jean, en rouge : E53 Place. Schéma disponible à https://cmapscloud.ihmc.us:443/rid=1XC9DD50R-1TK40PZ-1VXFC8 184
Figure 69 : Vestiaire de la fosse d'Arenberg – Juin 2018 187
Figure 70 : Instances de *E22 Man-Made Object* issues du témoignage de Jean. Les flèches violettes représentent la relation d'instanciation, les bleues de hiérarchie. Schéma réalisé avec OntoGraph sur Protégé.. 188
Figure 71 : À gauche, liste des instances de E7 Activity issue du témoignage de Jean. À droite, environnement sémantique de l'instance *football*. Captures d'écran issues de Protégé avec OntoGraph. 190
Figure 72 : Termes contenus dans le témoignage de Jean reliés à leurs données terminologiques 193
Figure 73 : Extrait des données de TerminoMines et ThesoMines contenues dans le témoignage de Jean 194
Figure 74 : Extrait des données d'OntoMines, TerminoMines et ThesoMines contenues dans le témoignage de Jean. Disponible à https://cmapscloud.ihmc.us:443/rid=1XBQ4P7XD-1GV9CSL-1DTD4R 197
Figure 75 : Maquette représentant une cage et explications par un ancien mineur, Edmond à Arenberg. Document déposé sur la plateforme OKAPI. 198
Figure 76 : Instanciation de l'activité de destruction des *E18 Physical Thing* par héritage de classes 201
Figure 77 : Extrait du fichier SKOS de ThesoMines v1.0 204
Figure 78 : Alignement du concept puits avec un concept distant dans ThesoMines 205

Figure 79 : Extrait d'OntoMines, instance « avantage social » 207

Table des tableaux

Tableau 1 : Typologie des SOC selon (Zacklad 2018) 26
Tableau 2 : critères d'évaluation des SOC dans (Zacklad 2010) et (Zacklad 2018) 27
Tableau 3 : Exemples de structurations organisées par type, objet, mode et type de format issues d'un corpus de textes scientifiques 46
Tableau 4 : Termes français et anglais issus du livre blanc ISO 25964-1 pour le domaine Système d'organisation de concepts 50
Tableau 5 : Ontologies documentaires 73
Tableau 6 : Plan de recherche du CHM par groupes, compagnies et concessions 103
Tableau 7 : Extrait de la liste des ouvrages sur la thématique MINES bibliothèque G. Lefebvre Université de Lille 109
Tableau 8 : Critères de distinction des ressources lexicales 111
Tableau 9 : Domaines du plan de classement du CERCHAR, édition de 1976 114
Tableau 10 : Domaines du plan de classement de l'INICHAR, édition de 1965 114
Tableau 11 : Catégories thématiques du thésaurus du Service National de Géologie et Minière du Chili 117
Tableau 12 : Termes vs Concepts 123
Tableau 13 : Le concept *barrou* et ses réalisations dans la langue 123
Tableau 14 : Concept *barrou* 124
Tableau 15 : Indexation thématique non contrôlée du vocabulaire 125
Tableau 16 : Fréquence d'apparition de quelques termes dans le corpus 129
Tableau 17 : Fonctionnalités et utilisation du logiciel Opentheso 130
Tableau 18 : Instanciation des classes par des connaissances issues du patrimoine minier 134
Tableau 19 : Classes CIDOC CRM attribuées aux descripteurs du thésaurus 135
Tableau 20 : définition des entités spatiales et temporelles et les acteurs 136
Tableau 21 : traduction des classes de l'anglais vers le français et note sur la présence ou non dans la version 5.0.4 136
Tableau 22 : Classes du CIDOC CRM v.6.2.1 qui n'ont pas de traduction en français http://cidoc-crm.org/versions/cidoc_crm_v6.2.1_with_translations.html 137
Tableau 23 : Regroupement des descripteurs par classe 138
Tableau 24 : Données disponibles (ou potentiellement déposables) sur TerminoMines, ThesoMines et OntoMines 141
Tableau 25 : Extrait de TerminoMines 144
Tableau 26 : Extraction manuelle de termes dans les articles de *La Voix du Nord* et indexation. Termes non présents dans la terminologie. 144 termes 145
Tableau 27 : Extraction terminologique à partir d'un article de presse. Relevé de la langue, des équivalents, des définitions, de l'étymologie et indexation thématique – 25 termes 147
Tableau 28 : Nature des données des SOC 162
Tableau 29 : choix des propriétés associées aux classes *E21 Person* et *E74 Group* pour décrire les acteurs du patrimoine minier 174
Tableau 30 : Acteurs de type *club* dans une partie du corpus Europresse 178
Tableau 31 : Choix des propriétés associées à la classe *E53 Place* pour décrire les lieux du patrimoine minier 180
Tableau 32 : Instanciation des *E53 Place* dans la transcription du témoignage de Jean 181
Tableau 33 : Choix des propriétés associées à la classe *E52 Time-Span* pour décrire les périodes du patrimoine minier 185
Tableau 34 : Propriétés du modèle OntoMines qui diffèrent selon le type de concept pour la classe *E7 Activity* 189
Tableau 35 : Exemples de relations hiérarchiques dans ThesoMines 195
Tableau 36 : Alignement des champs de TerminoMines avec les catégories de TBX 206
Tableau 37 : Passage des concepts de ThesoMines à OntoMines 208

Table des matières détaillée

REMERCIEMENTS **2**

SOMMAIRE **4**

RÉSUMÉ ET MOTS-CLÉS **6**

ABSTRACT AND KEYWORDS **7**

REMARQUES PRÉLIMINAIRES **8**

INTRODUCTION **9**

PARTIE 1. CADRE THÉORIQUE DE L'ORGANISATION DES CONNAISSANCES **14**

CHAPITRE 1. ORGANISATION, REPRÉSENTATION ET MODÉLISATION 15

1.1 Organisation des connaissances (OC) *15*

1.1.1 Enjeux actuels et généraux de l'organisation des connaissances 22

1.1.2 Approches en organisation des connaissances 22

1.1.1 Systèmes d'organisation des connaissances (KOS) 25

1.2 Représentation des connaissances (RC) *27*

1.2.1 Représentation mentale des connaissances 28

1.2.2 Représentation graphique des connaissances 30

1.2.2.1 Domaine de la terminologie 31

1.2.2.2 Domaine de l'information-documentation 32

1.2.2.3 Représentations ontologiques 39

1.2.2.4 Systèmes de représentation des connaissances 39

1.2.3 Représentation globale des connaissances et web sémantique 40

1.2.4 Langage de représentation des connaissances 41

1.2.5 Représentation des connaissances sensibles 43

1.3 Modélisation des connaissances *43*

1.3.1 Importance de la structuration 45

1.3.2 Problème des ontologies 47

CHAPITRE 2. APPROCHES DOCUMENTAIRES DE L'ORGANISATION DES CONNAISSANCES 49

2.1 Systèmes d'organisation des connaissances (SOC) *49*

2.2 Types et hétérogénéité des SOC *50*

2.2.1 Vocabulaire contrôlé vs vocabulaire libre 51

2.2.2 Langages documentaires 51

2.2.2.1 Classification documentaire 53

2.2.2.2 Taxinomie 55

2.2.2.3 Thésaurus et terminologie 56

2.2.2.4 Répertoire de vedettes matières (RVM) 59

2.2.2.5 Autres langages documentaires 60

2.3 SOC et Web sémantique *61*

2.3.1 Interopérabilité des SOC 61

2.3.2 SOC culturels (cultural KOS) et web sémantique 62

CHAPITRE 3. APPROCHES ONTOLOGIQUES DE L'ORGANISATION DES CONNAISSANCES 63

3.1 Domaines d'étude des ontologies *64*

3.2 Types d'ontologies *67*

3.2.1 Ontologies de haut niveau 67

3.2.2 Ontologies de domaine ou de spécialité 68

3.3 Travaux actuels sur les ontologies dans le champ de l'OC *70*

3.4 Rôle des ontologies *71*

3.4.1 Ontologies pour la description, le partage, l'échange et la réutilisation de connaissances 71

3.4.2 Ontologies pour la recherche d'information 76

3.4.3 Construction de dictionnaires électroniques 76

3.4.4 Ontologies pour inférer de la connaissance 77

3.5 Éléments d'un modèle ontologique *77*

3.5.1 Concept 77

3.5.2 Relation ou propriétés 78

3.5.3 Axiomes, inférences et règles de raisonnement 79

3.6 Différentes approches de conception d'une ontologie de domaine *79*

3.6.1 À partir d'un glossaire et ses définitions 79

3.6.2 À partir de textes 80

3.6.3 À partir d'un thésaurus ... 81
3.6.4 Approche hybride ... 82

PARTIE 2. LE PATRIMOINE MINIER COMME DOMAINE DE CONNAISSANCE ... 84

CHAPITRE 4. DÉFINIR LE PATRIMOINE MINIER ... 86
4.1 Un patrimoine à plusieurs facettes ... 86
4.1.1 Patrimoine culturel, naturel, matériel et immatériel ... 86
4.1.2 Patrimoine industriel ... 88
4.2 Un patrimoine inscrit à l'UNESCO ... 88
4.2.1 Un « paysage culturel » ... 88
4.2.2 Un espace marqué par deux temporalités ... 89
4.3 Un patrimoine langagier en danger ... 91
4.3.1 Picard ou ch'ti, chtimi, cht'i mi, rouchi ? ... 91
4.3.2 Documentation et archivage du matériel linguistique ... 92
4.3.3 Vocabulaire de la mine : enjeux linguistiques ... 93
4.4 Un patrimoine souterrain à rendre visible ... 94
4.5 Un patrimoine documentaire hétérogène et dispersé ... 95
CHAPITRE 5. VERS UNE MÉTHODOLOGIE DE CONCEPTION DES SOC DU DOMAINE MINIER ... 99
5.1 Constitution d'un corpus documentaire hétérogène ... 99
5.1.1 Rencontres avec les acteurs détenteurs de documentation ... 99
5.1.1.1 Archives audiovisuelles : INA ... 100
5.1.1.2 Centre de documentation : CHM de Lewarde ... 102
5.1.1.3 Organisme scientifique : BRGM Hauts-de-France ... 105
5.1.1.4 Organisme associatif : Mission Bassin minier (MBM) ... 107
5.1.1.5 Bibliothèques de recherche physiques ou numériques ... 108
5.1.2 Identification des ressources lexicales pour définir le domaine ... 110
5.1.2.1 Création du corpus de référence ... 110
5.1.2.2 Création du corpus d'étude ... 112
5.1.3 Identification des ressources classificatoires pertinentes ... 113
5.1.3.1 Plans et cadres de classement ... 113
5.1.3.2 Index thématiques ... 115
5.1.3.3 Terminologies ... 115
5.1.3.4 Thésaurus documentaires ... 116
5.1.4 Autres ressources ... 118
5.1.4.1 Ressources audiovisuelles ... 118
5.1.4.2 Ressources supplémentaires ... 119
5.1.5 Outils utilisés ... 120
5.2 Des ressources lexicales à la terminologie de la mine ... 120
5.2.1 Plusieurs terminologies, plusieurs vocabulaires miniers ... 120
5.2.2 Un langage spécifique à une activité et à un territoire ... 121
5.2.3 Des noms, des verbes, des expressions ? ... 122
5.2.4 Terme vs concept et relations sémantiques ... 122
5.3 De la terminologie minière au thésaurus ... 124
5.3.1 Structuration : relation hiérarchique ... 124
5.3.2 Relation associative ... 127
5.3.3 Termes descripteurs et non-descripteurs : relation d'équivalence ... 128
5.3.4 Enrichissement du thésaurus ... 129
5.3.5 Outil utilisé ... 130
5.4 Du thésaurus au modèle ontologique ... 131
5.4.1 Choix du modèle ontologique du CIDOC CRM ... 132
5.4.2 Appropriation du modèle ontologique CIDOC CRM ... 132
5.4.3 Confrontation avec d'autres travaux utilisant le modèle ... 134
5.4.1 Classes et propriétés représentatives du patrimoine minier ... 134
5.4.1.1 Passage des descripteurs de thésaurus aux classes du modèle ... 134
5.4.1.2 Description des entités nommées avec le CIDOC CRM ... 135
5.4.2 Traduction des classes et propriétés ... 136
5.4.3 Des descripteurs du thésaurus à l'ontologie ... 137
5.4.4 Représentation des modèles obtenus : cartes conceptuelles ... 138
5.4.5 Enregistrement des données ... 140
5.5 Du modèle ontologique réduit à l'ontologie de domaine instanciée ... 142
CHAPITRE 6. PRÉSENTATION DES TROIS SOC ... 144
6.1 Une terminologie de la mine : TerminoMines ... 144
6.2 Un thésaurus du patrimoine minier : ThesoMines ... 147

6.3 Une ontologie de la mine : OntoMines. ... *149*
6.3.1 Classes et propriétés du CIDOC CRM sélectionnées pour décrire le patrimoine minier ... 150
6.3.2 Représentation et formalisation des données ... 150
6.3.3 Extrait de retranscription instancié ... 155

PARTIE 3. ENJEUX CONCEPTUELS DES SOC DU PATRIMOINE MINIER ... 158

CHAPITRE 7. COUVERTURE INFORMATIONNELLE ... 159
7.1 Recouvrement et différenciation conceptuels ... *159*
7.2 Évaluation de la saturation ... *162*
7.2.1 Saturation absolue des SOC ... 162
7.2.2 Point de vue sur le domaine ... 163
7.3 Processus d'enrichissement : démarche évaluation et perfectionnement ... *165*
7.3.1 Test par indexation automatique d'un corpus ... 165
7.3.2 Test de la souplesse de la structure ... 167
CHAPITRE 8. FINESSE DE DESCRIPTION OU GRANULARITÉ ... 170
8.1 Les entités nommées ... *172*
8.1.1 Les acteurs ... 172
8.1.2 Les entités spatiales ... 179
8.1.3 Les entités temporelles ... 184
8.2 Les objets matériels ... *186*
8.3 Les activités ... *188*
CHAPITRE 9. EXPRESSIVITÉ DES FORMALISMES ... 192
9.1 La structure de la terminologie ... *192*
9.2 Les relations hiérarchiques du thésaurus ... *193*
9.3 Le réseau de l'ontologie ... *196*
9.3.1 Une couverture sémantique aussi générique que spécifique ... 196
9.3.2 Limites sémantiques du modèle et ajustements nécessaires ... 198
CHAPITRE 10. INTEROPÉRABILITÉ ... 203
10.1 Interopérabilité syntaxique ... *203*
10.1.1 ThesoMines et SKOS ... 203
10.1.2 TerminoMines et TBX ... 205
10.1.3 OntoMines et OWL/RDFS ... 206
10.2 Interopérabilité sémantique ... *207*
10.2.1 Entre SOC ... 207
10.2.2 Entre les ressources ... 208
10.2.3 Entre les domaines ... 208

CONCLUSION ... 211

BIBLIOGRAPHIE ... 219

Bibliographie scientifique ... *219*
Ressources et corpus utilisés ... *234*

ANNEXES ... 235

Annexe 1 : Extrait de retranscription de l'entretien effectué avec l'ancien mineur Jean à Arenberg ... 235
Annexe 2 : Extrait de retranscription de l'entretien effectué avec l'ancien mineur Martial à Givenchy-en-gohelle ... 245
Annexe 3 : Hiérarchie des classes issue de la version 5.0.4 du CIDOC CRM, instances relevées du patrimoine minier et note sur les classes selectionnées dans le projet TECTONIQ ... 255
Annexe 4 : Classes et propriétés sélectionnées pour représenter le patrimoine minier et liens vers schémas d'aide à l'instanciation ... 263
E1_CRM Entity ... 263
E7_Activity ... 263
E9_Move ... 267
E11_Modification ... 267
E13_Attribute Assignement ... 268
E65 Creation ... 268
E83_Type Creation ... 268
E64_End of existence, E69_Death et E6_Destruction ... 269
E39_Actor, E21_Person et E74 Group ... 270

E18_Physical Thing *270*
E24_Physical_Man-Made_Thing *271*
E22_Man-Made_Object *271*
E25_Man-Made_Feature *274*
E73_Information Object, E31_Document et E33_Linguistic Object *276*
E27_Site *276*
E92_Spacetime_Volume *277*
E52_Time-Span et E50 Date *278*
E53_Place *278*
E3_Condition State *279*
E72_Legal Object et E30_Right *279*
E54_Dimension *279*
E28_Conceptual Object *280*
Annexe 5 : Liste des vedettes matières RAMEAU identifiées permettant d'indexer des documents traitant d'aspects du domaine minier et alignement avec les descripteurs de ThesoMines *281*
Annexe 6 : Liste des indexations thématiques de la terminologie et alignement avec les descripteurs de ThesoMines *283*

LISTES ET TABLES **298**

Liste des sigles et acronymes utilisés *298*
Table des figures *299*
Table des tableaux *301*
Table des matières détaillée *302*