Conception d'ontologies formelles : une approche par élicitation de connaissances

Ouassila Labbani Narsis, Christophe Nicolle

CIAD UMR 7533, Université de Bourgogne, UB, F-21000 Dijon, France
{ouassila.narsis,cnicolle}@u-bourgogne.fr

RÉSUMÉ. *La construction d'ontologies formelles reste une démarche complexe pour les entreprises. Dans la littérature, cette construction se base sur les connaissances techniques et les spécialités des experts du domaine, sans plus de détails sur les méthodologies employées. Les éventuels problèmes de désaccords entre experts, d'expression des connaissances implicites rarement verbalisées, de qualification du résultat par des cas d'usages, ou simplement d'adhésion du groupe d'experts, restent actuellement sans réponse. Cet article propose une démarche méthodologique basée sur l'élicitation des connaissances pour la conception d'ontologies formelles, consensuelles et partagées. L'approche proposée est testée expérimentalement sur des projets de collaborations industrielles dans le domaine de la manufacture (associant des sources de connaissances d'entreprises multinationales) et dans le domaine de la viticulture (associant des connaissances explicites et des savoirs implicites acquis par l'observation).*

ABSTRACT. *Building formal ontologies remains a complex process for companies. In the literature, this process is based on the technical knowledge and expertise of domain experts, without further details on the used methodologies. Possible problems of disagreements between experts, expression of implicit knowledge related to high level know-how rarely verbalized, qualification of results by use cases, or simply adhesion of the group of experts, remain currently unsolved. This paper proposes a methodological approach based on knowledge elicitation for the conception of formal, consensual, and shared ontologies. The proposed approach is experimentally tested on industrial collaboration projects in the field of manufacturing (associating knowledge sources from multinational companies) and in the field of viticulture (associating explicit knowledge and implicit knowledge acquired through observation).*

MOTS-CLÉS : *Ingénierie de connaissances, Élicitation de connaissances, Gestion des connaissances, Conception d'ontologie.*

KEYWORDS: *Knowledge engineering, Knowledge elicitation, Knowledge management, Ontology design.*

---

1. Introduction et motivation

La transformation numérique représente un défi technologique majeur pour les entreprises. Elle se traduit par l'introduction de nouveaux outils et technologies trop souvent en rupture avec les usages. Cette transformation s'accompagne souvent d'une modification de l'organisation, des processus, mais aussi du rapport du salarié avec sa fonction. Ce dernier point peut sérieusement affecter la préservation et la transmission des savoir-faire acquis par expérience depuis plusieurs années. Le savoir-faire d'une entreprise est sa principale richesse et la captation numérique de cette richesse devient un enjeu stratégique. Cette captation est un processus articulé en plusieurs étapes : il faut identifier les savoir-faire, les recueillir, les homogénéiser, les combiner puis les formaliser. Cette dernière étape permet d'obtenir une description numérique compréhensible par l'humain et sans ambiguïté pour la machine. Dans ce processus, l'utilisation de la modélisation sémantique joue un rôle très important, d'où l'usage de formalismes et de langages basés sur la logique tels que les `ontologies`. En informatique, une ontologie désigne une description formelle et explicite des concepts et de leurs relations. Décrite en logique de description, son rôle est de capturer les connaissances du domaine de manière générique et d'en fournir une compréhension partagée et communément acceptée (Maedche *et al.*, 2000). La construction d'ontologies permet d'une part d'atteindre un consensus dans la compréhension commune des connaissances, et d'autre part, de faciliter la mise en place d'un modèle formel représentatif et flexible, indispensable à la description numérique du savoir-faire d'une entreprise. Néanmoins, créer une ontologie est un processus complexe. Au-delà d'une compréhension du domaine concerné et des aspects techniques de l'ingénierie ontologique, il requiert une volonté de collaboration qui dépasse les frontières disciplinaires (Kanza *et al.*, 2018). Malgré les efforts pour le développement de méthodologies et outils d'aide à la construction d'ontologies (Gruber, 1995 ; Jones *et al.*, 1998 ; Missikoff, Schiappelli, 2005), la démarche reste empirique sans méthode normalisée. Les approches existantes ne prennent pas en compte les enjeux humains, comme la gestion des savoirs implicites, la collaboration et la gestion des désaccords entre experts, ou encore l'adhésion des collaborateurs aux évolutions technologiques. Avant d'être un problème de formalisation, construire une ontologie est tout d'abord un problème de conception et de modélisation qui doit s'effectuer à partir de la description linguistique et informelle des savoir-faire des experts et des connaissances du domaine. Pour répondre à cet enjeu, nous proposons une démarche permettant de construire des ontologies formelles basées sur les savoir-faire. Nous montrons l'apport de l'élicitation de connaissances dans cette démarche, et comment structurer et formaliser les connaissances afin d'uniformiser l'hétérogénéité sémantique pour construire un espace numérique d'expression et de partage des connaissances.

2. Élicitation de connaissances pour le recueil des savoir-faire

La problématique d'acquisition et de formalisation des connaissances est au centre des recherches menées dans le domaine de l'*ingénierie de connaissances* (Studer *et al.*, 1998a). Cette discipline tire ses origines de l'intelligence artificielle et de la psy-

chologie cognitive. Elle propose des concepts, méthodes et techniques permettant d'acquérir, de modéliser, et de formaliser les connaissances d'un domaine dans un objectif d'opérationnalisation, de structuration et de communication. En ingénierie de connaissances, l'*élicitation de connaissances* correspond au processus permettant d'aider un expert métier à exprimer et formaliser ses connaissances dans un objectif de les conserver, les utiliser et les transmettre. C'est un long processus qui doit être adapté selon la nature des connaissances à acquérir, le domaine d'application, et les raisonnements à mettre en œuvre. C'est la phase la plus coûteuse et la plus délicate dans un processus de construction d'un système à bases de connaissances.

En milieu industriel, recueillir les savoir-faire est une étape importante dans la construction de solutions numériques, et nécessite un accès direct aux connaissances et à l'expertise humain. De nombreuses techniques d'élicitation de connaissances ont été proposées dans la littérature (Gavrilova, Andreeva, 2012 ; Iqbal, Suaib, 2014 ; Khan[1] *et al.*, 2014). La figure 1 représente une taxonomie des techniques les plus utilisées (Chen, 2012). La classification proposée est basée sur la nature des techniques selon les types de communication utilisés par l'analyste pour obtenir les connaissances des experts. Ces techniques d'extraction de connaissances sont pour la plupart issues du domaine de la psychologie du travail. Certaines sont dites *directes*, et consistent à interroger l'expert et à le faire s'exprimer oralement, comme les interviews. Tandis que les autres, dites *indirectes*, sont fondées sur l'interprétation des éléments recueillis, capturés indirectement à partir d'autres sources, comme pour l'analyse documentaire.

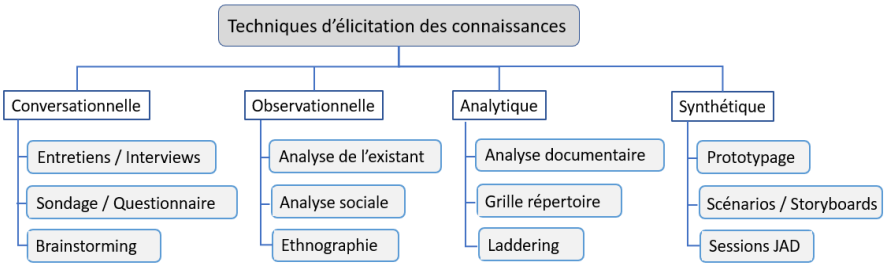

FIGURE 1. *Taxonomie des techniques d'élicitation de connaissances*

3. Enjeux des techniques d'élicitation de connaissances

L'élicitation de connaissances est une étape critique dans le processus de développement des systèmes à base de connaissances. Il est donc impératif d'appliquer les techniques appropriées pour mener à bien ce processus selon la nature des connaissances et du raisonnement à mettre en œuvre. Il existe évidemment des combinaisons de ces techniques où différentes approches sont utilisées à différentes étapes du

processus d'acquisition des connaissances. Néanmoins, le manque de méthodologies précises pour le choix et l'adaptation de ces techniques, ainsi que le grand écart entre le domaine de l'expert et celui de l'analyste rendent cette tâche difficile et source de plusieurs erreurs.

La conversation est l'une des formes d'interaction sociale les plus répandues. Les techniques conversationnelles sont donc très utiles pour l'acquisition des connaissances. Elles permettent également de découvrir les opinions, les sentiments et les objectifs des différents individus. Cependant, ces techniques demandent beaucoup de travail, comme pour l'organisation des réunions ou la production et l'analyse des transcriptions à partir des enregistrements (Christel, Kang, 1992 ; Sharma, Pandey, 2014). De plus, leur efficacité dépend fortement du comportement et de l'attitude des experts et de l'analyste où le contexte de la conversation joue un rôle très important (Byrd *et al.*, 1992).

Les techniques observationnelles sont très pratiques lorsque l'analyste manque d'expérience dans le domaine, et peuvent être utilisées pour la collecte de connaissances tacites liées à la pratique. Néanmoins, tout comme les techniques conversationnelles, les techniques d'observation sont également difficiles à maîtriser et nécessitent beaucoup de temps et d'organisation (Hudlicka, 1996).

La réutilisation de connaissances déjà disponibles permet de gagner du temps et de réduire les coûts. Les techniques analytiques sont alors utilisées pour la collecte des connaissances existantes qui ne sont pas directement exprimées, et provenant de sources différentes. Cependant, ces techniques sont principalement basées sur des données empiriques et de la documentation. Elles font appel à des connaissances antérieures et peuvent restreindre la vision globale du système. La possibilité de reproduire des erreurs est donc une menace constante pour ce type de techniques, et il ne faut les utiliser que dans des situations particulières et bien maitrisées.

Il est important de souligner que les techniques présentées sont basées sur l'hypothèse implicite que les experts et l'analyste sont coopératifs et sincères, que les experts sont disposés à partager leurs connaissances, et que l'analyste possède l'expérience et les compétences nécessaires pour mener à bien une session d'élicitation. Cependant, dans la réalité, l'élicitation de connaissances est un processus d'interaction sociale complexe qui doit faire face à plusieurs difficultés :

– **Réticence de l'expert.** Le processus d'acquisition de connaissances met l'expert dans des conditions différentes auxquelles il est habitué. Il peut être donc réticent à ce processus, ou agir différemment, perturbé par la présence de l'analyste.

– **Explicitation des connaissances implicites.** Certaines connaissances ne sont pas directement verbalisables par les experts, et sont difficilement accessibles et communicables.

– **Sélection et accès aux connaissances génériques.** Ces connaissances correspondent à un domaine particulier et sont évidentes pour les experts du même domaine. Elles sont indispensables au raisonnement, peuvent être rendues explicites, mais sont rarement exprimées.

– **Distorsions des connaissances expertes.** L'analyste est souvent étranger au domaine de l'expert. Un décalage peut alors se créer entre les connaissances de l'expert et celles transcrites du fait d'une mal compréhension des propos de l'expert.

– **Hétérogénéité sémantique.** Plusieurs experts du même domaine sont concernés par le processus d'élicitation. Il n'est donc pas exclu de se retrouver dans des situations où les experts ne sont pas d'accord. L'expert peut aussi manquer de précision dans l'expression de ses connaissances ou se tromper, ce qui peut engendrer des incohérences ou des ambigïtés sémantiques.

À ces difficultés s'ajoute le fait que la plupart des techniques d'élicitation de connaissances sont coûteuses, très complexes, et doivent faire face à une réticence générale, particulièrement dans l'industrie. De plus, il n'existe pas de méthodologies claires pour l'application de ces techniques qui sont souvent effectuées d'une manière ad-hoc, sans un processus prédéfini.

4. Élicitation de connaissances basée sur la didactique professionnelle

L'un des défis majeurs posés en élicitation de connaissances est de tendre vers le recueil le plus exhaustif possible des connaissances métiers et des exigences de la solution cible. Il est donc nécessaire de combiner plusieurs techniques afin de bien comprendre le domaine, structurer et adapter les entretiens, faire face aux éventuelles ambiguïtés et désaccords entre les experts, et assurer leur adhésion aux représentations numériques résultantes. La réussite d'un processus d'élicitation de connaissances dépend fortement de la méthodologie appliquée et des solutions mises en place pour faire face aux éventuelles difficultés.

Comme discuté précédemment, les techniques d'élicitation se reposent essentiellement sur la participation des experts métiers. De ce fait, la volonté de ces derniers à partager leurs connaissances et leur adhésion au projet jouent un rôle principal dans la réussite de l'opération. De plus, la nature des interactions humaines et sociales que les experts du domaine peuvent avoir avec leur interlocuteur peuvent fortement affecter leur comportement et l'expression de leur savoir-faire.

Pour répondre à ces verrous, nous avons développé, en collaboration avec des chercheurs dans le domaine des Sciences Humaines et Sociales, une approche d'élicitation de connaissances inspirée de la *didactique professionnelle*, et basée sur les apports de la psychologie du travail et du développement (Pastré, 2002). Notre approche est une combinaison des techniques conversationnelles, observationnelles et analytiques, et peut être adaptée selon le domaine d'application, la nature des connaissances à collecter, et les besoins du système à développer.

Notre démarche cible les connaissances implicites et explicites. Elle prend en compte l'état psychologique des experts lors des entretiens afin de garantir la clarté, la complétude et la cohérence des propos, et de résoudre tout conflit apparent. Elle permet de répondre à plusieurs problématiques d'élicitation de connaissances, comme

l'explicitation des connaissances implicites, l'adhésion des experts métier, et la résolution d'éventuels conflits liés à l'hétérogénéité sémantiques des connaissances. Les principales phases de notre approche sont présentées dans la figure 2.

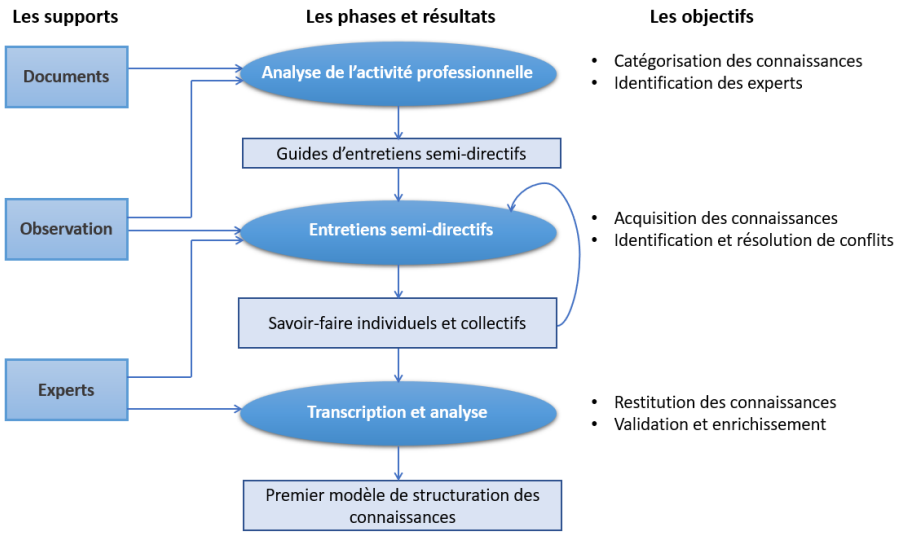

FIGURE 2. *Approche d'élicitation de connaissances pour le recueil des savoir-faire*

4.1. Analyse de l'activité professionnelle

Inspirée des approches analytiques, cette phase consiste à étudier les documents internes de l'entreprise et, dans certains cas, réaliser un état de l'art sur le domaine afin de collecter les connaissances génériques (réglementation, normes, processus, etc.). Ces connaissances sont indispensables au raisonnement, et ne sont toujours pas évidentes pour l'analyste. Il faut donc prendre conscience de leur existence et leur importance afin de les collecter et de les intégrer dans le système à développer.

Combinée à cette méthode analytique, nous proposons de réaliser une phase d'observation pré-entretien afin d'acquérir une meilleure compréhension du domaine d'application et des processus existants. L'objectif est de faire émerger un référentiel des connaissances et des modalités nécessaires pour accomplir une tâche précise. La finalité de cette phase est de construire, à partir de sources d'information hétérogènes (documents internes, état de l'art, observation), le dictionnaire des données de l'entreprise permettant de maitriser une partie du vocabulaire métier du domaine indispensable pour préparer la phase des entretiens et faciliter la communication.

À l'issue de cette phase, le dictionnaire de données, le glossaire métier, ainsi que la liste des processus et des experts concernés par la collecte des savoir-faire, sont pré-

sentés aux experts métier pour validation. Pour cela, nous utilisons généralement des plateformes de collaboration de type Wiki [1] afin de faciliter l'échange avec les experts. Cette phase permet d'assurer une compréhension commune et partagée du vocabulaire et des besoins du système à développer. Elle est indispensable dans un processus d'élicitation pour mieux préparer et définir les guides d'entretien et gagner en temps et en efficacité. Notre expérience dans ce domaine a montré une bonne recevabilité de la démarche et une forte motivation et collaboration par les experts métier.

4.2. Entretien semi-directif

Cette phase consiste à extraire la connaissance par la mise en place d'entretiens adaptés selon les besoins identifiés suite à la phase d'analyse. L'objectif est de définir des guides d'entretien adaptés, d'anticiper les temps nécessaires pour la réalisation de ces entretiens, et de définir en amont le planning selon la disponibilité des experts.

Dans les techniques conversationnelles, la conduite d'entretien est l'une des pratiques de collecte de connaissances les plus utilisées. Elle peut être directive, non-directive, ou semi-directive (Chevalier, Meyer, 2018). Dans notre approche, nous utilisons les entretiens semi-directifs basés sur un dialogue continu. Pour cela, un guide d'entretien est réalisé au préalable avec des questions directives afin de lister les points à aborder. Il se poursuit, selon la réponse de l'expert, par un ensemble de questions libres laissant la liberté à l'expert de s'exprimer. Dans notre approche, les guides d'entretien sont définis en partenariat avec des experts en psychologie de travail et en Sciences Humaines et Sociales afin de bien prendre en compte l'état psychologique des experts lors des entretiens, les aider à mieux exprimer leurs pensées, et faire face aux éventuels soucis de réticence.

Dans notre approche, une première étape d'entretiens individuels est réalisée. Elle permet de recueillir l'information générale concernant l'exercice des savoir-faire de manière individualisée. Dans cette phase d'entretien, les techniques d'observation entrent également en jeu lorsque la communication verbale devient insuffisante pour la collecte des connaissances tacites. Elles permettent également la vérification et la validation des connaissances recueillies.

Des entretiens collectifs pourraient être réalisés si l'équipe le juge pertinent. Ceux-ci seront effectués après analyse des entretiens individuels, et permettent de traiter des thèmes spécifiques, souvent relatifs à des sujets de confusion ou de désaccord entre experts. L'objectif de ce travail collaboratif est de lever l'ambiguïté et d'uniformiser les connaissances collectées. Dans notre approche, ce processus est réalisé par des analystes possédant des compétences interpersonnelles pour aider à établir un consensus entre des groupes hétérogènes d'experts. Ces compétences sociales sont aussi importantes que les techniques utilisées dans le processus d'élicitation.

---

1. Un wiki est une application web qui permet la création, la modification et l'illustration collaboratives de pages à l'intérieur d'un site web. Nous utilisons dans notre approche Semantic MediaWiki https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki

4.3. Transcription et analyse

La transcription des entretiens est la restitution des propos de l'expert en vue d'une analyse (Rioufreyt, 2016). Elle doit être fidèle, compréhensible et complète. Ainsi, pour garantir la clarté et l'uniformité des transcriptions, nous proposons la construction d'un modèle général et structuré (*template*) regroupant l'ensemble des métadonnées à associer pour chacun des entretiens transcrits. La mise en place d'un template commun pour la restitution aide également à la structuration et l'organisation des connaissances collectées.

Après la transcription sous forme textuelle des entretiens (enregistrés en audio), un volume important de propos des experts est collecté. Pour traiter le résultat, nous procédons à une *analyse thématique catégorielle* (Dany, 2016), une technique en sciences humaines et sociales permettant de décrire et de tirer des conclusions sur les caractéristiques d'une communication. Elle met l'accent sur l'identification, l'analyse, et l'interprétation des patterns de sujets ou thèmes au sein des textes examinés, en répondant aux questions basiques de type *Qui*, *Quoi*, *Quand*, *Comment*, etc.

Des analyses complémentaires, comme le calcul d'une Binomiale (théorème de Bernoulli, 1713) (Droesbeke *et al.*, 2002), peuvent être également réalisées selon les besoins du projet. Les avantages de ces méthodes d'analyse comprennent notamment la mise en évidence de points importants manquants pouvant servir de point de développement, d'approfondissement et de discussion.

L'analyse des entretiens permet également la collecte de règles logiques en lien avec le savoir-faire afin de les coder et les intégrer dans le modèle formel. Ces règles sont directement extraites du corpus d'entretiens, et correspondent à des phrases de type conditionnel comme : *"Si... alors"*, *"Quand..."*, *"Il faut..."*.

Les connaissances collectées et le résultat de l'analyse sont restitués et présentés aux experts pour validation avant formalisation et transformation numérique de ces connaissances. Par souci de déontologie, la restitution est anonyme, et aucun entretien n'a été intégralement transmis hormis à l'équipe de recherche.

5. Modélisation formelle des connaissances

Une ontologie est une représentation formelle de la connaissance. Elle est définie comme une spécification formelle et explicite d'une conceptualisation partagée d'un domaine d'intérêt (Studer *et al.*, 1998b). Elle fournit une structure commune et partagée des concepts, leurs propriétés et leurs relations, et permet la représentation des connaissances d'un domaine sous une forme interprétable par la machine. Par conséquent, une ontologie est destinée à être utilisée comme un composant logiciel dans des systèmes informatiques répondant à des objectifs opérationnels différents.

Au cours des dernières années, il y a eu un réel intérêt pour l'application des ontologies dans plusieurs domaines afin de résoudre des problèmes de modélisation liés à

la gestion de connaissances, tels que l'intelligence artificielle, la linguistique, le web sémantique, l'intégration de bases de données hétérogènes, etc.

De nombreuses méthodologies de construction d'ontologies ont été proposées dans la littérature. Elles se distinguent selon le type d'ontologie à développer (générique ou du domaine) et des données en entrée (corpus de textes, bases de connaissances, schémas de base de données, etc.). La plupart des méthodologies existantes proposent des recommandations et directives générales sous forme de guides à respecter dans le processus de développement (Gruber, 1995 ; Gómez-Pérez *et al.*, 1995 ; Noy *et al.*, 2004). Certaines méthodologies proposent de suivre des phases de construction inspirées des processus de développement en entreprise (Gruninger, Fox, 1996), tandis que d'autres méthodologies plus complètes tentent de combiner plusieurs approches pour construire une démarche unifiée (Uschold, 1996 ; De Nicola *et al.*, 2005).

Il est généralement admis que, pour être efficace et économiquement faisable, les ontologies doivent être développées de manière collaborative. Une première méthodologie de développement collaboratif d'ontologies a été proposée par Holsapple *et al.* (Holsapple, Joshi, 2002). Depuis, Construire des ontologies de manière collaborative est devenu un paradigme central définissant l'*ingénierie d'ontologie collaborative*, et plusieurs méthodologies ont été proposées dans le domaine (Simperl, Luczak-Rösch, 2014 ; Hayes *et al.*, 2005). Cette catégorie se base sur le fait qu'une ontologie doit faire l'objet d'un consensus et être acceptée par la communauté du même domaine. Elle fait intervenir plusieurs personnes qui doivent travailler ensemble afin de créer une compréhension commune et partagée de la structure et du sens des concepts de l'ontologie.

6. Enjeux des méthodologies de construction d'ontologies

Malgré les efforts et le nombre croissant de travaux sur la construction d'ontologies (Fernández-López, 1999 ; Pinto, Martins, 2004 ; Gómez-Pérez *et al.*, 2004), nous constatons que les méthodologies proposées sont souvent limitées à des recommandations et ne constituent pas une méthodologie complète. Elles sont développées de manière indépendante et se limitent souvent à des études de cas pour le développement d'une seule ontologie liée à un projet particulier. Le principal obstacle à l'application de ces méthodologies est qu'il n'y a pas d'indication claire sur le degré de généralité de ces techniques et méthodes individuelles. Par conséquent, il est difficile de décider quelles méthodologies sont susceptibles de s'appliquer pour un domaine particulier.

Impliquer différents acteurs et experts du domaine dans le processus de construction d'ontologie est l'ambition de l'ingénierie d'ontologie collaborative. Elle vise à construire une ontologie riche et complète en impliquant plusieurs acteurs ou groupes. Néanmoins, elle est principalement basée sur l'hypothèse qu'un compromis est toujours possible en cas de divergence d'opinions et de connaissances, ce qui ne reflète toujours pas la réalité. Les groupes de participant dans le processus sont généralement géographiquement décentralisés et ayant des domaines de compétences variés, rendant ainsi difficile la définition d'un consensus qui doit se faire de manière progressive et

asynchrone. Euzanet (Euzenat, 1995 ; 1996) a identifié un certain nombre de problèmes liés à la construction d'ontologies dans un contexte distribué, comme la gestion des interactions et la communication entre les différentes personnes, le contrôle de l'accès aux données, et la détection et la correction d'erreurs. D'autres limites de la construction d'ontologie collaborative sont également discutés dans (Simperl, Luczak-Rösch, 2014). Les méthodologies existantes sont souvent spécifiques à des projets particuliers, et sont difficiles à généraliser.

De manière générale, les approches existantes de construction d'ontologies restent empirique, sans méthode normalisée. Elles ne prennent pas en compte certains aspects humains, ne proposent pas de solutions pour assister et orienter les experts dans l'expression de leur savoir-faire, et ne font pas face aux réticences et appréhensions potentielles face à cette transformation technologique. Pour résoudre ce problème, nous proposons dans la section suivante une méthodologie collaborative pour le développement d'ontologies centrée sur l'expert et basée sur la technique d'élicitation de connaissances que nous avons proposée dans la section 4.

7. Du savoir-faire à l'ontologie formelle

Une fois que les savoir-faire sont clairement identifiés, il est nécessaire de les exprimer dans un langage compréhensible par la machine. Le passage d'une description informelle à une modélisation formelle des connaissances est un processus complexe et risqué. Afin de répondre à cette problématique, nous proposons une démarche pour structurer, modéliser et formaliser les connaissances. La figure 3 en représente les principales étapes. Après l'élicitation des connaissances pour le recueil du savoir-faire, une structuration et une modélisation semi-formelle sont réalisées, suivi d'une formalisation sémantique de ces savoirs avec intégration des ontologies existantes du domaine. Notre proposition est une hybridation des deux principales approches utilisées en in-

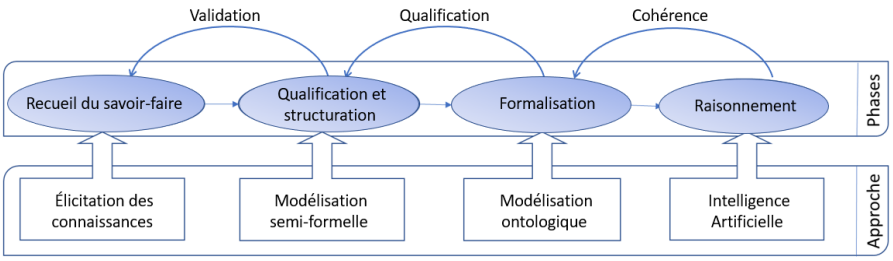

FIGURE 3. *Approche de formalisation des savoir-faire*

génierie de connaissances pour la construction de modèles. La première correspond à une démarche ascendante basée sur les techniques d'extraction et d'identification des connaissances, tandis que la deuxième est une démarche descendante qui privilège la réutilisation et l'adaptation de modèles existants (Aussenac-Gilles *et al.*, 2012).

7.1. Structuration et modélisation semi-formelle

L'objectif de cette phase est de passer d'une représentation informelle des savoir-faire, exprimés en langage naturel, à une structuration cohérente et une modélisation semi-formelle sous forme de cartes mentales et de diagrammes UML (Unified Modeling Language) (OMG, 2017). Ce standard de modélisation d'un domaine, selon le paradigme objets, propose un ensemble de diagrammes permettant l'expression des différentes facettes d'un savoir-faire : la définition des éléments statiques de la connaissance, la définition des éléments dynamiques associés (les usages) et une représentation des contraintes simples associée aux usages. Cette étape de modélisation est basée sur les résultats de la phase d'analyse et de restitution des connaissances présentée dans la section 4. Elle est réalisée en collaboration avec les experts métiers, qui sont les seuls à pouvoir valider les modèles proposés. La modélisation UML est principalement utilisée pour assurer la bonne compréhension des savoir-faire et les attentes des collaborateurs. Elle réduit considérablement le gap entre la description informelle des savoirs fournie par les experts, et leur formalisation numérique qui doit être compréhensible par la machine. De plus, UML est basé sur une représentation graphique facilement compréhensible par les participants, et largement utilisé en milieu industriel. Dans notre approche, le modèle UML proposé doit être suffisamment expressifs pour être compris par les experts du domaine non-informaticiens et suffisamment contraint pour correspondre à une logique informatique. Il permettra ainsi de mener des échanges entre les différents intervenants pour obtenir une modélisation validée par tous.

Nous utilisons dans notre approche les diagrammes de classes UML permettant de mettre en avant les concepts du domaine (classes), leurs propriétés (attributs) et leurs relations (association). Les diagrammes construits doivent être validés par les experts du domaine et révisés si nécessaires afin de représenter de manière cohérente et complète les connaissances recueillies lors de la phase d'élicitation de connaissances. Le langage UML est aussi maîtrisé par les personnels informatiques des entreprises, cela nous permet de les intégrer dans la démarche de transition numérique à partir de cette étape. Ils auront la charge, si ce n'est d'implémenter les modèles finaux présentés, au moins d'interconnecter la solution au système d'information numérique et de développer des services d'exploitations de ces connaissances.

7.2. Formalisation et modélisation ontologique

La modélisation UML répond aux besoins d'expression, mais pas au besoin d'une compréhension machine. Le langage UML ne permet qu'une représentation semi-formelle de la connaissance et ne prend pas en compte la mise en place des raisonnements associés au savoir-faire. Pour répondre à cette limitation, nous proposons de traduire cette représentation UML dans une équivalence sémantique formelle sous forme d'une ontologie (Mkhinini *et al.*, 2019). Plusieurs travaux scientifiques se sont intéressés à l'étude de la relation entre les modèles UML et les langages d'ontologie. Pour comprendre et se positionner par rapport à ces travaux, nous avons réalisé une revue de

littérature systématique (SLR : Systematic Literature Review) sur le sujet (Mkhinini *et al.*, 2020). Le résultat de cette étude a montré que la majorité des travaux existants proposent généralement une transformation d'un diagramme de classes UML vers une ontologie en se focalisant sur l'aspect syntaxique des éléments UML. De plus, ils ne prennent pas en compte l'intégration des ontologies existantes validées et utilisées par des experts du même domaine d'application. De notre expérience, la transformation des modèles UML en ontologie doit se baser, au-delà de la syntaxe, sur le contexte et le sens de chaque élément. Pour répondre à ce besoin et compléter le savoir décrit par les experts métier, nous sélectionnons et utilisons tout ou partie d'ontologies existantes disponibles dans des bases comme LOV [2] (Linked Open Vocabularies).

Pour ce faire, nous avons développé une approche basée sur la correspondance sémantique entre les modèles de connaissances permettant la réutilisation des ontologies existantes dans le domaine (Mkhinini *et al.*, 2019). Après validation par les experts métier des modèles UML générés à l'étape précédente, nous utilisons des règles de transformation inspirées des travaux existants pour passer des modèles UML représentants les savoir-faire à une représentation ontologique formelle. Nous proposons par la suite d'intégrer des ontologies existantes en identifiant les similarités sémantiques entre ces ontologies et les diagrammes de classes UML. Une étape d'alignement des concepts identifiés dans des sources externes (connaissance du domaine) et ceux définis en local (savoir-faire des experts de l'entreprise) est alors nécessaire pour mettre en correspondance les connaissances du domaine et les savoir-faire des experts métiers. Ainsi, l'ontologie finale définit le vocabulaire logique qui permet d'exprimer des faits et des connaissances du domaine et les connaissances métiers sur lesquelles nous pouvons raisonner. Cette approche permet l'identification des ensembles sémantiques correspondant à la description de chaque élément du savoir-faire en relation avec les connaissances du domaine. L'ontologie résultante servira de base pour construire des cas d'usages métiers et des outils numériques adaptés sur lesquels des règles de raisonnement pourront être ajoutées.

8. Expérimentation

Notre approche de formalisation d'ontologies basée sur l'élicitation de connaissances a été expérimentée dans le cadre de projets de collaborations industrielles.

Le premier projet, en collaboration avec une multinationale, dans le domaine du packaging de luxe [3], avait pour principal objectif la capitalisation et la digitalisation des savoir-faire de l'entreprise pour la construction de nouvelles solutions numériques. Ce projet englobait des processus métiers réalisés sur 4 pays (France, Etats-Unis, Brésil, Chine), dont 4 langues et usages différents. Dans ce projet, nous avons appliqué notre démarche d'élicitation de connaissances pour interviewer 55 collaborateurs représentant différents services de l'entreprise sur plusieurs sites dans les 4 pays. Cette

2. https://lov.linkeddata.es/dataset/lov/
3. Le contrat de confidentialité ne permet pas de dévoiler le nom de l'entreprise.

expérience nous a confronté aux problèmes d'expression du savoir-faire, souvent basée sur une terminologie diversifiée et complexe. Après transcription et analyse des connaissances collectées, nous avons identifié le vocabulaire du domaine, élaboré une vision globale de l'organisation des différents services de l'entreprise, et proposé une modélisation UML pour la structuration et l'unification de la terminologie et des usages. Après validation par les experts, cette modélisation a été par la suite utilisée pour construire une ontologie formelle basée sur le savoir-faire, et à laquelle nous avons intégré des concepts d'ontologies du domaine. L'ontologie résultat a été utilisée pour démontrer la faisabilité industrielle de notre approche et fournir un espace numérique collaboratif d'expression des savoir-faire pour l'entreprise. À partir de cet environnement numérique, nous avons développé une application de raisonnement artificiel permettant d'extraire l'ensemble des projets, expériences, risques, coûts, ressources matérielles, matières premières et ressources humaines proches des éléments décrits dans un cahier des charges client. Il était ainsi possible d'anticiper au plus juste le risque de prise en charge de réalisation de la demande client.

La deuxième expérimentation concerne un projet de collaboration dans le domaine de la viticulture [4]. L'objectif de ce projet était de fournir une plateforme de traçabilité complète du cycle de vie de la vigne en se basant sur les connaissances d'experts et des données de remontées de capteurs. L'application de notre approche d'élicitation de connaissances dans ce projet a permis la collecte des savoir-faire métier (mise en culture de la vigne, lutte contre les maladies, méthodes de vinification, etc.) des viticulteurs et des viniculteurs afin de comprendre le cycle de vie de la vigne et du vin, et les différents raisonnements et pratiques mis en œuvre dans le cadre de leur profession. Ainsi, pour comprendre le vocabulaire du domaine et construire les guides d'entretiens semi-directifs, un travail de veille bibliographique a été réalisé pour en savoir plus sur la filière vitivinicole et les risques qui y sont associés. Dans ce domaine, les viticulteurs se réfèrent beaucoup à l'expérience de leurs pairs, et l'observation de la nature et de ses phénomènes. Les aspects scientifiques et législatifs sont aussi très présents et doivent être pris en compte dans la prise de décisions. Du fait de la complexité du sujet exploré, il a été nécessaire de développer plusieurs grilles d'entretien permettant de prendre en considération les différents paramètres venant impacter le cycle de la vigne et la vinification. Ainsi, six grilles d'entretiens semi-directifs explorant différentes thématiques ont été développées, à savoir sur le cycle de la vigne, les maladies, les ravageurs, les risques climatiques, les risques physiologiques et la vinification. Un panel de 20 viticulteurs et de 5 professionnels du vin a été sélectionné pour passer les entretiens. Après restitution et analyse des savoirs collectés, nous avons proposé une structuration et une modélisation semi-formelle de ces savoirs sous forme de cartes mentales et de diagrammes UML, suivi d'une formalisation ontologique permettant de construire des solutions numériques capables de prédire l'état des vignes et de fournir aux viticulteurs des recommandations pour leur production (Belkaroui *et al.*, 2018). L'avenir dira si ce système permettra aux experts du domaine, comme ils le souhaitaient à l'origine du projet, d'anticiper les risques liés aux ravageurs et

4. https://www.europe-bfc.eu/beneficiaire/fui-23-wine-cloud/

maladies, de réduire les traitements chimiques dangereux pour l'environnement, et de maîtriser la modification des caractéristiques organoleptiques de leurs vins suite à l'influence du changement climatique.

Nous appliquons notre approche dans plusieurs projets de collaboration, y compris le projet ANR DALHAI [5], qui exige la collecte et la formalisation de connaissances expertes pour construire des solutions d'intelligence artificielle innovantes.

9. Conclusion

Cet article propose une approche méthodologique pour la construction d'une ontologie formelle basée sur le savoir-faire des experts et les connaissances du domaine. Nous avons défini une méthodologie d'acquisition des savoir-faire inspirée de la didactique professionnelle, et basée sur les apports de la psychologie du travail et du développement. Cette méthodologie est une combinaison des techniques analytiques, conversationnelles et observationnelles en élicitation de connaissances. Elle permet de collecter, qualifier et analyser les savoir-faire explicites et tacites, d'intégrer les connaissances génériques du domaine, et de faire face aux problématiques liées aux facteurs humains et à l'hétérogénéité sémantique des connaissances. Son principal objectif est de mettre l'humain qui détient la connaissance au centre des réflexions dans le développement des solutions numériques, et d'accompagner au mieux les analystes dans le processus de préparation, de collecte, d'analyse et de restitution de ces connaissances.

Les connaissances collectées sont par la suite organisées et structurées dans des modèles semi-formels, qui seront validés par les experts métier avant d'être formalisés dans une ontologie du domaine, consensuelle et partagée. Cette démarche permet de se doter d'une base de connaissance digitalisée et d'un ensemble d'outils numériques proches des experts et construits à base de leurs savoir-faire, leurs expériences et leurs attentes. Nous avons expérimenté notre approche dans le cadre de deux projets de collaboration industrielle. Le premier pour la digitalisation des savoir-faire d'une multinationale dans le domaine du packaging et le deuxième dans le domaine de la viticulture pour la prédiction du cycle de vie de la vigne. Ces deux expériences ont montré l'apport et la faisabilité industrielle de notre approche et remporté l'adhésion des experts du domaine comme celui du personnel IT.

Remerciements

Nous remercions le projet DALHAI pour le soutien apporté à ce travail scientifique.

Bibliographie

Aussenac-Gilles N., Charlet J., Reynaud C. (2012). Les enjeux de l'ingénierie des connaissances. *Information, Interaction, Intelligence-Le point sur le i (3)*, p. 244–266.

5. https://anr.fr/Projet-ANR-20-CE24-0001

Belkaroui R., Bertaux A., Labbani O., Hugol-Gential C., Nicolle C. (2018). Towards events ontology based on data sensors network for viticulture domain. In *Proceedings of the 8th international conference on the internet of things*, p. 1–7.

Byrd T. A., Cossick K. L., Zmud R. W. (1992). A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS quarterly*, p. 117–138.

Chen Y. (2012). *Requirement elicitation techniques.* https://www.umsl.edu/~sauterv/analysis/6840papers_f12/Chen/.

Chevalier F., Meyer V. (2018). Chapitre 6. les entretiens. *Les méthodes de recherche du DBA*.

Christel M. G., Kang K. C. (1992). *Issues in requirements elicitation*. Rapport technique. Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst.

Dany L. (2016). Analyse qualitative du contenu des représentations sociales. *Les représentations sociales*, p. 85–102.

De Nicola A., Missikoff M., Navigli R. (2005). A proposal for a unified process for ontology building: Upon. In *Database and expert systems applications: 16th international conference, dexa 2005, copenhagen, denmark, august 22-26, 2005. proceedings 16*, p. 655–664.

Droesbeke J.-J., Fine J., Saporta G. (2002). *Méthodes bayésiennes en statistique*. Editions Technip.

Euzenat J. (1995). *Building consensual knowledge bases: context and architecture.* IOS press.

Euzenat J. (1996). Corporate memory through cooperative creation of knowledge bases and hyper-documents. In *10th workshop on knowledge acquisition (kaw)*, p. 36–1.

Fernández-López M. (1999). Overview of methodologies for building ontologies.

Gavrilova T., Andreeva T. (2012). Knowledge elicitation techniques in a knowledge management context. *Journal of Knowledge Management*.

Gómez-Pérez A., Fernández-López M., Corcho O. (2004). Methodologies and methods for building ontologies. *Ontological Engineering: with examples from the áreas of Knowledge Management, e-Commerce and the Semantic Web*, p. 107–197.

Gómez-Pérez A., Juristo N., Pazos J. (1995). Evaluation and assessment of knowledge sharing technology. *Towards very large knowledge bases*, p. 289–296.

Gruber T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, vol. 43, n° 5-6, p. 907–928.

Gruninger M., Fox M. S. (1996). The logic of enterprise modelling. In *Modelling and methodologies for enterprise integration: Proceedings of the ifip tc5 working conference on models and methodologies for enterprise integration, queensland, australia, november 1995*, p. 140–157.

Hayes P., Eskridge T. C., Saavedra R., Reichherzer T., Mehrotra M., Bobrovnikoff D. (2005). Collaborative knowledge capture in ontologies. In *Proceedings of the 3rd international conference on knowledge capture*, p. 99–106.

Holsapple C. W., Joshi K. D. (2002). A collaborative approach to ontology design. *Communications of the ACM*, vol. 45, n° 2, p. 42–47.

Hudlicka E. (1996). Requirements elicitation with indirect knowledge elicitation techniques: comparison of three methods. In *Proceedings of the second international conference on requirements engineering*, p. 4–11.

Iqbal T., Suaib M. (2014). Requirement elicitation technique:-a review paper. *Int. J. Comput. Math. Sci*, vol. 3, n° 9, p. 1–6.

Jones D., Bench-Capon T., Visser P. (1998). Methodologies for ontology development.

Kanza S., Stolz A., Hepp M., Simperl E. (2018). What does an ontology engineering community look like? a systematic analysis of the schema. org community. In *European semantic web conference*, p. 335–350.

Khan[1] S., Dulloo A. B., Verma M. (2014). Systematic review of requirement elicitation techniques. *India*.

Maedche A., Schnurr H.-P., Staab S., Studer R. (2000). Representation language-neutral modeling of ontologies. In *Proceedings of the german workshop" modellierung-2000". koblenz, germany*, p. 129–142.

Missikoff M., Schiappelli F. (2005). A method for ontology modeling in the business domain. In *Emoi-interop*.

Mkhinini M. M., Labbani O., Nicolle C. (2019). Enterprise knowledge modeling, uml vs ontology: Formal evaluation. In *2019 ieee 15th international conference on intelligent computer communication and processing (iccp)*, p. 127–134.

Mkhinini M. M., Labbani-Narsis O., Nicolle C. (2020). Combining uml and ontology: An exploratory survey. *Computer Science Review*, vol. 35, p. 100–223.

Noy N. F., McGuinness D. L. *et al. (2004). Ontology development 101: a guide to creating your first ontology. 2001.* See http://protege. stanford. edu/publications.

OMG. (2017). Omg© unified modeling language (omg uml©). In Version 2.5.1,© 2017 object management group, inc. *Consulté sur https://www.omg.org/spec/UML/*

Pastré P. (2002). L'analyse du travail en didactique professionnelle. Revue française de pédagogie*, p. 9–17.*

Pinto H. S., Martins J. P. (2004). Ontologies: How can they be built? Knowledge and information systems*, vol. 6, p. 441–464.*

Rioufreyt T. (2016). La transcription d'entretiens en sciences sociales.

Sharma S., Pandey S. K. (2014). Requirements elicitation: Issues and challenges. In 2014 international conference on computing for sustainable global development (indiacom)*, p. 151–155.*

Simperl E., Luczak-Rösch M. (2014). Collaborative ontology engineering: a survey. The Knowledge Engineering Review*, vol. 29, n° 1, p. 101–131.*

Studer R., Benjamins V. R., Fensel D. (1998a). Knowledge engineering: Principles and methods. Data & knowledge engineering*, vol. 25, n° 1-2, p. 161–197.*

Studer R., Benjamins V. R., Fensel D. (1998b). Knowledge engineering: Principles and methods. Data & knowledge engineering*, vol. 25, n° 1-2, p. 161–197.*

Uschold M. (1996). Building ontologies: Towards a uni ed methodology. In Proceedings of 16th annual conference of the british computer society specialists group on expert systems.