Élaboration d'ontologies à partir de corpus en utilisant la méthode d'ingénierie des connaissances KOD

Jean-Marc Mercantini* — Nicole Tourigny — Eugène Chouraqui ***

**LSIS – UMR CNRS 6168*
Domaine universitaire Saint-Jérôme, Avenue Escadrille Normandie-Niemen
Université Aix-Marseille (U3), 13397 Marseille cedex 20 France
{Jean-Marc.Mercantini, Eugene.Chouraqui}@lsis.org

*** LSI-ERICAE Département d'informatique et de génie logiciel*
Pavillon Pouliot, Faculté des sciences et de génie
Université Laval, Québec, (Québec) Canada G1K 7P4
Nicole.Tourigny@ift.ulaval.ca

RÉSUMÉ. *Fondée sur l'anthropologie et la linguistique, la méthodologie KOD a d'abord été conçue pour élaborer des systèmes de connaissances. Son but était de guider le cogniticien lors de la construction d'un modèle conceptuel distinct du modèle d'implémentation. L'objectif de cette communication est de montrer comment KOD peut être utilisée pour élaborer des ontologies au niveau des connaissances à partir d'un corpus. KOD permet en outre de guider l'ontologiste dans la réalisation de sa tâche. Des exemples, tirés de nos expérimentations, illustrent la pertinence de KOD pour cette tâche. Le corpus s'est avéré d'une influence importante lors du processus de construction d'ontologies.*

ABSTRACT. *Founded on the fields of anthropology and linguistics, the KOD methodology was first elaborated for knowledge system building. Its main goal was to guide the knowledge engineer when building a conceptual model distinct from the implementation model. In this paper, we show how KOD can be applied to corpus for ontology elaboration at the knowledge level. KOD can guide the ontologist in the ontology building task. Some examples, drawn from our experimentations, illustrate the pertinence of KOD for this task. The corpus turned out to be a key factor of the ontology building process.*

MOTS-CLÉS : *ontologie, ingénierie des connaissances, modélisation des connaissances, méthode, corpus, linguistique, conceptualisation, KOD.*

KEYWORDS: *ontology, knowledge engineering, knowledge modeling, method, corpus, linguistic, conceptualization, KOD.*

1. Introduction

Une ontologie est un langage qui permet de spécifier explicitement une conceptualisation, définie comme une vue simplifiée d'un monde que l'on veut représenter dans un certain but en utilisant des concepts et leurs relations (Gruber, 1993). La nécessité d'établir plusieurs ontologies pour une conceptualisation donnée vient du fait que le monde ne s'interprète pas de façon unique pour l'ensemble des individus et des machines dites « intelligentes ». Sa décomposition, en objets discrets et relations entre ces objets, dépend de la façon de l'observer et des intentions des observateurs. La façon d'observer le monde et l'interprétation qui en est faite dépendent directement de la culture des observateurs et des moyens dont ils disposent pour l'observer. Ainsi, nous voyons apparaître les difficultés inhérentes à des situations de travail impliquant la coopération et la collaboration d'individus de cultures différentes (milieu social, métier, formation universitaire, etc.) ainsi que la mise en œuvre de machines intelligentes. Si ces difficultés peuvent être maîtrisées lors de situations « normales » de travail, elles peuvent devenir de réelles sources de danger lors de situations critiques où des décisions doivent être prises et exécutées sous fortes contraintes. Un des objectifs des ontologies est de faciliter les échanges de connaissances entre humains, entre humains et machines ainsi qu'entre humains par l'intermédiaire de machines (Uschold *et al*., 1996). Dans ce sens, il devient nécessaire de résoudre les difficultés engendrées par l'observation, la représentation et l'interprétation de situations (normales ou critiques) afin de faciliter la résolution de problèmes (l'intention). Dans cette communication, nous présentons nos travaux portant sur la construction d'ontologies en utilisant une méthode servant de base à la conception de systèmes d'information intelligents d'aide à la résolution de problèmes.

Le Génie Ontologique porte sur la construction d'ontologies, qui s'avère actuellement un sujet de recherche très important. Le Génie ontologique peut être vu comme la branche de l'Ingénierie des Connaissances (IC) qui exploite les principes de l'Ontologie (Philosophie) pour élaborer des ontologies (Guarino, 1995 ; Guarino *et al*., 1995). L'IC porte sur l'étude du processus de construction des systèmes de connaissances (SC) (Studer *et al*., 1998), lesquels sont exploités pour réaliser ou aider à réaliser des tâches humaines qui demandent beaucoup de connaissances, généralement difficiles à formaliser, et leur développement nécessite la modélisation et l'acquisition de ces connaissances; l'IC permet l'étude de concepts, de méthodes et de techniques pour effectuer ces activités (Charlet *et al*., 2000).

L'état de l'art sur les ontologies en général et sur le Génie Ontologique en particulier montre qu'un grand nombre de travaux ont été développés ou sont en cours de développement pour proposer des outils automatiques d'aide à la construction d'ontologies. Toutefois, il n'existe aucun outil entièrement automatisé qui, à partir d'un corpus initial décrivant un certain domaine, produise l'ontologie de ce domaine. Les techniques existantes sont qualifiées de semi-automatiques et

nécessitent l'intervention de l'expert en Génie Ontologique pour aboutir à un résultat pertinent. La plupart des outils proposés sont le résultat de l'intégration d'outils existants, ce qui pose le problème de la cohérence des résultats produits par l'enchaînement des outils. Les étapes les plus critiques sont celle de l'acquisition des connaissances à partir de sources documentaires pouvant être multimodales et celle de la structuration du domaine, avec en particulier la détermination des relations sémantiques.

L'étude des méthodes de construction d'ontologies a donné lieu à plusieurs travaux de synthèse, par exemple (Corcho *et al*., 2003), (Fernández-López *et al*., 2002; 1999) et (Pinto *et al*., 2004). Les méthodes qui ont fait l'objet de nombreuses publications et citations dans la littérature scientifique et qui peuvent être considérées comme étant les plus représentatives sont TOVE (Grüninger *et al*., 1995), ENTERPRISE (Uschold, 1996), (Uschold *et al*., 1995), METHONTOLOGY (Fernández *et al*., 1997), (Fernandez *et al*., 1999) et TERMINAE (Aussenac-Gilles *et al*., 2000). La principale critique formulée envers TOVE est qu'elle se limite à la formalisation des connaissances, sans proposer une étape préalable de modélisation conceptuelle de l'ontologie. Avec TOVE, l'expert doit en effet structurer le domaine en utilisant des concepts exprimés au moyen d'un langage formel (basé sur la logique du premier ordre), directement à partir du corpus.

Avec ENTERPRISE, l'expert peut définir un modèle conceptuel de l'ontologie en partant de sa connaissance des concepts du domaine. ENTERPRISE guide ensuite l'expert dans une démarche de généralisation et d'instanciation des concepts connus. Le modèle conceptuel est ensuite formalisé et implémenté. La méthode METHONTOLOGY situe le processus de construction ontologique dans un cadre d'IC et de Génie Logiciel. METHONTOLOGY permet de construire des ontologies au niveau des connaissances à l'aide d'un modèle conceptuel intermédiaire, sans nécessiter une connaissance a priori de concepts. La construction de l'ontologie est réalisée selon un cycle constitué de quatre étapes: spécification (but et utilisateurs visés de l'ontologie), conceptualisation (structuration du domaine au niveau des connaissances), formalisation (traduction automatique du modèle conceptuel en utilisant des traducteurs) et implémentation (expression du modèle formel à l'aide d'un langage d'implémentation). Des guides sont proposés, constitués d'un ensemble de tables prédéfinies facilitant l'acquisition des connaissances et leur conceptualisation. Une plateforme logicielle, ODE (Blázquez *et al*., 1998), a été développée pour guider l'utilisation de METHONTOLOGY.

Avec TERMINAE, le processus de construction ontologique est situé dans un cadre d'ingénierie des connaissances linguistiques. Il repose sur quatre étapes principales qui sont : la constitution du corpus, l'analyse linguistique, la normalisation et la formalisation. L'analyse linguistique permet, à partir du corpus textuel constitué, d'identifier la terminologie du domaine étudié grâce à l'utilisation d'outils de traitement automatique du langage naturel. Dans l'étape de normalisation, qu'on peut associer à l'étape de conceptualisation de METHONTOLOGY, l'expert du domaine choisit des concepts pour structurer et

valider la connaissance qui sera représentée dans l'ontologie, en se basant sur les résultats obtenus suite à l'analyse linguistique. TERMINAE est aussi une plateforme logicielle qui permet à l'expert de construire l'ontologie. TERMINAE fournit des guides méthodologiques.

Les travaux du Génie Ontologique ont fait ressortir la nécessité d'élaborer des guides pour construire des ontologies au niveau des connaissances à partir de sources documentaires, très souvent exprimées sous forme textuelle, mais qui peuvent aussi être multimodales. KOD (*Knowledge Oriented Design)*, développée par (Vogel, 1988), est justement une méthode d'IC dont le but initial était d'aider à construire un modèle conceptuel distinct du modèle d'implémentation lors de la conception de systèmes d'information intelligents d'aide à la résolution de problèmes. Fondée sur les travaux réalisés en anthropologie et en linguistique, KOD fournit des guides pour l'élaboration d'un modèle conceptuel à partir d'un corpus de connaissances.

L'objectif de la présente communication est de montrer comment KOD peut être utilisée pour aider à acquérir des connaissances à partir d'un corpus et à les conceptualiser pour élaborer des ontologies au niveau des connaissances, soit un niveau distinct de l'implémentation (Newell, 1982). Après avoir décrit la démarche généralement utilisée pour la construction d'ontologies, nous expliquerons comment KOD peut être appliquée à un corpus particulier pour identifier la terminologie du domaine étudié et formuler un modèle conceptuel. Des exemples permettront d'illustrer cette application. Enfin, nous conclurons sur la pertinence de cette méthode pour la construction d'ontologies et sur l'influence du corpus sur le déroulement du processus de modélisation.

2. Démarche générale pour l'élaboration d'une ontologie

Les travaux réalisés en Génie Ontologique ont permis de mettre en évidence cinq étapes principales pour le développement d'une ontologie (Gandon, 2002) :

1. Spécification de l'ontologie. Cette étape a pour but de fournir une description claire du problème étudié ainsi que la façon de le résoudre. Elle permet de préciser l'objectif, la portée et le degré de granularité de l'ontologie qui sera construite.

2. Définition du corpus. Il s'agit de sélectionner parmi les différentes sources de connaissance celles qui permettront de répondre aux objectifs de l'étude, définis dans l'étape de spécification.

3. Étude linguistique du corpus. Cette étape consiste à analyser le corpus pour en extraire les termes porteurs de connaissance ainsi que les relations qui les lient.

4. Conceptualisation. Lors de cette étape, il s'agit de transformer les termes obtenus suite à l'étude linguistique du corpus : les termes seront transformés en concepts et les relations lexicales en relations sémantiques. Au terme de cette étape, un modèle conceptuel est obtenu.

5. Formalisation. L'objectif de cette étape est d'exprimer au moyen d'un langage formel le modèle conceptuel obtenu au terme de l'étape précédente.

3. KOD : une méthode d'ingénierie des connaissances

Nous nous plaçons dans le cas de figure du développement d'une ontologie devant servir de base à la spécification d'un outil informatique d'aide à la résolution d'un problème dans un domaine donné. Ainsi, l'ontologie ne se limite pas à structurer ce domaine mais elle le structure au regard du problème à résoudre et compte tenu de la méthode de résolution de ce problème. Dans ce sens, il est important que la méthode d'élaboration de l'ontologie aide le cogniticien à conceptualiser le triplet Td : < Domaine, Problème, Méthode >. L'ontologie réalise la cohérence de ce triplet pour servir de base à la conception d'un artéfact.

KOD s'inscrit dans la famille des méthodes d'IC ayant pour but de guider le cogniticien dans sa tâche d'élaboration de SC. KOD a été conçue pour introduire un modèle explicite entre la formulation du problème en langue naturelle et sa représentation dans le métalangage informatique choisi. KOD repose sur une démarche inductive qui, sur la base d'un corpus constitué de documents, d'observations et de discours d'experts, oblige à exprimer de façon explicite le modèle conceptuel des connaissances (aussi appelé modèle cognitif) des experts.

Les principales caractéristiques de cette méthode sont qu'elle repose sur des principes issus de la linguistique et de l'anthropologie. Ses fondements linguistiques la rendent bien adaptée pour l'acquisition de connaissances exprimées en langage naturel. Ainsi, elle propose un cadre méthodologique pour guider la collecte des termes et les organiser à partir d'une analyse terminologique. Grâce à ses fondements anthropologiques, le cogniticien dispose d'un cadre méthodologique facilitant l'analyse sémantique de la terminologie utilisée pour produire un modèle cognitif. Enfin, KOD guide le travail du cogniticien depuis l'extraction de la connaissance jusqu'à l'élaboration du modèle informatique.

La mise en œuvre de la méthode KOD repose sur l'élaboration de trois modèles successifs : le modèle pratique, le modèle cognitif et le modèle informatique (Tableau 1). Chacun de ces modèles est élaboré selon trois paradigmes : <Représentation, Action, Interprétation>. Le paradigme Représentation permet de modéliser l'univers tel que l'expert se le représente, cet univers étant constitué d'objets concrets ou abstraits en interrelation. KOD offre les outils méthodologiques pour élaborer la structure de cet univers de connaissance suivant ce paradigme. Le paradigme Action permet de modéliser le comportement actif de certains objets. Ainsi, les plans d'actions conçus par des opérateurs humains, aussi bien que ceux d'opérateurs artificiels, seront modélisés dans un même format. Le paradigme Interprétation / Intention permet de modéliser les raisonnements mis en œuvre par l'expert pour interpréter des situations et projeter des plans d'action en fonction de ses intentions.

Le modèle pratique résulte de la représentation d'un discours ou d'un document exprimé dans les termes du domaine, au moyen de taxèmes (structures statiques des objets), d'actèmes (structures dynamiques des objets) et d'inférences (base de la structure cognitive de la tâche).

Modèles/ Paradigmes	Représentation	Action	Interprétation
Pratique	Taxèmes	Actèmes	Inférences
Cognitif	Taxinomies	Actinomies	Schémas
Informatique	Objets	Méthodes	Règles

Tableau 1. *KOD : les trois niveaux de modélisation selon les trois paradigmes*

Le modèle cognitif est construit par abstraction du modèle pratique. Il est composé de taxinomies, d'actinomies et de schémas d'interprétation. Le modèle informatique résulte de la formalisation du modèle cognitif dans un langage formel indépendant de tout langage de programmation.

4. KOD : élaboration des modèles

La première étape de modélisation consiste à élaborer un modèle pratique, à partir d'un corpus Mp (Figure 1), pour extraire les termes et les relations qui les lient au moyen d'une analyse de façon à constituer un langage terminologique. Les termes du langage sont classés en <taxèmes, actèmes, inférences> en accord avec les trois paradigmes de la méthode (Tableau 1). Les taxèmes permettent de représenter les objets et les concepts considérés par les experts du domaine. Les actèmes décrivent les activités pouvant être effectuées par les experts et provoquant des changements d'états au niveau des objets et des concepts. Les inférences sont les éléments grâce auxquels les experts construisent leurs raisonnements pour interpréter une situation et projeter leur intention d'action.

L'étape suivante est l'élaboration du modèle cognitif, effectuée à partir du modèle pratique, au moyen d'un langage conceptuel basé sur la terminologie du domaine. Le processus consiste à : (i) analyser les termes synonymes et homonymes ; (ii) transformer les termes résultants en concepts et les relations lexicales en relations sémantiques. En accord avec les trois paradigmes de la méthode, le modèle cognitif est organisé en <taxinomies, actinomies, schémas de raisonnement> (Figure 1). Les taxinomies sont le résultat de la classification des taxèmes. Elles se présentent sous forme d'une structure arborescente hiérarchique de type « sorte-de » ou « est-un » liant les concepts et les objets du domaine. Les actinomies sont le résultat d'une organisation ordonnée des actèmes définissant un plan d'action. Les schémas de raisonnement définissent une structure d'inférence modélisant le raisonnement de l'expert lorsqu'il planifie ses actions.

La troisième étape porte sur l'élaboration du modèle informatique qui nécessite d'exprimer au préalable le modèle conceptuel au moyen d'un langage formel pour

constituer un modèle formel. Le choix de ce langage formel dépend des propriétés du modèle conceptuel. L'opération de formalisation consiste à intégrer les éléments du modèle conceptuel dans la définition de classes et d'objets devant être utilisés pour le développement du logiciel. L'étape d'implémentation qui suit consiste à traduire le modèle formel dans un langage de programmation.

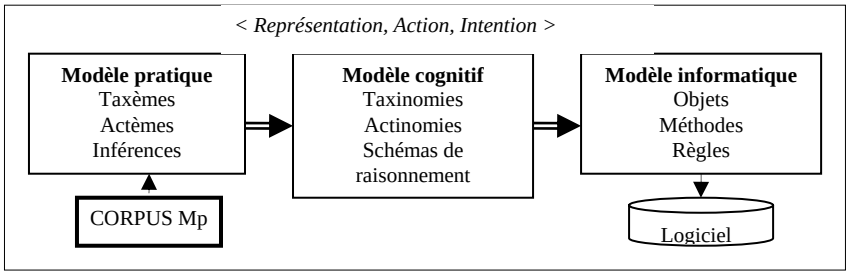


Figure 1. *Les trois niveaux de modélisation selon KOD*

Le processus de modélisation avec KOD peut également être détaillé comme sur la figure 2. Le corpus étant constitué de plusieurs documents (D1, D2, …, Dn), chacun fait l'objet d'une étude linguistique et les résultats permettent de construire l'ensemble des modèles pratiques (MP1, MP2, …, MPn).

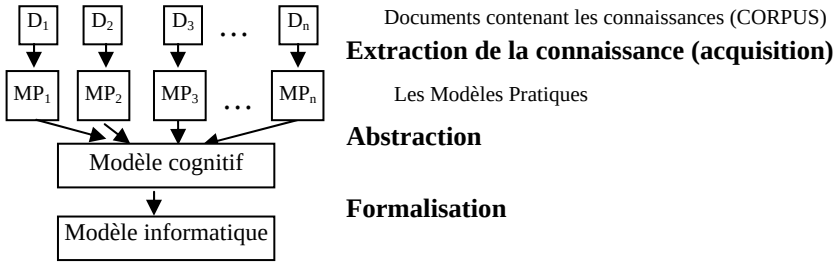


Figure 2. *KOD : processus de modélisation*

5. KOD et le cycle de vie de construction des ontologies

Le développement d'une ontologie peut être vu comme une tâche requérant une grande quantité de connaissances. Dans ce sens, l'utilisation de KOD semble a priori pertinente puisque l'un des objectifs d'une méthode d'IC est d'aider à définir les connaissances d'un domaine étudié. Dans la présente section, nous expliquons

comment KOD peut permettre de réaliser chacune des étapes du processus de développement d'une ontologie.

La projection de la méthode KOD sur la démarche générale d'élaboration d'une ontologie montre qu'elle guide la constitution du corpus et qu'elle fournit les outils opérationnels pour couvrir les étapes 3 (étude linguistique) et 4 (conceptualisation). Dans ce qui suit, nous reprenons chaque étape du cycle de vie de développement d'une ontologie, que nous mettons en correspondance avec les étapes du processus proposé par KOD. Nous développerons de façon plus spécifique : (i) la constitution du corpus, (ii) l'étude linguistique et (iii) la conceptualisation.

Dans le cadre de travaux de recherche antérieurs, nous avons pu mettre en œuvre cette méthode (Mercantini *et al.*, 2000 ; 2003 ; 2004) dans les domaines de l'accidentologie routière, la sécurité des sites industriels urbains et l'étude des erreurs de conduite d'installations industrielles. Pour illustrer chacune des étapes du processus, nous avons extrait des exemples de ces études de cas.

5.1. *Spécification de l'ontologie*

La méthode KOD ne propose pas d'outils facilitant la spécification d'une ontologie. Pour mener à bien cette étape, de nombreux auteurs préconisent de s'appuyer sur la notion de scénario (Gandon, 2002), (Uschold *et al.*, 1996), (Caroll, 1997) avec les objectifs de préciser et de justifier le bien fondé de construire une ontologie, son usage anticipé et les destinataires envisagés. Nous ne développerons pas davantage cette étape mais nous illustrons ce propos en donnant les résumés des scénarii qui ont été rédigés dans le cadre de nos trois études de cas. Le plan des scénarii est conforme au triplet Td : < Domaine, Problème, Méthode >.

Etude de cas 1 : Le Domaine concerné est celui des installations industrielles de distribution d'électricité. Le Problème est celui de comprendre le comportement cognitif des opérateurs humains lorsqu'ils sont confrontés à des situations critiques, qu'elles soient dues à un défaut matériel ou à une erreur de conduite. La Méthode de résolution du problème a consisté à construire un simulateur de situations critiques en vue d'immerger les opérateurs dans des situations de travail simulées. L'objectif était de recueillir les modes de raisonnement des opérateurs en fonction des scénarii simulés afin de les modéliser. L'ontologie a été construite à partir d'une base de rapports d'incidents/accidents afin d'élaborer des modèles génériques de scénario d'accident et de situations critiques.

Etude de cas 2 : Le Domaine est celui des Sites Industriels Urbains. Le Problème est celui d'aider les gestionnaires de tels sites à produire des plans d'intervention en cas d'accidents majeurs et de les valider. La Méthode de résolution de problème a consisté à développer un outil de simulation de scénarii d'accidents et de plans d'urgence. Le gestionnaire est alors capable d'évaluer la pertinence de son plan au regard des scénarii d'accident les plus probables. L'ontologie, construite à partir

d'une base de plans d'urgence existants, a permis de proposer un modèle générique de scénario d'accident et un modèle générique de plan d'urgence.

Etude de cas 3 : Le Domaine est celui des accidents de la circulation routière. Le Problème est celui d'assister les experts en accidentologie dans leurs diagnostics d'accidents. La Méthode de résolution du problème a consisté d'une part à développer un Modèle Générique d'Accident (MGA) et d'autre part à développer un Processus de recherche des Causes (PRC). Une première ontologie a été développée à partir d'une base de rapports d'accidents en vue de produire le MGA et une deuxième ontologie a été développée en vue de produire le PRC.

5.2. *Définition du corpus*

La définition du corpus s'effectue sur la base de la spécification de l'ontologie et en considérant les propriétés des modèles pratiques et conceptuels résultant de l'application de la méthode KOD. Ainsi, les documents à collecter doivent à la fois être représentatifs du triplet < Domaine, Problème, Méthode > et répondre aux critères de pertinence exigés par les trois paradigmes < Représentation, Action, Interprétation/Intention >.

5.3. *Étude linguistique du corpus*

L'application de la méthode KOD pour effectuer l'étude linguistique conduit à l'élaboration de Modèles Pratiques suivant deux étapes : (i) une analyse terminologique des documents du corpus, suivie (ii) d'une modélisation sous forme de Taxèmes, d'Actèmes et d'Inférences.

L'analyse terminologique a pour objet d'identifier dans les documents du corpus les termes et les relations utilisés pour décrire les éléments du domaine ainsi que leurs comportements, en considérant le double point de vue du problème abordé et de la méthode de résolution considérée. L'analyse consiste à paraphraser les documents du corpus pour obtenir des phrases simples permettant de qualifier les termes employés. Les termes dont il est question sont représentatifs des trois paradigmes de la méthode. Nous obtenons ainsi un langage terminologique où les termes peuvent être des objets, des valeurs, des relations liant les objets aux valeurs, des actions et des inférences.

La modélisation sous forme de Taxèmes consiste à organiser les termes représentant les objets et les concepts du triplet Td au moyen de prédicats binaires de type < Objet, Attribut, Valeur >. L'attribut définit une relation qui lie l'objet à une valeur. On définit cinq types de relation prédicative : Classifiante, Identifiante, Descriptive, Structurelle et Situative. Dans ce qui suit, nous donnons deux exemples : l'un à partir de texte et l'autre à partir d'un tableau.

Exemple cas 1 : verbalisation à partir d'un texte

Extrait du texte original : « … pour aligner la position de la manette du switch 101 du 21Y5 sur le panneau de commande, avec le signal rouge, et l'ouvrir ensuite … »

Paraphrases : (1) Le 21Y5 contient le switch 101. (2) Le switch 101 possède un indicateur rouge. (3) Le switch 101 possède une manette. (4) Le switch 101 est situé sur le panneau de commande. (5) Le switch 101 a une position ouverte. (6) Le switch 101 a une position fermée.

Les Taxèmes issus de ces paraphrases sont indiqués dans le tableau 2.

Réf.	Taxèmes	Relations
(1)	< 21Y5, composé de, switch 101>	Structurelle
(2)	< switch 101, composé de, indicateur rouge >	Structurelle
(3)	< switch 101, composé de, manette >	Structurelle
(4)	< switch 101, est sur, panneau de commande >	Situative
(5)	< switch 101, position, ouvert >	Descriptive
(6)	< switch 101, position, fermée >	Descriptive

Tableau 2. *Taxèmes de l'exemple cas 1*

Exemple cas 2 : verbalisation à partir d'un tableau (ou d'une figure)

RECAPITULATIF des STOCKAGES de PRODUITS
INFLAMMABLES – EXPLOSIFS – TOXIQUES – CORROSIFS

Chantier : C 11/CO-PRODUITS

Points Sensibles	Repère Appareil	Produits	Identification		Quantité maxi de Stockage	Inflam.	Expl.	Tox.	Corro
			Matière	Danger					
7	R 021.09	Ricino récupéré			55	X			
	R 021.12	Acide Chlorhydrique	1789	88	6			X	X
	R 021.13	Méthanol dilué	1230	336	25	X	X	X	
	R 026.01	Méthanol rectifié	1230	336	25	X	X	X	
	R 021.22	Soude 30 %	1824	88	11				X

…

Paraphrases :

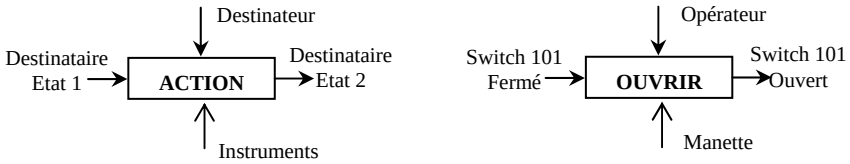
(1) C 11/CO-produits est un Chantier **(2)** PS7 est un Point Sensible **(3)** C 11/CO-produits est composé des Points Sensibles PS6, PS7, PS8, PS9, PS2 **(4)** PS7 contient les appareils R 021.09, R 021.12, R 021.13, R 026.01, R 021.22 **(5)** L'Acide Chlorhydrique est un Produit Dangereux **(6)** Le R 021.09 est un Conteneur **(7)** Le R 021.09 contient du Ricino- récupéré **(8)** Le R 021.09 possède une quantité maxi de stockage de 55 m3 **(10)** Le R 021.12 possède une quantité maxi de stockage de 6 m3 **(11)** Le R.021.12 contient de l'Acide Chlorhydrique **(12)** L'Acide Chlorhydrique est Toxique

Le tableau 3 présente des Taxèmes issus du processus d'analyse de cet exemple.

Réf.	Taxèmes	Relations
(6)	< R 021.09, TYPE, Conteneur >	Classifiante
(11)	< R 021.12, CONTIENT, Acide Chlorhydrique >	Descriptive
(10)	< R 021.12, CAPACITE, 6 m3 >	Descriptive
(12)	< Acide Chlorhydrique, EST, Toxique >	Identifiante

Tableau 3. *Quelques Taxèmes issus de l'exemple cas 2*

L'obtention des actèmes consiste à identifier dans les documents les verbes qui représentent les activités effectuées par des opérateurs humains ou artificiels. La modélisation sous forme d'actèmes (Figure 3) consiste alors à organiser les termes au sein d'une structure de type 7-uplet modélisant les activités associées aux éléments du triplet Td : < Destinateur, Action, Destinataire, Propriétés, Etat1_Dest, Etat2_Dest, Instruments >. L'action est effectuée par le Destinateur au moyen d'Instruments et elle s'applique sur le Destinataire qui subit un changement d'état (Etat1_Dest → Etat2_Dest). Tous les éléments du 7-uplet sont des taxèmes.



OUVRIR (action sur le Switch 101 – 14T1 / 14T2)	
Eléménts	**Valeurs**
Destinateur	(Opérateur salle de contrôle, Opérateur tableau)
Destinataire	(Switch 101 – 14T1, Switch 101 – 14T2)
Etat 1 (Destinataire) 1. Position manette 2. Position contact 3. Etat indicateur rouge 4. Etat indicateur vert	 1. (fermée, ouverte) 2. (fermée, ouverte) 3. (allumé, éteint, grillé) 4. (allumé, éteint, grillé)
Etat 2 (Destinataire) 1. Position manette 2. Position contact 3. Etat indicateur rouge 4. Etat indicateur vert	 1. (fermée, ouverte) 2. (fermée, ouverte) 3. (allumé, éteint, grillé) 4. (allumé, éteint, grillé)
Instruments	(manette-14T1, manette-14T2)

Figure 3. *Représentation générique d'un actème et exemple de l'actème OUVRIR appliqué au Switch 101(extraite du cas 1)*

La modélisation sous forme d'inférences consiste à représenter les éléments du corpus qui caractérisent les activités cognitives des personnes ou des machines. Ainsi, dans le cadre de l'étude de cas 3, nous avons pu classer les inférences des conducteurs automobiles dans les quatre grandes catégories suivantes :

1. Les inférences de type *interprétation* qui sont produites par les conducteurs pour donner un sens aux objets qu'ils observent.

2. Les inférences de type *reconstruction* de situation qui permettent aux conducteurs d'estimer une situation complète à partir d'une observation parcellaire.

3. Les inférences de type *anticipation* qui permettent aux conducteurs d'estimer l'évolution d'une situation à partir d'une observation présente.

4. Les inférences de type *décision* qui permettent aux conducteurs de choisir une action de conduite en fonction de la situation à laquelle ils estiment être confrontés.

Un exemple d'inférence de type *reconstruction* (extrait d'un entretien) est le suivant :

« Le camion s'est levé tout doucement du milieu. »
« La voiture qui le suivait a marqué le stop et démarré derrière le camion. »
« La route était assez large pour que la voiture double le camion. »
« La voiture, elle a dû dépasser le camion. »

Dans cet extrait on comprend que le conducteur ne voit pas la voiture dépasser le camion mais que, suite à cette inférence, il s'est reconstruit une situation où la voiture dépasse le camion.

5.4. *Étude linguistique du corpus*

Cette étape consiste à élaborer le modèle cognitif à partir du modèle pratique. À l'issue de cette étape, on obtient 3 ensembles de concepts : ceux associés aux objets et leurs propriétés, ceux associés aux actions et ceux associés aux schémas de raisonnement. L'abstraction du modèle pratique en un modèle cognitif est basée sur l'opération de classification. On obtient ainsi des taxinomies, des actinomies et des schémas de raisonnement.

5.4.1. *Élaboration des taxinomies*

Cette étape repose sur l'analyse des taxèmes, qui consiste à spécifier la nature des attributs (ou relations) qui caractérisent chaque objet. De la nature de ces attributs vont dépendre la construction des taxinomies (relations « sorte-de » ou « est-un ») ou d'autres types de structure arborescente (relations « est-composé-de », « est-sur », etc.).

A titre d'exemple, reprenons quelques taxèmes du cas 2 :

< R 021.09, TYPE, Conteneur >

< R 021.12, TYPE, Conteneur >
…
< R 033.23, TYPE, Conteneur >

Chacun de ces taxèmes précise que les éléments R 0X.Y appartiennent à la catégorie « Conteneur » et par conséquent ce terme est représentatif d'un ensemble d'éléments qui ont des propriétés communes. Le terme « Conteneur » aura le statut de concept. Par conséquent l'attribut « TYPE » sera représenté, au niveau conceptuel, par la relation d'instanciation « est-un ». Considérons maintenant, le taxème suivant, issu du même cas 2 :

< Conteneur, TYPE, Equipement >.

Il signifie que le concept « Conteneur » appartient à une catégorie de niveau hiérarchique supérieur. Le terme Equipement définit également un concept et l'attribut TYPE lie deux concepts. Cette relation doit être différenciée de la précédente et sera notée « sorte-de ». Ces deux relations sont à la base de la construction des taxinomies. La figure 4 en donne un exemple où les flèches avec un trait continu représentent la relation « sorte de » et celles avec un trait pointillé représentent la relation « est-un ».

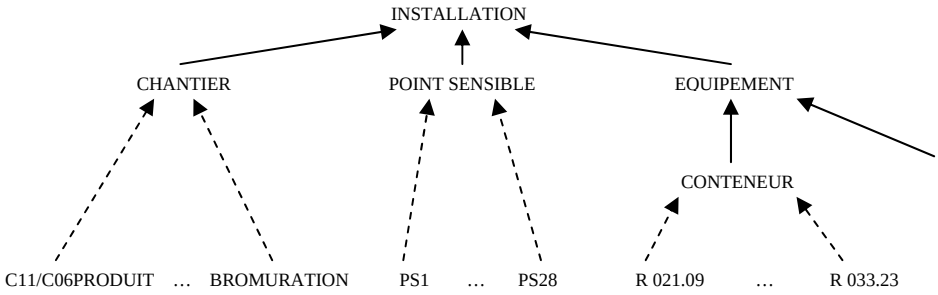


Figure 4. *Taxinomie partielle extraite du cas* 2

Pour les autres types de relations, considérons l'exemple suivant :

< Acide Chlorhydrique, est, Toxique >
< Acide Chlorhydrique, est, Corrosif >
< Méthanol dilué, est, Inflammable >
< Méthanol dilué, est, Explosif >.

Ces taxèmes sont représentatifs d'une relation identifiante. Le problème qui se pose vient du fait que la copule « EST » n'est pas assez précise. En effet, comment distinguer les relations précédentes de la relation : < chat, EST, noir > ? Il devient nécessaire d'interpréter la copule « EST » pour rendre explicite la relation qu'elle sous-entend. Nous obtenons alors, les taxèmes suivants :

< Acide Chlorhydrique, DANGEROSITE, Toxique >
< Acide Chlorhydrique, DANGEROSITE, Corrosif >
< Chat, COULEUR, noir >.

La définition d'un concept est obtenue en regroupant l'ensemble des connaissances le concernant. Ainsi, le concept « Conteneur » a été ainsi défini :

Conteneur	
Nom	: Identificateur
Repère	: {R 021.09, …, R 033.23}
Contient	: {Acide Chlorhydrique, Chlore, … Brome}
Capacité	: {6m3, …, 1000 m3}
Localisation	: {PS1, PS2, …, PS28}
Composé de	: Liste d'équipements
Défaillance	: Fuite
Réparation	: Colmatage
Port_Entrée	: Défaillance/Réparation
Port_Sortie	: Produit
Etat1	: Bon
Etat2	: Rupture
Pression	: Réel
Température	: Réel
Débit de Fuite	: Réel

On y reconnaît les trois groupes de connaissances en relation avec le triplet Td. Le premier est associé au domaine, le deuxième au problème (accident de type dégagement gazeux) et le troisième groupe à la méthode de résolution (simulation).

5.4.2. *Élaboration des actinomies*

Les actinomies sont obtenues par association d'actèmes dans le but de fournir une description du comportement des objets du domaine ou bien pour modéliser les processus associés à la méthode de résolution du problème. Le formalisme que nous avons adopté est le suivant (illustré sur la figure 5) :

< Point de vue, Liste d'actèmes effectués, Liste d'actèmes perçus, Liste des temps début et fin, Ordonnancement temporel, Liens entre les actèmes >

Point de vue : il identifie l' « agent » qui effectue l'actinomie.

Liste d'actèmes effectués : elle énumère les actèmes réalisés par l' « agent » dont le point de vue est pris en compte.

Liste d'actèmes perçus : elle représente les actèmes perçus par l' « agent » dont le point de vue est pris en compte.

Liste des temps de début et fin : elle représente l'ensemble des temps associés au début et à la fin de chaque actème qu'il soit effectué ou perçu.

Ordonnancement temporel : il définit une relation d'ordre total ou partiel entre les données temporelles.

Liens entre les actèmes : ils permettent d'établir les liens structurels entre les actèmes. Ils peuvent établir des relations de type (Destinataire, Destinateur), (Destinataire, Destinataire), (Destinateur, Destinataire) ou (Destinateur, Destinateur).

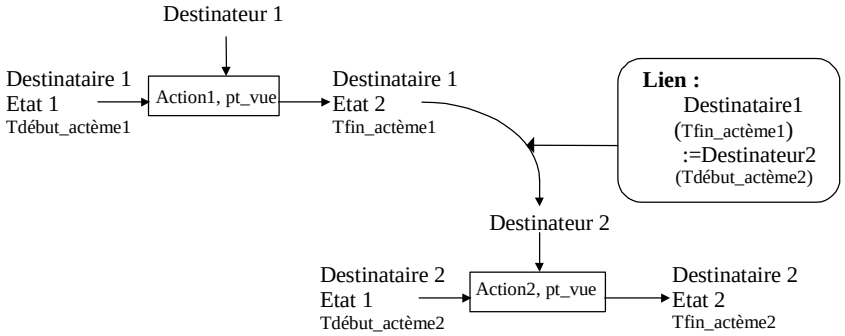


Figure 5. *Représentation générique d'une actinomie*

Dans la figure 6, l'actinomie modélise un scénario d'accident.

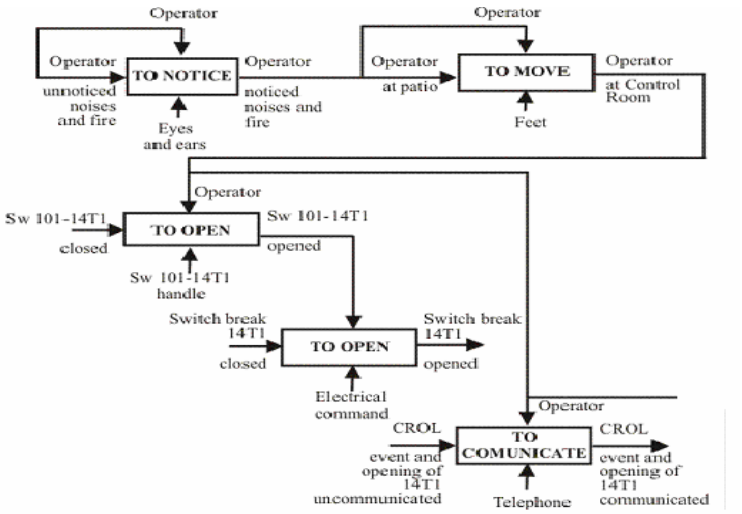


Figure 6. *Actinomie modélisant un scénario d'accident (extraite du cas 1)* (Mercantini *et al.*, 2004)

5.4.3. *Élaboration des schémas de raisonnement*

Le formalisme que nous avons adopté est le suivant :

< Identificateur, Liste d'inférences, Ordonnancement temporel, Liens entre inférences >

Identificateur : il identifie le schéma de raisonnement.

Liste d'inférences : elle énumère les inférences incluses dans le schéma de raisonnement.

Ordonnancement temporel : il donne l'ordonnancement temporel des inférences.

Liens entre inférences : ils donnent l'organisation structurelle du raisonnement.

Voici un exemple d'un schéma de raisonnement.

Considérons les extraits suivants d'un discours :

(1) " J'ai vu la voiture qui était loin. "
(2) " J'avais largement le temps de passer. "
(3) " Je suis passée."

Ces extraits peuvent être représentés de la façon suivante :

< pp1, Observation, <Pilote_1>, <Prise d'Information>, <Véhicule_2>, Valeur, Date>
< pci, Interprétation, <Distance (loin)>, Valeur, Date
< ph1, Hypothèse, <Pilote_1>, <Démarrer>, Valeur, Date >
< pca, Anticipation, <Pilote_1>, <Démarrer>, Réussite, Date >
< pcd, Décision, <Pilote_1>, <Démarrer>, Valeur, Date >

5.5 *Formalisation*

La formalisation d'une ontologie (modèle conceptuel ou cognitif) consiste avant tout à choisir ou à construire le langage formel capable d'intégrer toutes les propriétés de l'ontologie. Mais il doit également être pertinent au regard de l'usage qui doit être fait de l'ontologie. Ainsi, dans le cadre des études de cas devant aboutir à la construction de simulateurs, nous avons utilisé les langages formels suivants : *Discrete EVent Simulation systems* (DEVS) (Zeigler, 1984), *UML-RT*, Réseau *de Petri*. Pour d'autres études, nous avons utilisé la logique des prédicats ou *Objlog* (Faucher, 2001). La figure 7 donne le modèle formel des principaux concepts de l'ontologie qui a été développée dans le cadre de l'étude de cas 2.

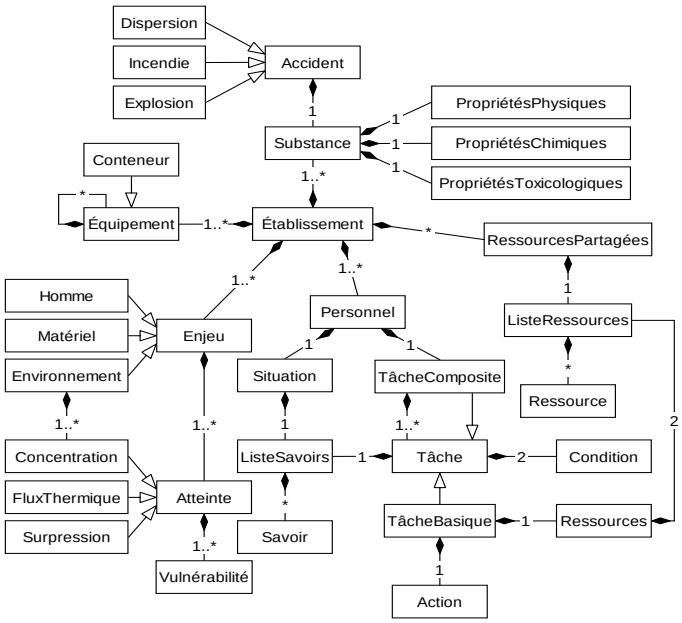


Figure 7. *Exemple de modèle formel (extrait de l'étude de cas 2)*

6. Conclusion et perspectives

Dans cette communication, nous avons décrit la méthode KOD et montré, sur des exemples tirés de nos expérimentations, comment elle peut aider à acquérir des connaissances à partir d'un corpus et les conceptualiser pour élaborer des ontologies au niveau des connaissances. Avec KOD, on peut effectivement constituer un corpus, identifier la terminologie utilisée et élaborer le modèle conceptuel des ontologies. KOD permet de couvrir le cycle généralement accepté pour la construction d'une ontologie, et plus particulièrement les phases de l'étude linguistique et de la conceptualisation.

KOD s'avère donc une méthode pertinente pour la construction d'ontologies. Fondée sur des travaux de recherche effectués en anthropologie et en linguistique, elle propose en effet un cadre méthodologique pour le recueil et l'organisation des connaissances, en fonction non seulement du domaine, mais aussi du problème à résoudre et de la façon de le résoudre. De plus, si des connaissances sont manquantes, elle peut guider le travail du cogniticien afin de les identifier le cas échéant.

Les principales difficultés rencontrées, lors de l'application de KOD aux trois cas que nous avons traités, ne sont pas directement induites par la méthode elle-même mais sont plutôt d'ordre général, à savoir :

- la constitution d'un corpus de documents suffisamment pertinents pour couvrir de façon complète et précise l'espace défini par le triplet Td. En effet, pour chacune des applications traitées, il a fallu compléter le corpus initial au moyen de documents plus précis voire au moyen d'enquêtes sur le terrain ;

- la lecture des documents et l'extraction des connaissances sont orientées suivant l'interprétation des spécifications de l'ontologie, ce qui pose le problème de la pertinence et de la complétude de l'ontologie ;

- la détermination de certaines relations sémantiques. En effet, si certaines relations posent peu de problèmes, comme par exemple les relations structurelles ou classifiantes, il n'en est pas de même pour les relations identifiantes qui traduisent un recouvrement sémantique partiel entre l'objet et la valeur. La difficulté réside alors, dans l'explicitation de cette relation qui résulte d'une démarche inductive et d'une connaissance profonde du triplet Td. L'exemple du paragraphe 5.4.1 illustre cette difficulté.

Nos travaux actuels consistent aussi à automatiser l'opération de formalisation, en particulier vers les langages formels orientés pour la simulation (DEVS et Réseaux de Petri). Enfin, la mise en oeuvre de KOD pour la construction d'ontologies ayant été expérimentée, d'autres défis sont à venir, notamment l'utilisation d'approches complémentaires pour faciliter la recherche d'information dans le corpus, par exemple l'utilisation de patrons. Nos travaux futurs incluent également la mise en oeuvre d'une plateforme logicielle pour assister l'ontologiste dans sa tâche.

5. Bibliographie

Aussenac-Gilles N., Biébow B., Szulman S., «Revisiting ontology design: a method based on corpus analysis», *EKAW-2000, Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management*, R. Dieng et O. Corby (Ed.), LNAI 1937, Springer, 2000, p. 172-188.

Blázquez M., Fernández-López M., Garcia-Pinar J.M., Gómez-Pérez A., «Building ontologies at the knowledge level using the ontology design environment», *Proceedings of the Workshop on Knowledge Acquisition, Modelling and Management: KAW'98*, Banff, Canada, 1998.

Caroll J.M., «Scenario-based Design», Chapter 17, *Handbook of human-computer interaction*, M. Helander, T.K. Landauer, P. Prabhu (ed.), Second completely revised edition, Elsevier Science B.V, 1997.

Charlet J., Zacklad M., Kassel G., Bourigault D., « Ingénierie des connaissances: recherches et perspectives », *Ingénierie des connaissances. Évolutions récentes et nouveaux défis*, J. Charlet., M. Zacklad, G. Kassel, D. Bourigault (Ed.), Eyrolles, 2000, p. 1-22.

Corcho O., Fernández-López M., Gómez-Pérez A., «Methodologies, tools and languages for building ontologies: Where is their meeting point?», *Data & Knowledge Engineering*, vol. 46, n°1, 2003, p. 41-64.

Fernández-López M., Gómez-Pérez A., «Overview and analysis of methodologies for building ontologies», *The Knowledge Engineering Review*, vol. 17, n° 2, 2002, p. 129-156.

Fernández-López M., «Overview of methodologies for building ontologies», *Proceedings, IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5)*, Stockhom, Sweden, August 2, 1999, p, 4-1 à 4-13.

Fernández M., Gómez-Pérez A., Juristo N., «METHONTOLOGY: From ontological art towards ontological engineering, *Proceedings AAAI-97 Spring Symposium Series, Workshop on ontological engineering*, Stanford (California), 1997, p. 33-40.

Faucher C., «Easy definition of new facets in frame-based language Objlog+», *Data & Knowledge Engineering*. vol. 38, n° 3, 2001, p. 223-263.

Gandon F., «Ingénierie d'ontologie: une synthèse et un retour d'expérience», rapport de recherche n° 4396, mars 2002, INRIA Sophia-Antipolis.

Gruber, T.R., Toward principles for the design of ontologies used for knowledge sharing, Technical Report KSL-93-04, Knowledge Systems Laboratory, Stanford University. (Revised in August 1993)

Grüninger M., Fox M.S., «Methodology for the design and evaluation of ontologies», *Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95*, Montréal (Canada), 1995.

Guarino N., «Formal ontology, conceptual analysis and knowledge representation», *International Journal of Human-Computer Studies*, vol.43, n° 5-6, 1995, p. 625-640.

Guarino N., Giaretta P., «Ontologies and knowledge bases: towards a terminological clarification», *Towards very large knowledge bases*, N. J. I. Mars (Ed.), IOS Press, 1995. (cité par Gandon, 2002).

Mercantini J.-M., Loschmann R., Chouraqui E. «A provisional analysis method on safety of an urban industrial site», Safety in the Modern Society, People and Work Research Report 33, 2000, p 105- 109.

Mercantini J.M., Capus L., Chouraqui E., Tourigny N., «Knowledge engineering contributions in traffic road accident analysis», *Innovations in Knowledge Engineering*, R. Jain, A. Abraham, C. Faucher, , B. Jan van der Zwaag (Ed.) 2003, p 211-244.

Mercantini J.-M., Turnell M.F.Q.V, Guerrero C.V.S, Chouraqui E., Vieira F.A.Q., Pereira M.R.B, «Human centred modelling of incident scenarios», *IEEE SMC 2004, Proceedings of the International Conference on Systems, Man & Cybernetics*, The Hague, The Netherlands, October 10-13, 2004, p. 893-898.

Newell A., « The knowledge level », *Artificial Intelligence*, vol. 18, n° 1, 1982, p. 87-127.

Pinto H.S., Martins J.P., «Ontologies : how can they be built ? ", *Knowledge and Information Systems*», vol. 6, n° 4, 2004, p. 441-464.

Studer R., Benjamins V.R., Fensel D., « Knowledge engineering: principles and methods », *Data & Knowledge Engineering*, vol. 25, n° 1-2, 1998, p. 161-197.

Uschold M., «Building ontologies: towards a unified methodology», *Proceedings of Expert Systems 96*, Cambridge, December 16-18, 1996.

Uschold M., Grüninger M., «Ontologies: principles, methods and applications», The *Knowledge Engineering Review*, vol. 11, n° 2, 1996, p. 93-136.

Uschold M., King M., «Towards a methodology for building ontologies», *Proceedings of the IJCAI-95 Worshop on Basic Ontological Issues in Knowledge Sharing*, Montréal (Canada), 1995.

Vogel C., *Génie cognitif*, Masson (Sciences cognitives), Paris, 1988.

Zeigler B.P., Theory of Modelling and Simulation. Krieger Publishing Co., Inc., 1984.