

Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances.

Bruno Bachimont

*Direction de la Recherche Prospective
Institut National de l'Audiovisuel
4, Avenue de l'Europe
94366, Bry sur Marne Cedex
Tel : 01 49 83 23 18
Fax : 01 49 83 25 82
Email : bbachimont@ina.fr*

Résumé

La représentation formelle et symbolique des connaissances repose sur des langages formels composés d'une part de règles syntaxiques génératives pour la construction de formules et d'autre part de règles de composition sémantique associant aux formules construites une signification. La construction syntaxique assemble les formules à partir de primitives dont le sens se compose pour donner la signification des formules. La tâche des ontologies est définir quelles primitives munies de leur signification sont nécessaires pour la représentation des connaissances dans un contexte donné.

Définir une ontologie est une tâche de modélisation menée à partir de l'expression linguistique des connaissances. La modélisation s'effectue en trois étapes, correspondant à trois engagements : un engagement sémantique, fixant le sens linguistique des concepts, un engagement ontologique fixant leur sens formel et enfin un engagement computationnel déterminant leur exploitation effective.

Nous présentons ici ces trois types d'engagement et explicitons les étapes de l'élaboration d'une ontologie.

Mots clés : Ontologie, terminologie, engagement sémantique, engagement ontologique, sémantique formelle, sémantique différentielle.

Abstract

The symbolic and formal representation of knowledge relies on formal languages. Formal languages consist of generative syntactic rules, that build formula, and of semantic compositional rules that associate formula with meaning. The syntactic building of formula and their compositional semantics rely on primitives. The aim of ontologies is to define which primitives must be used in knowledge representation tasks.

Defining an ontology is a modelling task starting from the linguistic expression of knowledge. Ontology modelling is a three-step process. First, the semantic commitment specifies the linguistic meaning of concepts. Second the ontological commitment specifies their formal meaning. Finally, the computational commitment specifies their effective computer-based use.

This paper presents those three commitments and explains the main steps that must be followed in ontological modelling.

Key Words: Ontology, terminology, semantic commitment, ontological commitment, formal semantics, differential semantics.

1 Introduction

L'informatique a ceci de plaisant qu'elle permet d'automatiser bon nombre de tâches ou, tout au moins, d'en assister considérablement la réalisation. Il faut pour cela modéliser la tâche à effectuer ou le problème à résoudre de manière à en obtenir une représentation formelle. C'est cette représentation formelle qui est mise en œuvre par la machine pour obtenir le comportement souhaité ou le résultat recherché.

Si l'informatique étudie comment rendre effective la spécification formelle d'une résolution, elle ne préjuge pas de la manière d'obtenir une telle spécification. D'autres disciplines se chargent de modéliser le problème à résoudre ou la tâche à effectuer de manière à obtenir la représentation formelle à partir de laquelle l'informatique peut travailler.

L'ingénierie des connaissances fait partie de ces disciplines. Une manière de la caractériser est de la définir comme la modélisation formelle de problèmes pour lesquels les seules connaissances dont on dispose sont de nature linguistiques ou cognitives. L'objectif est de spécifier le comportement calculatoire d'une machine à partir de la donnée de tels problèmes. La modélisation formelle consiste dans la représentation à l'aide d'un langage formel des connaissances du domaine. Par la définition même de ce qu'est un langage formel, il est nécessaire de traiter le problème de la « modélisation ontologique ». On entend par ce terme le fait de définir quelles sont les primitives de représentation et leur signification qui seront utilisées pour la modélisation formelle des connaissances.

Cet article propose une définition des ontologies et une méthode pour les élaborer. Cette méthode comprend plusieurs étapes permettant de passer de l'expression linguistique des connaissances, telles qu'elles nous autres humains pouvons les considérer, à une représentation formelle et calculable des connaissances, propres à une exploitation informatique. Ces étapes permettent d'articuler des notions d'engagement sémantique avec celles plus communes d'engagement ontologique et de treillis de concept. Nous donnons un exemple et concluons sur les problèmes que posent la généralité des ontologies et la linguisticité des connaissances. Cette méthode a été utilisée pour la représentation de connaissances médicales et a permis d'élaborer une ontologie de plus de 2000 concepts effectivement utilisés par un système de compréhension de comptes rendus. Cette méthode est en cours d'expérimentation à l'INA pour la description structurée des contenus audiovisuels.

2 Ontologie et représentation des connaissances

Le problème posé par la représentation des connaissances est de formaliser dans un langage formel de représentation les connaissances permettant de traiter le problème à résoudre. Traditionnellement, un langage formel se définit à partir d'un alphabet comprenant des symboles de variables, des symboles de fonction et des symboles de prédicat et des symboles d'opérateurs logiques (connecteurs), des quantificateurs, etc.

Ces symboles sont des primitives à partir desquelles toutes les formules bien formées du langage sont construites selon des règles de construction syntaxique définies par le langage lui-même. Mais il ne suffit pas de préciser comment construire les formules du langage, il faut en outre préciser quelle signification ou sémantique elles peuvent recevoir, pour leur associer des connaissances du domaine. C'est pourquoi un langage formel se voit pourvu d'une sémantique formelle précisant quelles significations associer aux formules en fonction des significations des symboles qu'elles contiennent et de la manière dont ils sont assemblés.

Il faut distinguer les primitives logiques des primitives non logiques. Les primitives logiques ont leur sémantique spécifiée par le langage formel lui-même, et possèdent une

signification indépendante du domaine. C'est en quelque sorte une sémantique intégrée dans le langage. Les primitives non logiques possèdent une signification dépendante du domaine : non spécifiée par le langage, elle doit être explicitement déterminée en fonction du domaine.

Ainsi, la sémantique des connecteurs logiques et de la composition des symboles de fonction avec les symboles de variables est propre au langage et ne dépend pas de ce que les symboles représentent et donc du domaine concerné. Par exemple, la sémantique formelle associée au langage spécifie les tables de vérité des connecteurs une fois pour toutes. En revanche, les symboles de fonction et de prédicat n'ont pas de sémantique définie par le langage. Quelle fonction représente un symbole de fonction, quel prédicat représente un symbole de prédicat n'est pas spécifié par la définition du langage formel.

Prenons par exemple un langage formel du 1^{er} ordre. Soit la formule suivante : Douleur(x) ET Fièvre(x). La sémantique formelle des langages du 1^{er} ordre spécifie que P et Q sont des prédicats, donc des fonctions d'un domaine D dans les valeurs de vérité {Vrai, Faux}. ET est un connecteur logique dont la table de vérité spécifie que F1 et F2 est vraie ssi F1 et F2 sont toutes deux vraies. Mais cette sémantique ne spécifie pas quels prédicats Fièvre et Douleur doivent représenter dans le domaine, même si le libellé de ces prédicats en donne quelque idée.

La donnée d'un langage formel laisse donc ouvert le problème de définir quels symboles fonctionnels et relationnels sont nécessaires et quelle sémantique il convient de leur associer. Ce problème est celui des ontologies que l'on peut donc définir de la manière suivante :

Définir une ontologie pour la représentation des connaissances, c'est définir, pour un domaine et un problème donnés, la signature fonctionnelle et relationnelle d'un langage formel de représentation et la sémantique associée.

Contrairement aux définitions que l'on peut rencontrer çà et là dans la littérature (par exemple (Gruber, 1993)), cette définition a l'avantage d'être à la fois précise et rigoureuse et de renvoyer directement à la nature même de la formalisation et la représentation des connaissances.

Ce n'est qu'une fois l'ontologie définie qu'il est possible d'associer une sémantique dans le domaine aux constructions syntaxiques du langage. Il n'est possible de représenter des connaissances, c'est-à-dire d'exprimer dans un langage formel les connaissances du domaine, que si les formules et les primitives non logiques qu'elles contiennent sont pourvues d'une sémantique qui permette de savoir quelle connaissance est assumée par cette formule. Autrement dit, la représentation des connaissances n'est possible que quand le problème des ontologies est résolu.

3 Définir des primitives : l'engagement sémantique

Le problème est de définir les primitives non logiques d'un langage de représentation et la sémantique associée. Il faut pour cela déterminer quelles sont les notions élémentaires à partir desquelles toutes les connaissances du domaine sont construites. Autrement dit, définir les primitives non logiques du langage, c'est trouver les primitives du domaine.

Le problème est que, de manière générale, il n'existe pas de primitives dans un domaine. Il est impossible de trouver, par exemple en médecine, des notions à partir desquelles toutes les connaissances sont construites. Les notions sont elles-mêmes des connaissances permettant de définir d'autres connaissances, mais sont également définies par elles. Le problème est donc qu'il n'existe pas de primitives générales d'un domaine et qu'il est nécessaire pourtant de disposer de primitives pour entreprendre la représentation des connaissances.

Il faut par conséquent modéliser les primitives nécessaires pour la formalisation et la représentation du problème à résoudre et des connaissances s'y rapportant. La définition d'une ontologie n'est donc pas la caractérisation ou la détermination de primitives déjà existantes dans un domaine, mais la modélisation ou construction de primitives pour la

résolution du problème. Comment construire alors des primitives ? Le principe est de repartir de l'expression linguistique des connaissances du domaine.

1.1 La normalisation sémantique

Les primitives nécessaires à la représentation des connaissances doivent être modélisées à partir des données empiriques dont on dispose, à savoir l'expression linguistique des connaissances. Le travail de modélisation doit s'effectuer à partir de documents attestés dans la pratique d'un domaine et rassemblés en un corpus.

Le corpus se constitue de documents produits dans le contexte où le problème à résoudre se pose. Ce sont par exemple des documentations techniques, des ouvrages de références, des documents de travail, des manuels propres au domaine ou à l'industrie concernée, ou bien encore la transcription d'interviews menées avec des spécialistes.

Parmi ces documents, il faut sélectionner ceux à partir desquels l'analyse sera menée pour élaborer l'ontologie. La sélection s'effectue en fonction de critères qui relèvent tant de la méthode d'analyse qui sera utilisée (analyse de corpus par exemple) que du problème à résoudre (ne retenir que les documents pertinents pour le problème à résoudre).

La constitution d'un corpus est très délicate de manière générale car le corpus conditionne largement le type et la nature des traitements que l'on peut effectuer sans que l'on ait forcément loisir de choisir le type de données le plus adéquat. Le choix d'un corpus introduit des biais sans qu'il soit toujours loisible de les apprécier.

Le corpus comporte l'expression des notions qu'il faut modéliser. Une première tentative est donc de considérer que les unités linguistiques utilisées sont des concepts. Autrement dit, on définit des concepts par des libellés linguistiques. Par exemple, l'expression de « virose » paraît importante (selon un expert, ou selon les résultats d'une analyse de corpus proposant des candidats termes (Bourigault, 1994)) et l'on retient le libellé de « virose » pour définir un concept. Cela présente un avantage et un inconvénient.

L'avantage est que le concept reçoit d'emblée une interprétabilité dans le domaine par les spécialistes qui l'utilisent ou le consultent. Par exemple, un médecin considérant un concept dont le libellé est « angioplastie », ou une documentaliste utilisant un concept dont le libellé est « personnage politique », associent toujours une signification à ces libellés qui leur permettent de l'utiliser. L'inconvénient est que, si ces libellés sont interprétables, rien n'impose qu'ils soient interprétés de la même manière ou à tout le moins de manière cohérente et compatible entre plusieurs spécialistes. En effet, le libellé « histoire de la maladie » évoque quelque chose de différent pour chaque médecin : comment le différencier de cet autre libellé qu'est par exemple « antécédents du malade » ? Comment s'assurer que tout spécialiste mobilisera ces libellés de la même manière ? Le danger n'est pas en effet que ce que dira l'un ne sera pas compris par l'autre : tout médecin sait ce qu'est l'histoire de la maladie et les antécédents de la maladie. Mais rien n'assure qu'un médecin A utilisera ces notions comme son confrère B.

Par conséquent, il est nécessaire de contraindre l'interprétation spontanée que fait tout spécialiste de ces libellés pour que, respectant ces contraintes d'interprétation, tout spécialiste associe les mêmes significations que ses confrères à un libellé. Sinon, ce libellé ne peut être considéré comme un concept, et ne correspond donc pas aux primitives recherchées.

Le problème est en effet que tout spécialiste, confronté au libellé d'un concept, l'interprète en fonction du contexte dans lequel il se trouve. Par conséquent, ce libellé ne peut être utilisé comme une primitive puisque, par définition, une primitive doit être définie de manière non contextuelle, indépendamment des autres primitives. En contraignant l'interprétation effectuée par les spécialistes, la même signification est associée quel que soit le contexte, c'est-à-dire indépendamment du contexte. C'est à cette condition que le libellé, pourvu de

cette signification, peut fonctionner comme une primitive et être mobilisé pour la représentation formelle des connaissances.

Le problème est donc de partir de la sémantique de la langue naturelle pour arriver à la définition non contextuelle d'un libellé. De quelle nature est la sémantique de la langue naturelle ? Nous nous appuyons pour la caractériser sur les travaux de F. Rastier (Rastier et al., 1994) et adoptons le paradigme différentiel. En effet, plusieurs paradigmes se présentent dès lors que l'on veut caractériser la sémantique des langues :

1. Le paradigme *référentiel*, qui associe à chaque unité linguistique une référence, un objet pris dans un univers de référence (référence extensionnelle) ou un concept (référence intensionnelle) ; comprendre un mot, c'est trouver l'objet qui lui correspond.
2. Le paradigme *psychologique* qui associe à chaque unité linguistique une représentation psychologique ou une « image » mentale ; comprendre un mot, c'est s'en faire une représentation.
3. Le paradigme *différentiel*, qui associe à chaque unité linguistique les unités voisines en langue (celles qui sont utilisées en même temps qu'elle dans les contextes d'usage) pour la définir par les identités et différences qu'elle entretient avec ses voisines.

Les paradigmes référentiel et psychologique sont *extralinguistiques* dans la mesure où ils associent une unité linguistique avec une entité non linguistique (objet ou représentation). Cela pose le problème de définir ce que sont les objets ou les représentations associés aux unités linguistiques. Cela entraîne la linguistique soit dans une ontologie (au sens philosophique) où il faut définir les objets du monde, soit dans la gnoséologie (théorie de la connaissance et de la cognition) où il faut définir ce que sont les représentations psychologiques associées à la compréhension linguistique. Au lieu de d'assigner un domaine précis et autonome à la linguistique pour la pourvoir ensuite d'une théorie, les paradigmes référentiel et psychologique plonge la linguistique dans des problèmes plus vastes qui remettent à un horizon indéterminé la détermination d'un objet et d'une méthode. La linguistique n'est pas une théorie de l'Être ni une théorie de la connaissance et n'a pas à l'être, contrairement à ce qu'affirment par exemple la philosophie analytique ou formelle d'une part et les sciences cognitives d'autre part.

Le paradigme différentiel est intralinguistique : il définit une unité linguistique par d'autres unités linguistiques. Les différences et identités entre unités sont définies dans la langue et peuvent être décrites à partir d'un corpus qui atteste l'usage et définit les voisins en contexte. La sémantique différentielle détermine le signifié des unités linguistiques en termes de traits différentiels. De manière théorique, un trait différentiel est l'extrémité d'une relation binaire d'opposition ou d'identité, que l'on appelle *sème*. On définit ainsi plusieurs types de sèmes. Les sèmes *génériques* attestent l'appartenance d'une unité à une classe. Les sèmes génériques sont des traits de signification possédés par plusieurs unités. Cette commune possession détermine une classe définie par ces traits. Par exemple, les scalpels et les bistouris sont des instruments permettant d'inciser la peau. Les sèmes *spécifiques* déterminent les différences permettant de distinguer les unités linguistiques appartenant à une même classe. Ainsi, un scalpel n'est pas la même chose qu'un bistouri, car il est pour les morts et le bistouri pour les vivants ; l'un est un instrument chirurgical (bistouri), l'autre un instrument de dissection (scalpel). Il existe d'autre type de sèmes. Ce sont en particulier les sèmes inhérents et afférents, les premiers étant des sèmes attribués par défaut à un terme, mais éventuellement contredit par le contexte comme par exemple dans « la merlette blanche » où le trait inhérent de noirceur des merles est annulé par le contexte, les seconds étant des sèmes attribués par ce même contexte.

La caractérisation des sèmes permet ainsi de déterminer le signifié des unités employées dans un contexte donné. Mais on voit par exemple qu'une même unité ne possédera pas

toujours le même signifié selon que le contexte annule ou non ses sèmes inhérents et active ou non des sèmes afférents. Par conséquent, la description sémantique selon le paradigme différentiel permet de décrire le fonctionnement en corpus des libellés de concept, mais pas de dégager des primitives. Pour cela il faut donc passer d'une description contextuelle à la détermination de signifiés non contextuels.

La détermination de signifiés non contextuels est une normalisation sémantique : on fixe parmi les significations possibles que peut recevoir une unité en contexte celle qui doit être à chaque fois associée. Cela revient en fait à choisir un contexte de référence dans lequel par principe les termes doivent s'interpréter. C'est dans ce contexte que l'attribution des sèmes s'effectue. Autrement dit, il s'agit d'annuler l'effet du contexte : les sèmes inhérents ne peuvent plus s'annuler, les sèmes afférents ne doivent pas s'activer. On retrouve ainsi les conditions d'une sémantique pour laquelle les unités signifient toujours la même chose quels que soient les énoncés qui les contiennent. Ce qui n'est donc pas vrai d'un point de vue linguistique doit être imposé par la normalisation sémantique.

La normalisation linguistique est le choix d'un contexte de référence, celui de la tâche ou du problème qui motive l'élaboration d'une représentation formelle des connaissances. Le point de vue de la tâche permet au modélisateur de fixer quelle doit être la signification de l'unité linguistique considérée.

Qui peut effectuer la normalisation sémantique ? Qui faut-il mettre derrière le terme de « modélisateur » ? Indiscutablement, il faut un locuteur de la langue du domaine, qui explicite le sens des termes et les décline selon le contexte ; c'est un spécialiste ou expert. Par ailleurs, il faut structurer ces sens et significations, comme le fait un linguiste, c'est l'ingénieur de la connaissance. On retrouve ainsi le schéma désormais classique de l'expert et du cogniticien avec quelques nuances cependant : l'expert ne donne pas ses connaissances personnelles et individuelles du domaine, mais il intervient au même titre qu'un autre spécialiste du domaine. Il ne délivre pas ses connaissances propres, mais ce que signifient, selon lui, les expressions linguistiques du corpus. Par ailleurs, l'ingénieur de la connaissance n'est pas là pour faire accoucher l'expert de connaissances dont il n'a pas forcément conscience, mais pour structurer et mettre en forme selon une méthodologie les expressions linguistiques pour que, normalisées, elles puissent tenir lieu d'expressions de concepts.

De quelle méthodologie dispose l'ingénieur de la connaissance ? La difficulté est que les choix effectués pour chaque unité soient cohérents entre eux et permettent de construire un système de représentation. C'est là le rôle d'une méthodologie. La méthodologie que nous proposons ici repose sur l'organisation générale des unités en un réseau d'identités et de différences. Ce sont les propriétés structurelles de ce réseau qui permettent de contraindre l'interprétation des unités définies dans le réseau : la position d'une unité dans le réseau prescrit comment la comprendre et lui prescrit une signification qui pourra dès lors lui être associée quel que soit le contexte où elle se rencontre.

1.2 La structuration globale : l'arbre ontologique

La sémantique différentielle permet de décrire les unités entre elles par les identités qui les unissent et les différences qui les distinguent. Or, les unités décrites partagent toutes le fait d'être comparables entre elles : elles partagent le fait de dire ou d'être quelque chose pour que l'on puisse déterminer dans un second temps qu'elles ne sont pas la même chose. Par conséquent, toutes les unités se déterminent à partir d'une unité générique ultime, une unité racine, à laquelle elles appartiennent.

Par ailleurs, toutes les unités se déterminent d'une part par l'unité générique à laquelle elles appartiennent et d'autre part par les différences qui les distinguent. Cela signifie que le réseau des unités est un réseau d'héritage de propriétés où les unités filles héritent des sèmes d'une unité générique mère.

La question qui se pose alors est de savoir quelle peut être la structure d'un tel réseau d'héritage. En fait, la structure ne peut être qu'arborescente. Supposons en effet qu'une unité U ait plusieurs unités parentes directes, disons 2. Ces 2 unités parentes sont distinctes, elles se déterminent par identités et différences entre elles. Si elles ne se caractérisent que par les identités qu'elles possèdent l'une avec l'autre, cela implique que l'une est générique par rapport à l'autre et donc que l'une seulement des deux est une parente directe. Elles se caractérisent donc par les oppositions qui les distinguent. Cela implique qu'elles se déterminent par des propriétés ou traits mutuellement exclusifs : par exemple, que le scalpel est pour les morts, c'est-à-dire pas pour les vivants (bistouri). Si ces unités sont parentes directes de U, U doit hériter de ces propriétés mutuellement exclusives. Ce qui est contradictoire : par exemple, un instrument qui hérite de scalpel et bistouri doit vérifier le fait d'être pour les vivants et pas pour les morts, et pour les morts et pas pour les vivants. Par conséquent, une même unité ne peut avoir qu'une et une seule unité parente. Le réseau doit donc être un arbre.

1.3 La structuration locale : les principes différentiels

Dès lors que la structure du réseau est un arbre, il devient aisé de déterminer la signification que doit posséder une unité en fonction de sa position dans l'arbre. En effet, selon le paradigme différentiel que nous avons adopté, la signification d'un nœud se détermine en fonction de ses plus proches voisins. Or, dans un arbre, les plus proches voisins sont les d'une part l'unité parente et d'autre part les unités sœurs. Il faut donc déterminer la signification d'un nœud en fonction de son parent et de ses frères. On définit pour cela 4 principes fondamentaux, les principes différentiels, qui imposent d'explicitier en fonction des voisins les identités et différences qui définissent le nœud. Ces principes sont :

- Le principe de communauté avec le père :
Toute unité se détermine par l'identité qu'elle possède avec l'unité parente. Il faut expliciter en quoi l'unité fille est identique à l'unité parente. C'est, *mutatis mutandis*, le principe aristotélien de définition par le genre proche.
- Le principe de différence avec le père :
Toute unité se distingue de l'unité parente, sinon il n'y aurait pas lieu de la définir. Il faut donc expliciter la différence qui la distingue de l'unité parente. C'est, *mutatis mutandis*, le principe aristotélien de définition par la différence spécifique.
- Le principe de différence avec les frères :
Toute unité se distingue de ses frères sinon il n'y aurait pas lieu de la définir. Il faut donc expliciter la différence de l'unité avec chacune des unités sœurs. Ce principe n'est pas aristotélien, mais provient du paradigme différentiel choisi.
- Le principe de communauté avec les frères :
Toutes les unités filles d'une unité parente possède par définition un même trait générique, celle qu'elles partagent avec l'unité parente. Mais il faut établir une autre communauté entre les unités filles ; c'est celle qui permet de définir des différences mutuellement exclusives entre les unités filles. Par exemple, l'unité parente est « être humain », et les unités filles sont « homme » et « femme ». Ces unités partagent le fait d'être des humains. Mais cette propriété ne permet pas de définir en quoi sont différents les hommes et les femmes. On choisit alors comme principe de communauté la sexualité, où l'on peut attribuer à « homme » le trait masculin et à « femme » le trait féminin. Ces deux traits sont mutuellement exclusifs car ce sont deux valeurs possibles d'une même propriété.

Le dernier principe n'a pas exactement la même utilité que les 3 autres. En fait, il n'intervient que pour rendre le 3^e principe possible. Il repose sur le fait qu'il ne suffit pas de poser des propriétés différentes pour caractériser les différences les unités filles, mais qu'il faut savoir en quoi elles sont différentes.

1.4 Concepts et relations

Nous venons de préciser comment définir des concepts. Ces concepts se définissent en termes d'identités et de différences, c'est-à-dire en termes de relations. Les relations doivent également être définies dans l'ontologie. Cependant, elles ne se définissent pas de la même manière que les concepts, car, unissant des concepts, elles se caractérisent à partir d'eux. Si l'on ne retient que des relations binaires, les relations se définissent de la manière suivante :

- Une relation se définit par les concepts qu'elle relie : par exemple, « être animé » et « action » ; ces concepts constituent la *signature sémantique* de la relation.
- Elle se définit en outre par un contenu sémantique intrinsèque articulant les deux concepts : par exemple, le fait que l'être animé est l'agent de l'action. La sémantique intrinsèque de la relation est spécifiée vis-à-vis des autres relations possédant la même signature sémantique selon les principes différentiels vus plus haut. Par exemple, la relation « patient » entre être animé et action se définit par identité et différence avec la relation agent. L'identité, c'est le fait d'avoir la même signature, la différence, c'est le fait de subir l'action plutôt que de l'exercer. L'identité n'est pas forcément réduite au fait d'avoir la même signature : par exemple, la relation « agent volontaire » et « agent involontaire » possèdent comme identité outre la même signature le fait d'avoir un père commun, la relation « agent ».

Autrement dit, chaque signature sémantique est potentiellement la racine d'un arbre différentiel de relations possédant la même signature et spécifiées selon les principes différentiels. Les signatures sémantiques constituent également un arbre : on a donc un arbre de relations venant compléter l'arbre des concepts.

1.5 Engagement sémantique et modélisation des primitives

Les principes différentiels associés à un nœud de l'arbre ontologique explicitent en termes linguistiques, empruntés au domaine, ce qu'il faut comprendre à la lecture du libellé du nœud. Les principes différentiels constituent donc une grille de lecture et sont des prescriptions interprétatives qu'il faut suivre pour savoir comment interpréter le libellé. C'est donc le respect de ces principes qui permet de considérer ce libellé non pas comme une unité linguistique dont le sens varie selon le contexte de son utilisation, mais comme une primitive au sens invariable.

En effet, quand un utilisateur aborde l'arbre ontologique, ce qu'il considère, c'est avant tout une structure arborescente de termes dont le libellé est emprunté à la langue du domaine. Par conséquent, l'utilisateur aborde les concepts comme des unités linguistiques, comme des mots, qu'il interprète selon ses propres habitudes. Par conséquent, à chaque utilisateur son interprétation. Pour surmonter la variabilité de ces interprétations, les principes différentiels prescrivent à l'utilisateur comment préciser et ajuster le sens qu'il attribue aux libellés qu'il considère. Ainsi, les principes n'attribuent pas du sens à des libellés qui en serait par ailleurs dépourvus. Mais les principes raffinent et ajustent un sens que les utilisateurs attribuent spontanément aux nœuds de l'arbre ontologique, puisqu'ils sont des locuteurs de la langue du domaine et que les libellés des nœuds sont empruntés à cette langue.

Le contexte interprétatif d'un nœud n'est pas constitué seulement des principes différentiels associés à ce nœud, mais aussi des principes associés aux nœuds composant la branche le reliant à la racine de l'arbre ontologique. Ainsi, le sens d'un libellé est déterminé à partir de sa position dans l'arbre. On obtient un réseau dans lequel la position d'un nœud conditionne sa signification. La signification définie par la position dans l'arbre est invariable selon les contextes. Le libellé peut alors être utilisé comme une primitive. En respectant les principes différentiels, en s'engageant à suivre la sémantique qu'ils prescrivent, les nœuds de l'arbre ontologique correspondent à des concepts pouvant être utilisés comme des primitives de modélisation et de formalisation. Nous venons donc de définir l'engagement sémantique à

la base de l'ontologie : ensemble des prescriptions interprétatives qu'il faut respecter pour que le libellé fonctionne comme une primitive.

1.6 Ontologie régionale et engagement sémantique

On peut contester notre usage du paradigme différentiel pour modéliser une ontologie fondée sur l'engagement sémantique. En effet, le paradigme différentiel, intralinguistique, récuse de ce fait qu'il soit possible de déterminer le sens d'une unité linguistique par rapport à autre chose que le linguistique lui-même. Le sens d'une unité linguistique se définit par d'autres unités linguistiques, et non par un concept, une référence ou une représentation psychologique. Par conséquent, la seule donnée disponible pour déterminer le sens est l'usage linguistique lui-même, c'est-à-dire l'usage en contexte. Le paradigme différentiel a par conséquent comme but de définir le sens en contexte et de rendre compte des variations contextuelles. Autrement dit, derrière un mot, il n'y a pas un type, existant *a priori*, dont il faudrait décrire les altérations en contexte, mais seulement des occurrences, dont on abstrait, à des fins théoriques ou lexicographiques, un type : les définitions du dictionnaires ne sont que des constructions théoriques du linguiste, et non des réalités précédant l'usage en contexte des unités linguistiques (Rastier, 1987).

Or, notre méthodologie détourne d'une certaine manière le paradigme linguistique puisque, de la description du sens en contexte, on veut normaliser un sens non contextuel pour dégager des primitives de modélisation.

La contradiction n'est qu'apparente. En effet, la normalisation sémantique n'est pas une entreprise visant à retrouver derrière les variations contextuelles une signification unique, *le* sens de l'unité linguistique. Cela reviendrait en effet à sortir du paradigme différentiel puisque cela consiste à affirmer que l'unité linguistique signifie *a priori* un concept avant d'être plongée dans un contexte. Au contraire, la normalisation sémantique construit *une* signification en privilégiant un contexte particulier, celui de la tâche considérée. La normalisation part donc de la donnée empirique disponible, le corpus, pour analyser les variations contextuelles qu'il comporte du point de vue d'un contexte privilégié pour déterminer les sens possibles en une signification adaptée à la tâche.

C'est pourquoi l'engagement sémantique ne dégage pas le sens universel des termes, une ontologie universelle qui serait cachée derrière leurs usages variés. L'engagement sémantique dégage une ontologie valable seulement localement, *régionalement*, dans le cadre d'un domaine et d'une tâche. C'est pourquoi l'engagement sémantique normalise une *ontologie régionale*, et reflète une structure seulement valable pour accomplir des actions dans un domaine donné, et ne construit pas une ontologie universelle ou formelle, qui refléterait les lois universelles de la pensée (l'ontologie formelle de Husserl reprise par la philosophie formelle de Cocchiarella (Cocchiarella, 1991) et Guarino (Guarino, 1995)) ou du réel.

4 Formaliser des connaissances : l'engagement ontologique

L'arbre ontologique et l'engagement sémantique qu'il explicite mettent à disposition des primitives. Il est alors possible de représenter formellement des connaissances puisque la modélisation ontologique a dégagé ce que la formalisation des connaissances présuppose comme prérequis. Il devient possible de définir une sémantique formelle pour les concepts et de retrouver les propriétés auxquelles on est habitué en représentation des connaissances.

En effet, en fixant la signification des concepts, l'arbre ontologique fixe la référence des termes. Autrement dit, il devient possible de définir un terme par une sémantique référentielle, c'est-à-dire une sémantique formelle. Considérons à présent l'ensemble des concepts compris non plus comme des termes linguistiques dont il faut prescrire l'interprétation, se conformant à un engagement sémantique, mais comme des primitives définies par cet engagement sémantique. Appelons les concepts avant l'engagement sémantique les concepts sémantiques,

et « concepts formels » ceux correspondant au respect de l'engagement sémantique. Chaque concept possède donc une sémantique formelle qui lui associe un ensemble de référent. La question qui se pose est alors la suivante : quelle est la structure qui relie les concepts entre eux ? Plusieurs éléments de réponse se dégagent.

Premièrement, les concepts formels vérifient les relations d'identité unissant les concepts sémantiques. En effet, les concepts sémantiques s'interdéfinissent par identités et différences. L'identité correspond au fait qu'une notion est comprise dans une autre : la notion de bistouri comprend celle d'instrument d'incision. Par conséquent, tout objet qui est un bistouri est un instrument d'incision : le concept formel de scalpel hérite de celui d'instrument d'incision, et son extension est comprise dans celle d'instrument d'incision.

Deuxièmement, la différence entre concepts sémantiques, où deux notions s'excluent mutuellement, ne se répercute pas directement sur les concepts formels. Prenons un exemple. On a défini les notions d'acteur et d'être humain. Ces notions s'excluent : la notion d'acteur correspond à un rôle, celle d'être humain à une entité biologique. Or, une entité biologique n'est pas un rôle, même s'il est dans sa nature de pouvoir remplir un rôle. Mais, dès lors que l'on adopte une sémantique formelle, on ne considère plus des notions, mais des extensions d'objet. Par conséquent, un acteur, c'est l'ensemble des objets qui sont des acteurs, c'est-à-dire qui jouent un rôle. Les êtres humains sont également un ensemble d'objets. Le fait que les notions s'excluent implique que les extensions sont différentes. Mais pas qu'elles soient disjointes ! Et, à l'évidence, un objet qui est un homme peut être un objet qui est un acteur, même si la notion d'acteur n'est pas celle d'homme. Cela implique que l'on retrouve dans les relations unissant les concepts formels les relations d'héritage des concepts sémantiques, mais pas les exclusions. La structure des concepts formels n'est plus obligatoirement un arbre, mais plus généralement une structure de treillis. Cela se comprend d'ailleurs facilement pour la raison suivante : si la sémantique des concepts est référentielle, les relations entre les concepts sont des relations entre ensemble. La structure des concepts formels doit correspondre à la structure algébrique des ensembles, c'est-à-dire un treillis.

La formalisation des relations se définit alors facilement : une relation se définit comme un sous-ensemble du produit cartésien ces extensions des concepts composant sa signature.

On retrouve ainsi les principales affirmations de la littérature, en particulier celles de Sowa (Sowa, 1984) à propos de la structure des types de concepts, jouant le rôle d'ontologie pour ses graphes conceptuels, à savoir qu'il s'agit d'un treillis de types. La raison en est que la sémantique des concepts de Sowa n'est pas une sémantique linguistique différentielle comme nous l'avons définie, mais une sémantique intensionnelle définie en termes de sémantique référentielle (l'intension d'un concept, c'est le fait d'avoir une extension dans différents mondes possibles). On retrouve également la notion d'engagement ontologique proposée par Guarino (Guarino, 1994) : pour utiliser un concept de manière conforme à son sens, il faut respecter un engagement ontologique. La sémantique adoptée par Guarino est une sémantique référentielle où un concept est associé à une extension d'objets. Respecter le sens d'un concept, c'est s'engager à ce que lui corresponde une extension d'objets existants dans l'univers d'interprétation. Il s'agit donc bien d'un engagement *ontologique*, puisque c'est l'existence d'objets qui est prescrite par le sens du concept.

L'engagement ontologique permet ainsi de définir ce que nous appelons une *ontologie formelle* ou *ontologie référentielle*. Une ontologie référentielle est constituée de prédicats formels pourvus d'une sémantique référentielle ou extensionnelle. Il ne faut pas confondre notre ontologie référentielle avec l'ontologie formelle de la philosophie formelle qui y voit une théorie du quelque chose en général (Cocchiarella, 1991). Héritée de Husserl (Husserl, 1957, Husserl, 1950), l'ontologie formelle caractérise les lois reliant que doivent vérifier nécessairement toute théorie, indépendamment du domaine. C'est par exemple que « le tout est plus grand que la partie » et autres lois portant sur les structures à travers lesquelles nous

pensons le monde (théorie des ensembles, logique, etc.). Ce sont par conséquent les lois que vérifient les choses *en général*, indépendamment du fait de savoir de quelles choses particulières il s'agit. Une chose, si elle est une chose, vérifie nécessairement les lois de l'ontologie formelle : c'est le *modèle* de la chose en général. L'ontologie référentielle définie ici *n'est pas un modèle*, en particulier un modèle du domaine. C'est la formalisation du modèle sémantique que constitue l'ontologie régionale. Pour être un modèle, l'ontologie formelle doit s'adosser sur une ontologie régionale qui détermine quelle signification donner aux libellés des prédicats formels pour qu'ils puissent être compris comme des primitives de modélisation. Par conséquent, il y aura autant d'ontologies référentielles que d'ontologies régionales. Alors que, par définition, il ne peut y avoir qu'une seule ontologie formelle pour la philosophie formelle.

5 Utiliser une formalisation

Formaliser les connaissances ne suffit pas, il faut les utiliser dans un système opérationnel. Or, un système n'utilise pas les concepts en fonction de leur interprétation sémantique, accessible seulement aux locuteurs de la langue (c'est-à-dire, jusqu'à nouvel ordre, aux humains), ni encore en fonction de leur interprétation formelle, puisque le système n'accède pas aux objets constituant l'extension des concepts. Un système ne peut exploiter un concept qu'en fonction des opérations ou règles qu'il peut lui associer. La sémantique permettant à un système d'utiliser un concept est par conséquent la spécification informatique des opérations applicables au concept. Cette spécification est variable selon le type de système considéré. Ce peut être par exemple un graphe conceptuel qui, explicitant le contenu d'un concept, spécifie en fait les opérations de jointure que l'on peut effectuer. Nous appelons cette sémantique « sémantique computationnelle », car elle considère chaque concept comme la spécification d'un calcul. Nous n'insistons pas plus, car cette dimension est bien connue dans la mesure où tout système à base de connaissance est en fait et avant tout un système associant une sémantique computationnelle à des concepts exprimés à travers des identifiants syntaxiques exploités dans le calcul. En associant une sémantique computationnelle aux concepts de l'ontologie, on définit une *ontologie computationnelle*.

6 Un exemple

Prenons un exemple, totalement de circonstance, retraçant les étapes et les objets que nous venons de présenter. Il s'agit d'une tâche d'indexation de contenus audiovisuels où il faut déterminer les primitives de description. Le but de la tâche est de permettre de retrouver les séquences audiovisuelles à travers une requête portant sur les index. On part d'un corpus constitué de paraphrases linguistiques de contenus AV.

Corpus	John Wayne est acteur. Il apparaît dans plusieurs segments temporels du film à analyser avec plusieurs autres acteurs. Mais ces autres acteurs sont mineurs.
Analyse du corpus	1 ^{ère} phrase : la notion d'acteur apparaît. Elle peut être comprise comme une propriété de John Wayne, ou comme son type. C'est donc soit un rôle tenu par un être humain, soit un type d'être humain. 2 ^{ème} phrase : un acteur est objet qui apparaît durant des segments temporels avec d'autres objets. 3 ^{ème} phrase : les autres acteurs étant mineurs, il n'interviendront sans doute pas pour l'analyse et la description des segments. Par conséquent, seul John Wayne est pertinent pour la description.
Normalisation	Plusieurs choix sont possibles, selon le but de la tâche. S'il faut être

sémantique	capable de retrouver les images en fonction des acteurs mineurs, les acteurs, et en particulier John Wayne, sont des objets physiques localisés spatialement dans l'image. Si au contraire, seul John Wayne est important, c'est en fait un certain type de segment temporel, ceux où apparaît justement John Wayne. Donc le choix porte sur : John Wayne est un être humain dont la fonction est d'être acteur. John Wayne est un segment temporel d'un film.
Arbre ontologique	<pre> graph TD Entité --> Concepts Entité --> Relation Concepts -.-> Fonction Concepts -.-> EtreHumain[Etre humain] Concepts -.-> LocalisationAV[Localisation AV] Fonction --> Acteur Fonction --> Realisateur[Réalisateur] EtreHumain --> JohnWayne[John Wayne] LocalisationAV --> Segment LocalisationAV --> Objet Relation -.-> Apparaît </pre>
Principes	Les flèches en pointillés signifient que des nœuds manquent pour avoir un arbre ontologique bien formé. Illustrons donc les principes sur la « fonction » avec le concept « acteur » (resp. « réalisateur ») : P1 : indique une fonction d'un être humain dans un film P2 : jouer un rôle du film P3 : le réalisateur fait le film, l'acteur y joue P4 : apparaître <i>dans</i> le film versus travailler <i>sur</i> le film
Treillis formel	C'est le même arbre auquel on rajoute des définitions formelles aux concepts : la liste des acteurs pour « Acteur », la liste des êtres humains pour « Etre humain ». La liste des acteurs recoupe celle des êtres humains, on peut prendre leur intersection. On peut rajouter alors comme concept formel défini le concept « Personne-Acteur » ayant pour pères les concepts formels « Acteur » et « Etre humain », c'est-à-dire des concepts dont la sémantique est référentielle et déterminée par les engagements sémantique et ontologique. En revanche, « Personne-Acteur » n'est pas un concept sémantique : ce qui importe n'est pas son libellé linguistique dont on normalise la signification mais l'extension qui le définit comme concept formel. Les concepts formels définis se déterminent uniquement en fonction d'un engagement ontologique.
Treillis computationnel	C'est le treillis formel auquel on rajoute du code permettant l'opérationnalisation des concepts : [Etre-humain : JohnWayne]—(a_pour_fonction)→[acteur]

7 Expérimentations

La théorie présentée ici a été expérimentée dans le cadre de plusieurs expérimentations. Une première expérimentation fut la constitution d'une ontologie pour le système MENELAS (Zweigenbaum and Consortium-Menelas, 1995), système d'analyse de comptes rendus d'hospitalisation rédigés en langue naturelle. Ce système, construit à l'aide des graphes

conceptuels, était doté d'un arbre des types constitué selon la méthode décrite plus haut. Cet arbre comprenait environ 2000 concepts.

Plus récemment, nous avons mis en œuvre cette méthode dans les projets que nous menons à l'INA pour la description et la structuration des contenus audiovisuels (Projet Eurodelphes, MM). Il ne s'agit plus alors d'avoir des primitives pour le raisonnement, mais des primitives pour la description des contenus. Dans le cadre des travaux de normalisation menés par MPEG-7, nous avons proposé de considérer la déclaration de descripteurs et de leur sémantique par l'intermédiaire d'une ontologie (Bachimont, 1998).

8 Discussion

1.1 Problème des concepts non linguistiques

Plusieurs questions se posent. La première concerne le fait que toute connaissance portant sur le problème à traiter n'est pas toujours exprimable linguistiquement. En pratique, il faut souvent être capable de mêler la qualification linguistique d'un concept à une manière non linguistique de le caractériser. Par exemple, pour qualifier certains concepts portant sur la description des sons et de la musique, il n'est pas suffisant d'avoir une qualification linguistique. Ainsi, le concept de timbre se définit comme ce qui n'est ni la hauteur, ni l'intensité du son. Les principes différentiels risquent alors d'être de peu d'aide pour déterminer le contenu de la notion de timbre. En fait, il vaut mieux écouter le son, ou alors renvoyer à une méthode objective de définition du timbre comme un algorithme d'analyse. Le même problème se rencontre avec nombre d'attributs perceptifs, où la langue peine parfois à qualifier précisément les percepts.

Notre structuration en trois niveaux permet néanmoins de ne pas éluder le problème et d'en amorcer une solution. En effet, le vocabulaire descriptif des percepts, le timbre, le goût, etc., est fort riche. Ce vocabulaire se définit et se structure au niveau des concepts sémantiques. Néanmoins, ces concepts ne prennent leur sens complet que s'ils sont définis au niveau computationnel par les algorithmes qui peuvent les objectiver ou les renvoyer à un système de valeurs de mesure où ils peuvent se différencier plus nettement. Si l'on prend l'exemple des couleurs, la manière la plus objective de les différencier est la longueur d'onde. Mais une description linguistique est néanmoins pertinente et de toute manière fort usitée en pratique.

Mais il ne faut pas espérer de solution dans le cadre ontologique à ce problème fort complexe qui n'est rien moins que le rapport entre le perceptif et le discursif, le physique et le symbolique. Notons simplement que les définitions proposées ici n'éliminent pas le problème et permettent, quand par ailleurs on dispose des connaissances nécessaires, d'articuler les niveaux linguistiques et calculatoires.

1.2 Problème de la réutilisabilité des ontologies

Une seconde question concerne la genericité d'une ontologie. Elle est par définition limitée puisque la normalisation sémantique ne peut s'effectuer que dans la perspective d'une tâche donnée. Par conséquent, une ontologie ne peut être générique, elle est par essence toujours spécifique à une tâche. Cependant, cela ne préjuge pas de la question de savoir comment adapter une ontologie à une nouvelle tâche, ni de celle de savoir si une ontologie ne peut pas être adaptée à plusieurs tâches, prises en compte lors de la normalisation.

Cette question sur le plan théorique est assez complexe et ne peut être traitée ici. Nous renvoyons pour le fond donc à (Bachimont, 1996). Notons simplement que, en pratique, la réutilisabilité est souvent voire toujours possible. Par exemple, dans le projet Eurodelphes cité plus haut, l'ontologie de MENELAS, dans sa partie abstraite, a pu être réutilisée en totalité. Les autres parties n'étaient pas erronées pour la tâche de Delphes, mais sans objet. La raison

est que la description était suffisamment fine pour que le point de vue de Delphes puisse se représenter sans avoir à modifier le contexte. En effet une ontologie exprime des concepts de manière systématique, à certain niveau de finesse. Par exemple, on peut estimer qu'un député, un maire sont des types d'être humain, comme un médecin. On définit alors ces notions comme des spécialisations de « être humain ». Cependant, s'il faut évoquer un député qui est aussi un maire, on rencontre la difficulté que ces notions doivent s'exclure mutuellement. Par conséquent, le choix de modélisation effectué interdit qu'un maire soit un député. En revanche, si on définit ces notions comme des rôles politiques tenus par un être humain, le problème disparaît : la notion de maire est bien distincte de celle de député et elles s'excluent donc. Mais rien interdit qu'un homme soit député-maire... On voit donc que selon le niveau de détail, l'ontologie est robuste ou non. Par conséquent, nous formulons l'hypothèse suivante : *une ontologie peut être générique pour un ensemble de tâches données si elle possède un niveau de finesse de description permettant aux points de vue de chaque tâche de se représenter. Mais il n'existe pas de niveau de description permettant de décrire toutes les tâches du domaine. Il n'y a donc pas de généralité universelle ou absolue, mais de généralité relative à des tâches.*

Dans (Charlet et al., 1994), nous avons montré comment une ontologie, UMLS (McCray, 1989), n'avait pu être réutilisée pour le projet MENELAS. Cela provenait en particulier que UMLS ne reposait pas sur une systématique et une cohérence conforme aux prescriptions que nous faisons ici. Autrement dit, la tâche pour laquelle avait été conçu UMLS, essentiellement une tâche d'indexation, pouvait être réalisée sans adopter une méthodologie rigoureuse. Cependant, le travail effectué n'était valable que pour cette tâche. On peut alors formuler l'hypothèse complémentaire suivante : une ontologie sera d'autant plus facilement réutilisable qu'elle aura été conçue de manière systématique, selon les principes proposés ici.

9 Conclusion

Une ontologie est le résultat d'une modélisation. La modélisation porte sur la caractérisation de primitives pour la représentation formelle des connaissances. Ces primitives ne sont pas des données du domaine qu'il suffirait de déterminer, mais des constructions théoriques pour les besoins de la modélisation. Une ontologie se caractérise selon trois niveaux :

1. Le niveau sémantique ou interprétatif : l'ontologie régionale

L'ontologie est un arbre de concepts sémantiques. Un concept sémantique se caractérise par un libellé linguistique, emprunté à la langue du domaine, dont l'interprétation est contrainte par les principes différentiels, ceux qui lui sont directement associés plus ceux de ses ancêtres dans l'arbre. Ces principes correspondent à l'engagement sémantique qu'il faut respecter pour que le libellé ait un sens univoque et non contextuel et puisse être utilisé comme une primitive de représentation. Deux concepts sémantiques sont identiques si l'interprétation du libellé à travers les principes différentiels aboutit, pour chaque concept, à un sens équivalent.

2. Le niveau formel ou référentiel : l'ontologie référentielle

L'ontologie est un treillis de concepts formels. Les concepts formels se caractérisent par un libellé dont la sémantique se définit par une extension d'objets. Les concepts formels sont soit des concepts sémantiques dont on reprend le libellé et auquel on associe des référents conformément à l'engagement sémantique, soit de nouveaux concepts définis formellement par intersection de concepts formels déjà définis. Chaque concept formel est défini par un engagement ontologique qui spécifie quels objets doivent exister dans le domaine pour utiliser le concept conformément à sa signification formelle. Deux concepts formels sont identiques s'ils possèdent toujours la même extension.

3. Le niveau opérationnel ou computationnel : l'ontologie computationnelle

L'ontologie est un treillis de concepts computationnels. Les concepts computationnels se caractérisent par les opérations qu'il est possible de leur appliquer : ces opérations leur confèrent une sémantique dans le système effectif construit. L'engagement est alors computationnel : ce sont les opérations qu'il doit être possible d'appliquer au concept. Deux concepts computationnels sont identiques s'ils possèdent le même potentiel inférentiel.

Une ontologie est dépendante non seulement du domaine mais aussi de la tâche visée. En effet, c'est le contexte de la tâche qui permet de fixer les traits de signification pertinents des concepts sémantiques de manière à annuler les effets du contexte. Par conséquent, il ne peut exister d'ontologie universelle, ni même dans un domaine, une ontologie valant pour toutes les tâches possibles. En revanche, pour des tâches définies au préalable, il est possible de construire une ontologie valant pour toutes ces tâches. Il ressort, à la pratique, qu'il est toujours possible d'adapter une ontologie, sans qu'il soit possible de la réutiliser telle quelle.

10 Bibliographie

- Bachimont, B. (1996) *Herméneutique matérielle et Artéfacture : des machines qui pensent aux machines qui donnent à penser ; Critique du formalisme en intelligence artificielle*, Ecole Polytechnique.
- Bachimont, B. (1998) Ontologies and MPEG-7: an Editorial Perspective, In *VSMM'98 Workshop "Ontologies and MPEG-7"*Gifu, Japan.
- Bourigault, D. (1994) *LEXTER : un Logiciel d'EXtraction de Terminologie. Application à l'acquisition des connaissances à partir des textes*, Ecole des Hautes Etudes en Sciences Sociales.
- Charlet, J., Bachimont, B., Bouaud, J. and Zweigenbaum, P. (1994) Ontologie et réutilisabilité : expérience et discussion, In *Proceedings of the JAC'94 Conference*.
- Cocchiarella, N. (1991) Formal Ontology, In *Handbook of Metaphysics and Ontology*(Eds, Burkhardt, H. and Smith, B.) Philosophia Verlag, München.
- Gruber, T. (1993) A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, **5**, 199--220.
- Guarino, N. (1994) The Ontological Level, In *Philosophy and the Cognitive Sciences*(Eds, Casati, R., Smith, B. and White, G.) Hölder-Pichler-Tempsky, Vienna.
- Guarino, N. (1995) Formal ontology, conceptual analysis and knowledge representation, *International Journal of Human-Computer Studies*, **43**, 625--640.
- Husserl, E. (1950) *Idées Directrices pour une Phénoménologie*, Gallimard, Paris.
- Husserl, E. (1957) *Logique Formelle et Logique Transcendantale*, Presses Universitaires de France, Paris.
- McCray, A. T. (1989) The UMLS Semantic Network, In *SCAMC* Washington D.C., pp. 503-507.
- Rastier, F. (1987) *Sémantique Interprétative*, Presses Universitaires de France, Paris.
- Rastier, F., Cavazza, M. and Abeillé, A. (1994) *Sémantique pour l'analyse*, Masson, Paris.
- Sowa, J. (1984) *Conceptual Structures*, Addison Wesley.
- Zweigenbaum, P. and Consortium-Menelas (1995) Menelas: Coding and Information Retrieval from Natural Language Patient Discharge Summaries, In *Advances in Health Telematics*(Eds, Laires, M. F., Ladeira, M. J. and Christensen, J. P.) IOS Press, Amsterdam, pp. 82--89.

ENGAGEMENT SÉMANTIQUE ET ENGAGEMENT ONTOLOGIQUE : CONCEPTION ET RÉALISATION D'ONTOLOGIES EN INGÉNIERIE DES CONNAISSANCES.....	1
<i>Direction de la Recherche Prospective</i>	<i>1</i>
<i>Institut National de l'Audiovisuel.....</i>	<i>1</i>
<i>4, Avenue de l'Europe</i>	<i>1</i>
<i>94366, Bry sur Marne Cedex.....</i>	<i>1</i>
<i>Tel : 01 49 83 23 18.....</i>	<i>1</i>
<i>Fax : 01 49 83 25 82</i>	<i>1</i>
<i>Email : bbachimont@ina.fr.....</i>	<i>1</i>
RÉSUMÉ.....	1
ABSTRACT	1
1 INTRODUCTION.....	2
2 ONTOLOGIE ET REPRÉSENTATION DES CONNAISSANCES	2
3 DÉFINIR DES PRIMITIVES : L'ENGAGEMENT SÉMANTIQUE.....	3
1.1 LA NORMALISATION SÉMANTIQUE	4
1.2 LA STRUCTURATION GLOBALE : L'ARBRE ONTOLOGIQUE	6
1.3 LA STRUCTURATION LOCALE : LES PRINCIPES DIFFÉRENTIELS	7
1.4 CONCEPTS ET RELATIONS	8
1.5 ENGAGEMENT SÉMANTIQUE ET MODÉLISATION DES PRIMITIVES.....	8
1.6 ONTOLOGIE RÉGIONALE ET ENGAGEMENT SÉMANTIQUE.....	9
4 FORMALISER DES CONNAISSANCES : L'ENGAGEMENT ONTOLOGIQUE	9
5 UTILISER UNE FORMALISATION	11
6 UN EXEMPLE	11
7 EXPÉRIMENTATIONS	12
8 DISCUSSION	13
1.7 PROBLÈME DES CONCEPTS NON LINGUISTIQUES	13
1.8 PROBLÈME DE LA RÉUTILISABILITÉ DES ONTOOGIES	13
9 CONCLUSION.....	14
10 BIBLIOGRAPHIE.....	15