

A quoi servent les ontologies fondationnelles ?

Gunnar Declerck¹, Audrey Baneyx², Xavier Aimé³, Jean
Charlet^{1,4}

¹ INSERM UMRS 872 ÉQ.20, Ingénierie des connaissances en santé, Paris, France
{Gunnar.Declerck, Jean.Charlet}@crc.jussieu.fr

² SCIENCES PO, médialab, 75007, Paris, France
Audrey.Baneyx@sciences-po.fr

³ ORPHANET, Plateforme Maladies Rares, Hôpital Broussais, Paris, France
Xavier.Aime@inserm.fr

⁴ ASSISTANCE PUBLIQUE – HÔPITAUX DE PARIS

Résumé : Ce texte se propose de discuter l'idée que les ontologies fondationnelles sont utiles, voire nécessaires, au bon fonctionnement des systèmes de traitement de contenus, en particulier à leur interopérabilité sémantique. Après un rappel des principales caractéristiques des ontologies, nous proposerons une liste des grandes fonctions aujourd'hui attribuées ou au moins attendues des OF, pour en discuter ensuite le bien-fondé. Nous montrerons que la possibilité de réaliser une ontologie intégrant des primitives et définitions formelles suffisamment générales et génériques (universelles) pour décrire la sémantique des concepts de domaines spécialisés du savoir est loin d'être acquise ; et que (ii) même si une telle ontologie s'avérait réalisable, il n'est pas certain qu'elle permettrait d'assurer l'IS entre systèmes, c'est-à-dire un échange de données en préservant le sens. **Mots-clés** : Ontologies fondationnelles, Ingénierie des connaissances, Interopérabilité sémantique.

1 Introduction

De plus en plus de chercheurs dans le champ de l'Ingénierie des Connaissances véhiculent aujourd'hui l'idée qu'un recours aux ontologies fondationnelles (OF) est bénéfique, voire indispensable, au bon fonctionnement des systèmes de traitement de contenus, en premier lieu à leur interopérabilité sémantique (IS). Les OF offriraient de solutionner le fameux problème, posé à l'origine par le Web, dit de la tour de Babel (Smith, 2003), ou du cloisonnement des silos de données (Smith & Ceusters, 2010). Certains

auteurs en font ainsi un véritable cheval de bataille et mettent en place différentes actions pour en promouvoir l'usage.

B. Smith et W. Ceusters, initiateurs de la fondation *Open Biological and Biomedical Ontologies* (OBO)¹ sont sans doute actuellement les plus actifs à cet égard. Selon eux, la principale finalité des OF est d'améliorer l'intégration des données en apportant une solution au problème de la multiplicité des langages utilisés pour leur codage. Ce problème, dont on ne saurait sous-estimer la complexité, était déjà censé pouvoir être résolu par les ontologies de domaine. Mais la prolifération d'ontologies de domaine concurrentes, conçues suivant des perspectives et principes de formalisation de la sémantique hétérogènes, n'a fait que le repousser. Le problème de l'IS s'est posé non plus au niveau des données, mais des descripteurs ontologiques utilisés pour les qualifier, différents concepts, issus d'ontologies hétérogènes, pouvant être utilisés pour coder les mêmes données (Smith & Ceusters, 2010). Selon Smith et Ceusters, c'est ce problème que les OF doivent à terme permettre de résoudre, en minimisant le nombre d'ontologies de domaine utilisées pour le codage des données, tout en maximisant leur consistance mutuelle.

D'autres chercheurs, tels ceux impliqués dans le projet *Semantic Health*, parmi lesquels Alan Rector, ont toutefois un avis plus mesuré sur la question, et se permettent de nourrir des doutes quant à l'utilité réelle des OF, voire des ontologies tout court (cf. Tsujii & Ananiadou (2005) ; Stroetmann *et al.* (2009) sec. 4.2). Des tentatives sont également faites pour résoudre le problème de l'IS par le bas, en alignant directement les concepts d'ontologies (ou terminologies) de domaine concurrentes ou disciplinairement voisines, donc sans recourir aux principes de méta-description qu'offrirait une OF (cf. Hamdi *et al.* (2011); Pavel & Euzenat (2011))². Et de nombreuses initiatives sont lancées au niveau de la communauté européenne pour aligner des ontologies ou terminologies médicales utilisées en concurrence pour le codage des données³.

Que penser des espoirs investis dans les OF face à ces tentatives ? Les doutes à l'égard des OF sont-ils légitimes ? Ou bien les OF sont-elles la solution au problème de l'hétérogénéité des descripteurs sémantiques des données ? C'est à ces questions que nous proposons d'apporter des éléments de réponse dans ce texte. Après un bref rappel des principales ca-

1. <http://obofoundry.org/crit.shtml>

2. Ainsi de la banque de mappings du site Bioportal du NCBO (*National Center for Biomedical Ontology*) <http://bioportal.bioontology.org/>

3. Voir par exemple <http://www.salusproject.eu/> ou <http://www.ehr4cr.eu/>

ractéristiques des ontologies, nous proposerons pour ce faire une liste des grandes fonctions aujourd’hui attribuées ou au moins attendues des OF, pour en discuter ensuite le bien-fondé. Pour soutenir notre argumentation, nous utiliserons des travaux issus du domaine biomédical (dont nous sommes le plus familier), mais les idées présentées valent *a fortiori* quel que soit le domaine considéré.

2 Des ontologies de domaine aux ontologies fondationnelles

2.1 Les trois niveaux de généralité des ontologies

Les ontologies sont des artéfacts symboliques dont la principale fonction est de fournir une représentation de la sémantique des primitives conceptuelles d’un domaine qui soit manipulable par des machines. Elles se présentent le plus souvent sous la forme d’ensembles hiérarchisés de concepts, organisés en arbre ou en treillis, et articulés par des relations sémantiques explicites, notamment de subsomption et de méréonymie. L’organisation topologique des ontologies est essentielle car c’est à travers elle qu’est formalisée la signification : ce sont les relations que les concepts entretiennent les uns avec les autres qui en spécifient le sens.

On distingue généralement trois niveaux dans les ontologies (ou trois types d’ontologies), selon le degré d’abstraction et de généralité du domaine conceptuel couvert par leurs descriptions (Stenzhorn *et al.*, 2007). Du moins abstrait au plus abstrait, on trouve : a) les **ontologies de domaine**, b) les **ontologies noyaux** (*core-domain ontologies*), c) les **ontologies fondationnelles** (*top ontologies* ou *upper-level ontologies*). Les concepts présents dans les ontologies de domaine et les ontologies noyaux sont propres à une discipline. L’ontologie noyau a la charge des concepts les plus généraux : si la discipline est décomposable en sous-disciplines, elle couvrira les concepts qui leur sont transverses. Par exemple, une ontologie noyau de la médecine couvrira le champ de la cardiologie, de la cancérologie et de la psychiatrie. Y seront présents les concepts de maladie, patient, traitement, protéine, tissu, etc. En revanche, des concepts plus spécifiques, tels que « démence sénile » ou « trouble obsessionnel compulsif » relèveront d’une ontologie de domaine particulière : la psychiatrie.

La frontière entre ontologie de domaine et ontologie noyau est toutefois difficile à objectiver. Celle entre OF et ontologie noyau est plus facilement identifiable : l’OF ne contient que des concepts suffisamment généraux pour être utilisés dans *plusieurs* disciplines. L’OF est en ce sens transdisciplinaire ou si l’on préfère méta-disciplinaire : elle ne relève pas d’une

discipline particulière, sinon de la Métaphysique. Plus précisément : elle est censée répertorier et décrire les concepts utilisables dans *toutes* les disciplines, par exemple les concepts d'événement, de processus, de qualité, de segment temporel, de région spatiale, d'objet animé, d'objet intentionnel, etc. Elle ne contient aucun concept qui soit particulier à une discipline.

Les ontologies de domaine sont à première vue celles dont l'utilité est la plus évidente, par exemple pour le traitement de contenus textuels numériques. Une ontologie de domaine de la cardiologie permettra d'indexer un dossier patient, parce que les concepts qu'elle contient sont suffisamment spécifiques pour être présents dans les documents manipulés, et pertinents pour un professionnel du domaine. Les ontologies noyaux, en raison de leur caractère général, peuvent plus difficilement endosser un tel rôle. En revanche, elles peuvent permettre de structurer l'ontologie de domaine qu'elles subsument, de manière : a) à homogénéiser les principes de description sémantique utilisés pour définir les concepts du domaine (rôle de *standardisation*) ; b) à y naviguer plus aisément (rôle d'*orientation*).

2.2 Les fonctions attribuées aux ontologies fondationnelles

Les OF font depuis plusieurs années déjà l'objet de nombreuses études, discussions, voire querelles dans la communauté de l'ingénierie des connaissances (Guarino, 1998). L'attention qu'on leur porte vient de l'importance des problèmes qu'elles sont censées pouvoir solutionner. Au moins cinq fonctions leur sont généralement attribuées : (a) améliorer l'IS (Pease *et al.*, 2002; Masolo *et al.*, 2003; Smith & Ceusters, 2010) ; (b) fédérer (soit directement, soit indirectement, par le biais des ontologies noyaux qu'elles subsument) différentes ontologies de domaine, en les intégrant à un système de représentation sémantique unique omni-englobant (Masolo *et al.*, 2003; Stenzhorn *et al.*, 2007) ; (c) faciliter le processus de conception des ontologies de domaine (Borgo & Leitão, 2004; Pease *et al.*, 2002) ; (d) en améliorer la qualité et l'utilisabilité, voire la ré-utilisabilité, la maintenance et l'évolutivité (Keet, 2011; Smith & Ceusters, 2010) ; (e) expliciter les choix de conceptualisation qui sont faits de manière implicite dans les ontologies de domaine (*e.g.* leurs engagements ontologiques) et offrir à ce titre un cadre de comparaison (Masolo *et al.*, 2003; Guizzardi, 2006).

Les fonctions (a) et (b) sont celles qu'on trouve le plus souvent mentionnées dans la littérature. Et ce sont également celles pour lesquelles l'utilité des OF est la moins remise en cause. Ces fonctions sont également fréquemment attribuées aux ontologies noyaux. Par exemple, l'ontologie noyau Biotop, conçue à partir de l'OF BFO (*Basic Formal Onto-*

logy), est explicitement présentée par ses concepteurs comme devant servir à unifier les ontologies du domaine biomédical, de manière à assurer l'IS entre les données issues des champs disciplinaires afférents (Stenzhorn *et al.*, 2007). Sans cette action unificatrice, les ontologies de domaine restent cloisonnées. Etant conçues suivant des principes de modélisation des connaissances et de formalisation sémantique différents (sans même parler des langages de description), elles ne peuvent être utilisées en synergie au sein d'un même système. Ce cloisonnement se répercute sur les données qu'elles permettent d'indexer et de manipuler (compartmentation des données en silos), et donc, par exemple, sur la possibilité d'en extraire de manière (semi-)automatique des connaissances.

Il importe de le mentionner, les points (a) et (b) de la liste précédente réfèrent à des qualités *extrinsèques* des ontologies de domaine conçues à l'aide d'OF : l'usage d'une même OF pour guider la conception de différentes ontologies de domaine rend celles-ci sémantiquement compatibles. Il en va de même pour le point (e). Les points (c) et (d) réfèrent, en revanche, à des qualités *intrinsèques* des ontologies de domaine conçues à l'aide d'OF : elles sont mieux structurées, plus cohérentes, plus utilisables, plus faciles à mettre à jour. Dans la suite, nous traiterons uniquement des affirmations (a) et (b), en nous focalisant sur la question de l'IS. C'est en effet là le principal attendu des OF aujourd'hui.

3 Comment les ontologies fondationnelles peuvent assurer l'interopérabilité sémantique

3.1 Le problème de l'interopérabilité sémantique

Le plus large consensus au sujet de la fonction des OF concerne l'IS, soit la possibilité pour les machines d'*échanger de l'information en préservant leur signification* (Landgrebe & Smith 2011, p. 139 ; voir aussi Stroetmann *et al.* 2009, p. 17), cette signification pouvant être définie en termes purement opératoires, c'est-à-dire relativement aux opérations qui peuvent être réalisées sur la base des données « interprétées ». Cette caractérisation a l'avantage de replacer la notion d'interopérabilité dans son sens premier, qui réfère non à une capacité d'intercompréhension *stricto sensu*, mais d'inter-coordination (Bittner *et al.*, 2005). Elle permet également de faire droit au fonctionnement non sémantique des machines, en décrivant leur fonctionnement sans recourir à une métaphore psychologique. Les machines étant incapables d'accéder à l'ordre sémantique, elles manipulent les éléments d'information de manière purement mécanique, sans

en comprendre le sens : les opérations de manipulation sont gouvernées par la seule syntaxe, et c'est uniquement en représentant la sémantique dans la syntaxe (rôle des langages formels) que l'on peut donner l'impression que les manipulations mécaniques qu'effectuent les machines sont gouvernées par le sens des données (*semantic driven*).

Dans la grande majorité des articles consacrés aux OF dont nous avons pu prendre connaissance, les chercheurs sont toutefois avares en explications quant à la manière *précise* dont les OF vont permettre d'assurer ou d'améliorer l'IS entre systèmes hétérogènes. On se contente le plus souvent de tenir la chose pour acquise, ou de donner des explications par trop superficielles, en affirmant que l'usage d'une OF va permettre de subsumer différentes ontologies de domaine, et fournir ainsi une représentation formelle unifiée de la sémantique de leurs domaines respectifs.

Selon Borgo & Leitao (2004), les OF permettent d'assurer l'IS entre systèmes car elles remplissent deux fonctions : 1) réduire les risques de mésinterprétation des entités composant l'univers de discours considéré, par une description sémantique explicite des concepts dont ces entités sont l'instanciation linguistique ; 2) et donc mettre en place une transparence conceptuelle assurant la possibilité d'un partage de l'information. De même, Masolo *et al.* (2003) expliquent, dans le cadre du projet WonderWeb : « A product procurement process involving multiple agents with distributed lightweight ontologies may be carried out in an efficient way by using simple terminological services, but the risk of semantic mismatch can be minimized only if the agents rely on explicit, axiomatised ontologies, which serve to ensure mutual compatibility of the respective models in such a way as to check the extent of real agreement. » Remplir ces conditions exige notamment que les ontologies utilisées soient suffisamment exhaustives pour couvrir l'ensemble des termes du domaine (Borgo & Leitão, 2004). Les auteurs ne poussent toutefois pas plus loin leurs explications. En particulier, ils n'expliquent pas *comment* les descriptions conceptuelles explicites proposées dans les OF doivent permettre d'assurer l'IS entre systèmes, sachant notamment que les données traitées par ces systèmes ne correspondent pas directement aux concepts de l'OF - trop généraux - mais à des concepts relevant le plus souvent d'ontologies de domaine.

3.2 L'usage de primitives universelles pour décrire par combinatoire le sens des concepts de domaine

Une réponse plus précise à cette question est avancée par les auteurs (anonymes : appelons-les X) de la page du wikipedia anglophone consacrée à l'IS⁴. X commencent par marquer qu'une condition de l'IS est qu'un *standard commun* soit utilisé pour la spécification du sens des données échangées entre les machines, et que ce standard offre une spécification suffisamment détaillée de la sémantique pour neutraliser tout risque d'ambiguïté (par exemple le problème de la polysémie). Les ontologies, parce qu'elles font usage de langages formels comme les logiques de description (OWL-DL), répondent à ces critères. Maintenant, en quoi une ontologie *fondationnelle* est-elle ici requise ? Pourquoi une ontologie *de domaine* ne suffit-elle pas ? X répondent à cette question en ayant en vue ce qu'ils appellent l'interopérabilité sémantique générale (*General Semantic Interoperability*) (ISG), soit une forme d'IS ne se réduisant pas à un domaine déterminé du savoir (par exemple le champ de la médecine), mais prétendant valoir quel que soit le domaine considéré. La réponse qu'ils proposent requiert par ailleurs de prendre un nouvel élément en considération : l'évolution du savoir et du sens. Selon X, la réalisation d'une ontologie couvrant tous les termes de tous les domaines possibles est une tâche vaine, car de nouveaux termes sont régulièrement créés et d'anciens termes peuvent voir leur signification remaniée. Ces évolutions ne peuvent bien entendu pas être anticipées. En revanche, il est, selon X, possible d'en spécifier la signification si l'on dispose d'un ensemble suffisamment exhaustif de concepts primitifs, qu'il suffira alors de combiner de la manière idoine. C'est précisément ce qu'une OF bien conçue devrait permettre de faire⁵. L'idée est à première vue séduisante. Mettons à disposition un alphabet suffisamment puissant pour exprimer le sens de n'importe quel concept. Que tout le monde accepte de l'utiliser : la signification de tout concept nouvellement créé pourra être explicitée à l'aide de ce seul alphabet et l'IS

4. Semantic interoperability. In Wikipedia. Dernière modification le 20 janvier 2012. Date de consultation : 30 avril 2012. http://en.wikipedia.org/wiki/Semantic_interoperability.

5. « For the problem of General Semantic Interoperability, some form of foundation ontology ('upper ontology') is required that is sufficiently comprehensive to provide the defining concepts for more specialized ontologies in multiple domains. [...] The need for a single comprehensive all-inclusive ontology to support Semantic Interoperability can be avoided by designing the common foundation ontology as a set of basic ('primitive') concepts that can be combined to create the logical descriptions of the meanings of terms used in local domain ontologies or local databases. »

des systèmes sera assurée. Si l'on ne peut assurer la tâche d'explicitier le sens de tous les concepts en usage à un moment t , on peut en revanche assurer que tout sens nouveau soit *explicitable* dans le formalisme dont on dispose⁶.

Toutefois, on n'a pas encore résolu à ce stade (même théoriquement) le problème de l'ISG, ou de l'IS réduite à un domaine déterminé du savoir. Car si l'on fait usage d'une ontologie donnée pour faire communiquer deux systèmes (fût-elle chapeautée par une OF capable d'exprimer par combinatoire la sémantique de n'importe quel nouveau concept), et qu'une information connue d'un système S1 est transmise à un système S2 ne disposant pas d'une description explicite de sa sémantique, encore faudrait-il déterminer comment on peut accéder à cette description, même si on peut *en droit* la générer à l'aide des primitives de l'OF utilisée par S2. S1 doit nécessairement transmettre à S2, *en plus des données*, la description logique des termes inconnus de S2, description faisant usage de primitives auxquelles S2, utilisant la même OF que S1, a accès.

Pour les auteurs, une telle démarche à l'intérêt de limiter la nécessité d'établir un consensus préalable aux seuls concepts primitifs de l'OF. Ils estiment par ailleurs qu'une telle OF pourrait remplir cette fonction à l'aide de moins de 10 000 éléments, classes et relations confondues.

4 Discussion

4.1 Les ontologies fondationnelles sont-elles suffisantes pour assurer l'interopérabilité sémantique ?

Cette conception de la manière dont une OF peut contribuer à assurer l'IS entre systèmes est-elle viable ? Nous l'avons vu, un des objectifs reconnus des OF est de rendre possible l'intégration d'informations exprimées à l'aide d'ontologies de domaine différentes, en subordonnant ces ontologies à des principes de formalisation de la sémantique standardisés. Or, est-il possible d'isoler et de décrire des primitives conceptuelles capables d'exprimer par combinatoire le sens de n'importe quel concept, quel que soit le domaine considéré, ou même, plus raisonnablement, le sens de n'importe quel concept au sein d'un domaine donné ? Différents éléments portent clairement à mettre en doute une telle possibilité.

(i) Tout d'abord, la subordination de différentes ontologies de domaine

6. Notons-le, un tel projet n'est pas neuf : c'est le rêve de nombreux logiciens, de G.W. Leibniz à G. Frege.

à des principes de description sémantique identiques⁷ est rendue difficile par l'hétérogénéité des ontologies de domaine, développées très souvent de manière autonome, pour répondre à des besoins applicatifs particuliers (et ce, même si l'on se cantonne à une même discipline : disons la médecine, ou une même sous-discipline : disons la chirurgie). Les ontologies sont chaque fois des représentations partielles et partiales d'un monde complexe. Leur conceptualisation est influencée par la culture et l'expérience des concepteurs et dépend de la tâche pour laquelle elles ont été construites. Ce caractère hétérogène peut s'exercer à différents niveaux et être aussi bien d'ordre terminologique que structurel, conceptuel ou formel (Hoffmann, 2005). Les créateurs de Biotop remarquent fort justement : « Each ontology embraces only some distinct scenario with a mere partial view of the overall scientific field » (Stenzhorn *et al.*, 2007). Mais ce n'est peut-être pas pour rien, a-t-on envie d'ajouter. Peut-être cette vue partielle et partielle du domaine est-elle justement ce qui fait que ce domaine disciplinaire est ce qu'il est, et se distingue des autres. Lui imposer une stratégie de formalisation de la sémantique et un découpage du monde qui la rende compatible avec d'autres ontologies de domaine, fussent-ce des domaines disciplinaires adjacents, n'est-ce pas risquer de la rendre inutilisable, car « abstraite » ?

(ii) Ensuite, il ne semble pas exister de terrain conceptuel neutre qui puisse servir de pont entre différentes ontologies de domaine, et certainement pas qui ne soit sujet à discussion et controverse. Les OF sont censées proposer une représentation sémantique qui ne soit pas surdéterminée par un champ ou domaine d'application particulier. Mais est-ce véritablement possible ? Les concepts généraux des OF ne recouvrent-ils pas eux-mêmes des significations différentes selon le champ de connaissances qu'ils modélisent ? Que le projet de représentation des catégories les plus abstraites et universelles qui anime les concepteurs d'OF soit loin d'être trivial nous est notamment montré par les abondantes querelles qui animent la communauté de la Métaphysique contemporaine, où une multitude de camps se livre une guerre d'arguments sans merci, sans pouvoir se mettre d'accord sur les formes et propriétés de l'étant, ou même sur la légitimité de décrire l'étant en termes de propriétés (Nef, 2009). Il en va de même dans le champ pourtant plus spécifique des ontologies médicales, comme en té-

7. Relativement à cette question, il conviendrait en réalité de distinguer : 1) l'usage d'une OF pour standardiser des ontologies de domaine / noyaux déjà créées ; 2) l'usage d'une OF pour guider la conception d'une nouvelle ontologie de domaine / noyau. Le développement d'une OF dans la perspective du second usage semble beaucoup plus défendable que dans le cas du premier.

moigne le débat qui a récemment opposé Smith, Ceusters et Merrill (Smith & Ceusters, 2010; Merrill, 2010).

(iii) Plus fondamentalement encore, on peut douter du fait que la possibilité d'isoler et de décrire des primitives conceptuelles capables d'exprimer, par simple combinatoire, le sens de n'importe quel concept (d'un domaine donné ou de tous les domaines possibles, peu importe), permette de décrire ce sens avec le degré de précision exigé pour assurer l'IS entre des systèmes échangeant des données. En effet, le caractère général des primitives décrites dans les OF ne rend-il pas d'emblée vaine pareille entreprise ? Les primitives ne peuvent par principe offrir de décrire que les caractéristiques générales des concepts du domaine, donc justement pas la couche de signification qui rend ces concepts pertinents et utilisables par les acteurs. On peut bien dire qu'une « tumeur cancéreuse du poumon » est un perdurant indépendant (*independant continuant* dans BFO). Mais en aucune façon pareille définition ne saurait suffire à assurer l'IS entre deux systèmes échangeant des données médicales. Dans ce cas, en quoi les OF offrent-elles une solution au problème de l'IS, et une alternative à l'approche en termes de mappings (Smith & Ceusters, 2010) ?

(iv) Finalement, il est également important de marquer que l'IS ne peut être assurée par le modèle décrit dans la section précédente que si les concepts de l'ontologie de domaine définis par combinatoire à l'aide des primitives de l'OF sont explicitement reliés aux *termes* utilisés dans les documents pour encoder les données (principe des ressources termino-ontologiques). La possibilité d'exprimer le sens d'un nouveau concept par combinaison des primitives disponibles dans l'OF ne suffit donc pas en tant que telle à assurer l'IS, c'est-à-dire la possibilité pour un système S2 d'interpréter tel élément d'information transmis par S1 en adéquation avec le sens que possède cet élément pour S1. Encore faut-il que le concept nouvellement créé soit associé avec les *formes linguistiques* (le vocabulaire) sous lesquelles il est susceptible d'apparaître dans les données échangées. Or, on aurait tort de sous-estimer la difficulté d'une telle opération. En un sens, ce problème n'est pas celui des OF. Celles-ci prétendent rendre possible une description univoque et partagée/consensuelle de la sémantique des concepts d'un domaine, et non pas apporter une solution au problème de l'interprétation automatique des données à l'aide d'une ontologie de domaine. Toutefois, ces deux problèmes ne peuvent être totalement séparés, car les OF ne peuvent jouer leur rôle dans l'IS qu'à la condition préalable que le problème de la représentation linguistique des concepts ait pu trouver une solution viable (Tsujii & Ananiadou, 2005).

4.2 Existe-t-il des catégories universelles ?

Une des questions centrales qui ressort des réflexions précédentes est la suivante : comment objectiver dans une OF les catégories les plus abstraites, mais surtout les plus universelles, celles susceptibles de satisfaire tout le monde, de l'astrophysicien au chimiste moléculaire (pour se cantonner aux sciences de la Nature) ? La distinction de l'endurant et du perdurant est-elle universellement légitime ? Et surtout, comment déterminer si elle l'est ? Quelles modalités d'évaluation nous faut-il adopter pour déterminer si les choix de conceptualisation faits dans les OF (les catégories identifiées et les manières de les définir) sont légitimes ?

Deux grandes stratégies de conception d'OF sont aujourd'hui adoptées : l'approche *réaliste* (parfois qualifiée de révisionniste) et l'approche *conceptualiste* (ou descriptiviste ou orientée sens commun). La première prétend fixer ses choix de modélisation sur le tableau du réel construit par les sciences de la Nature (en particulier la physique et la biologie). La seconde prétend fixer les siens par une analyse des schèmes de rationalisation préscientifiques dont font usage les individus dans le langage, la perception ou le raisonnement ordinaire (idée de sens commun ou de science naïve). Si cette seconde approche s'appuie sur une discipline, c'est donc des sciences humaines qu'il s'agit, de la psychologie cognitive aux sciences sociales. L'approche réaliste est par exemple promue par les fondateurs de la fondation OBO (Smith & Ceusters, 2010). L'approche conceptualiste est adoptée par les auteurs de l'OF DOLCE, qui « vise à capturer les catégories ontologiques qui sous-tendent le langage naturel et le sens commun humain » (Masolo *et al.*, 2003), ou par les concepteurs de l'OF OntoMénélas (Zweigenbaum *et al.*, 1995) ainsi que, dans une certaine mesure, par ceux de l'OF SUMO⁸. C'est également en un sens celle de Guarino (1997) quand il met en avant un critère de sens commun et de plausibilité cognitive dans ses recommandations pour la conception d'ontologies.

Un prérequis au succès de l'approche conceptualiste est donc *qu'il existe* des catégories cognitives générales suffisamment partagées pour se retrouver d'un individu à l'autre. Cette hypothèse est-elle légitimée par les sciences humaines ? Avant toute chose, les modèles de catégorisation dont font usage nos systèmes cognitifs suivent-ils de près ou de loin la logique des OF ? Existe-t-il au niveau de notre mémoire sémantique des catégories de haut-niveau que nous utilisons pour partitionner la réalité en régions générales d'objets ?

8. <http://www.ontologyportal.org/>

L'idée que la mémoire sémantique humaine est organisée suivant les mêmes principes que les ontologies computationnelles modernes n'est pas neuve. Dès les années 60, Collins & Quillian avançaient l'idée que les significations sont stockées dans la mémoire humaine sous la forme de réseaux de concepts organisés hiérarchiquement en fonction de leur degré de généralité. Cette manière de représenter le sens aurait notamment l'avantage d'être économique du point de vue des ressources cognitives nécessaires au stockage de l'information. Le système n'a pas à spécifier, pour chaque concept particulier appartenant à une classe générale donnée, les propriétés du concept particulier qui sont également celles de la classe. La logique de l'héritage permet de ne spécifier cette information qu'une seule fois, au niveau de la classe. Par exemple, il est inutile de spécifier que le pigeon sait voler ou a des plumes, il suffit de spécifier que les oiseaux ont ces propriétés et que le pigeon est un oiseau. Depuis, de nombreux travaux ont proposé d'envisager la mémoire sémantique comme un système modulaire, organisé en plusieurs sous-systèmes sémantiques, séparés en fonction soit du mode d'acquisition des connaissances, soit de la catégorie, soit des deux à la fois. Allport (1985) a proposé un modèle reposant sur des super-catégories en adéquation avec des régions cérébrales spécialisées impliquant l'action, l'intention ou la perception. Et, selon Sieroff (2009), des aires spécifiques du cerveau seraient dédiées au traitement de domaines de connaissance importants sur le plan de l'évolution (comme les animaux, les végétaux, les parties du corps ou les outils). Plusieurs travaux ont également avancé que la distinction fondamentale entre Objet naturel / Artéfact / Objet abstrait, fréquemment représentée dans les OF, était présente au niveau de la mémoire sémantique.

Ces travaux ne suggèrent toutefois l'existence dans la mémoire sémantique humaine que de quelques grandes catégories (en aucun cas assez nombreuses pour peupler une OF utilisable), ils n'en précisent pas le sens (c'est-à-dire les conditions d'appartenance d'une instance à la classe catégorielle considérée), et le champ d'observation qu'ils couvrent ne permet en aucune façon de se prononcer sur leur possible « universalité ».

A l'inverse, des travaux dans le champ de l'anthropologie cognitive, notamment les travaux fondateurs de S.A. Tyler (1969), suggèrent plutôt une immense variabilité de catégories d'un individu à l'autre, chaque communauté possédant un système particulier de perception et d'organisation des phénomènes. J.-F. Dortier (2003) cite à ce propos le cas bien connu des couleurs. Le lexique associé aux couleurs varie considérablement d'un groupe social à l'autre : les esquimaux ont, par exemple, une plus grande

diversité de mots pour désigner les nuances de blanc que les amérindiens qui, eux, ont un lexique plus riche pour dénoter les nuances de vert.

Ce relativisme culturel se trouve certes parfois nuancé. Selon G. Jahoda (1989), si les catégories changent d'un groupe à l'autre, les processus cognitifs mis en œuvre sont communs (rassemblement par traits, notion de meilleur exemplaire). Et pour en revenir à la question des couleurs, la catégorisation des teintes de base paraît offrir une certaine stabilité, quelle que soit la culture des individus interrogés. Il n'en demeure pas moins toutefois que parler de catégories générales partagées est fort contestable. Même si certaines distinctions semblent universellement partagées (ou presque), ce qu'elles recouvrent, l'usage qui en est fait par les individus, ou le système de signification global dans lequel elles s'inscrivent semblent de leur côté éminemment variables. Ainsi, note Dortier (2003), les distinctions entre humain et non-humain existent dans toutes les cultures, mais les contours varient d'un peuple à l'autre. Si certaines catégories, comme les couleurs, semblent pouvoir entrer dans le cadre d'universaux, il semble donc hasardeux de vouloir généraliser ce constat. Dans le même ordre d'idées, l'expérience individuelle joue manifestement un rôle très important dans la détermination de la représentativité des sous-catégories vis-à-vis d'une catégorie donnée (Barsalou, 1982; Cordier, 1993).

A notre sens, la situation n'est guère différente en ce qui concerne l'approche réaliste adoptée pour la conception des OF. En effet, un prérequis au succès de l'approche réaliste est que des catégories universelles d'étant dont les définitions fassent consensus puissent être extraites du tableau que les sciences de la Nature brossent du réel. Or, cette possibilité est loin d'être assurée, comme en témoignent les nombreux débats qui portent actuellement sur la question (Smith & Ceusters, 2010; Merrill, 2010).

5 Conclusion

Nous l'avons vu, l'idée aujourd'hui répandue qu'une OF capable de « couvrir » différentes ontologies de domaine permet d'assurer (voire est l'unique manière d'assurer) l'IS entre systèmes faisant usage de ces ontologies pour échanger des données, est sujette à caution. La possibilité de réaliser une ontologie intégrant des primitives et définitions formelles suffisamment générales et génériques (universelles) pour décrire la sémantique des concepts de domaines spécialisés du savoir est loin d'être acquise. Et même si une telle ontologie s'avérait réalisable, il n'est pas certain qu'elle permettrait d'assurer l'IS entre systèmes, c'est-à-dire un échange de don-

nées en préservant le sens. Une description formelle de la sémantique des données est sans doute une manière d'assurer un consensus sur leur sens, et donc de préserver celui-ci lors des échanges. Mais, rien n'assure que pareille description puisse être réalisée autrement que de manière locale et intra-disciplinaire.

Par les éléments de réflexion proposés dans ce texte, nous ne voudrions toutefois pas discréditer les OF et leur dénier tout rôle possible dans l'IS. Notre but était uniquement ici de creuser une question trop vite résolue dans la plupart des travaux contemporains sur les OF. Dans notre laboratoire, nous avons nous-mêmes recours à une OF (le haut de l'ontologie de MÉNÉLAS⁹) pour la majorité des ontologies (médicales) de domaine que nous développons. L'usage d'OntoMénélas facilite leur conception et améliore leur qualité et leur maintenabilité (points c et d dans la section 2.2). Nous poursuivons également un travail d'analyse d'OntoMénélas et d'autres OF en testant leurs usages. Nous avons ainsi comparé la situation de réalisation d'une ontologie de domaine à l'aide d'une OF (utilisation *a priori*) à celle où l'ontologie de domaine est réalisée sans l'OF mais rattachée à cette dernière ensuite (utilisation *a posteriori*). Il est difficile à ce stade de tirer des conclusions définitives, nous constatons toutefois que l'utilisation *a priori* de l'OF est plus aisée que son utilisation *a posteriori*.

Références

- ALLPORT D. (1985). Distributed memory, modular subsystems and dysphagia. *Current Perspectives in Dysphagia*, p. 32–60.
- BARSALOU L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, **10**(1), 82–93.
- BITTNER T., DONNELLY M. & WINTER S. (2005). Ontology and semantic interoperability. In D. PROSPERI & S. ZLATANOVA, Eds., *Large-scale 3D data integration : Challenges and Opportunities*, p. 139–160. CRCpress.
- BORGIO S. & LEITÃO P. (2004). The role of foundational ontologies in manufacturing domain applications. volume 3290 of *Lecture Notes in Computer Science*, p. 670–688 : Springer.
- CORDIER F. (1993). *Les représentations cognitives privilégiées, typicalité et niveau de base*. Presses Universitaires de Lille.
- DORTIER J. (2003). *Le cerveau et la pensée*. Sciences Humaines. (ouvrage collectif sous la direction de J.F. Dortier).

9. <http://purl.oclc.org/NET/spim/ontologies/public/OntoMenas/>

- GUARINO N. (1998). *Formal Ontology in Information Systems : Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. Amsterdam, The Netherlands, The Netherlands : IOS Press, 1st edition.
- GUIZZARDI G. (2006). The role of foundational ontology for conceptual modeling and domain ontology representation. In *7th International Baltic Conference on Databases and Information Systems, Vilnius, Lithuania*.
- HAMDI F., SAFAR B. & REYNAUD C. (2011). Utiliser des résultats d'alignement pour enrichir une ontologie. In G. V. D.A. ZHIGHED, Ed., *11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'2011)*, volume 20 of *RNTI*, p. 407–412, Brest, France : Hermann.
- HOFFMANN P. (2005). *Similarité sémantique inter-ontologies basée sur le contexte*. PhD thesis, Université Claude Bernard – Lyon 1.
- JAHODA G. (1989). *Psychologie et Anthropologie*. Armand Colin.
- KEET C. M. (2011). The use of foundational ontologies in ontology development : An empirical assessment. In G. ANTONIOU, M. GROBELNIK, E. P. B. SIMPERL, B. PARSIA, D. PLEXOUSAKIS, P. D. LEENHEER & J. PAN, Eds., *European Semantic Web Conférence (1)*, volume 6643 of *Lecture Notes in Computer Science*, p. 321–335 : Springer.
- LANDGREBE J. & SMITH B. (2011). The hl7 approach to semantic interoperability. In *ICBO : International Conference on Biomedical Ontology, July 28-30, Buffalo, NY, USA*.
- MASOLO C., BORGIO S., GANGEMI A., GUARINO N. & OLTRAMARI A. (2003). *WonderWeb Deliverable D18 Ontology Library (final)*. Rapport interne, IST Project 2001-33052 WonderWeb : Ontology Infrastructure for the Semantic Web.
- MERRILL G. H. (2010). Ontological realism : Methodology or misdirection? *Applied Ontology*, **5**(2), 79–108.
- NEF F. (2009). *Traité d'ontologie pour les non-philosophes (et les philosophes)*. Gallimard, coll. folio-essais.
- PAVEL S. & EUZENAT J. (2011). Ontology matching : State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, **99**(Pre-Prints).
- PEASE A., NILES I. & LI J. (2002). The suggested upper merged ontology : A large ontology for the semantic web and its applications. In *Working Notes of the AAAI-2002, Workshop on Ontologies and the Semantic Web*.
- SIÉROFF E. (2009). *La neuropsychologie : Approche cognitive des syndromes cliniques*. Armand Colin.
- SMITH B. (2003). Ontology. In L. FLORIDI, Ed., *Blackwell Guide to the Philosophy of Computing and Information*. Oxford : Blackwell.
- SMITH B. & CEUSTERS W. (2010). Ontological realism : A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, **5**(3-4), 139–188.

- STENZHORN H., BEISSWANGER E. & SCHULZ S. (2007). Towards a top-domain ontology for linking biomedical ontologies. In K. A. KUHN, J. R. WARREN & T.-Y. LEONG, Eds., *MedInfo*, volume 129 of *Studies in Health Technology and Informatics*, p. 1225–1229 : IOS Press.
- STROETMANN V. E., KALRA D., LEWALLE P., RODRIGUES J., STROETMANN K., SURJAN G., USTUN B., VIRTANEN M. & ZANSTRA P. (2009). *Semantic Interoperability for Better Health and Safer Healthcare : Research and Development Roadmap for Europe*. Semantichealth report, European Commission (Ed.); Luxembourg : Office for Official Publications of the European Communities.
- TSUJII J. & ANANIADOU S. (2005). Thesaurus or logical ontology, which one do we need for text mining ? *Language Resources and Evaluation*, **39**(1), 77–90.
- TYLER S. (1969). *Cognitive Anthropology*. Holt and Rinehart and Winston.
- ZWEIGENBAUM P., BACHIMONT B., BOUAUD J., CHARLET J. & BOISVIEUX J.-F. (1995). Issues in the structuring and acquisition of an ontology for medical language understanding. *Methodology of Information in Medicine*, **34**(1/2).