

Validation de la sémantique d'un langage iconique médical à l'aide d'une ontologie : méthodes et applications

Jean-Baptiste Lamy, Lina F Soualmia, Alain Venot, Catherine Duclos

▶ **To cite this version:**

Jean-Baptiste Lamy, Lina F Soualmia, Alain Venot, Catherine Duclos. Validation de la sémantique d'un langage iconique médical à l'aide d'une ontologie : méthodes et applications. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.51-62. hal-01004802

HAL Id: hal-01004802
https://hal.science/hal-01004802

Submitted on 12 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Validation de la sémantique d'un langage iconique médical à l'aide d'une ontologie : méthodes et applications*

Jean-Baptiste Lamy[1], Lina F. Soualmia[1,2], Alain Venot[1], Catherine Duclos[1]

[1] LIMICS, Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny, France, INSERM UMRS 1142, UPMC Université Paris 6, Sorbonne Universités, Paris
{jean-baptiste.lamy,alain.venot}@univ-paris13.fr, catherine.duclos@avc.ap-hop-paris.fr

[2] TIBS, LITIS EA 4108, Institute of Biomedical Research, Rouen, France
lina.soualmia@chu-rouen.fr

Résumé : Les langages iconiques permettent de représenter des concepts par la combinaison de primitives graphiques (couleurs, pictogrammes...). Les exemples sont nombreux, des panneaux routiers aux icônes des interfaces informatiques. Cependant, ces langages ne définissent pas la sémantique de leurs icônes, ce qui pose plusieurs problèmes : combinaisons inconsistantes de pictogrammes, interprétations différentes d'une même icône par deux personnes différentes, difficultés à aligner les icônes avec des ressources existantes...

Le langage iconique VCM (Visualisation des Concepts en Médecine) permet de représenter par des icônes les principaux antécédents, maladies, traitements,... Dans cet article, nous proposons de valider la sémantique du langage iconique VCM à l'aide d'une ontologie. Trois applications de cette ontologie sont décrites : la vérification de la consistance des icônes constituées, l'alignement semi-automatique avec une terminologie médicale, et la génération d'un lexique des pictogrammes.

Mots-clés : Icônes, Langage iconique, Ontologies, Alignement, Médecine

1 Introduction

Il est bien connu qu'un "bon schéma vaut mieux qu'un long discours". C'est pourquoi de nombreux icônes, symboles ou pictogrammes ont été proposés (Dreyfuss, 1984) dans des domaines aussi variés que les interfaces homme-machine, la signalisation dans les lieux publics ou l'étiquetage des produits chimiques. Cependant le nombre d'icônes qu'un être humain peut mémoriser n'est pas infini et, lorsque l'on veut représenter un grand nombre de concepts, il n'est pas possible d'avoir une icône par concept. Une solution est alors de concevoir un *langage iconique*, dans lequel un grand nombre d'icônes peut être créé en combinant un nombre restreint de primitives tels que des couleurs ou des pictogrammes, en suivant une grammaire spécifique. Un exemple bien connu est la signalisation routière, où les panneaux sont composés de plusieurs éléments (cercle rouge, pictogramme,...). VCM (Visualisation des Concepts en Médecine, Lamy *et al.* 2008, 2009) est un langage iconique permettant de représenter les principaux concepts médicaux (maladies, traitements,...) par des icônes. Ce langage a pour objectif de faciliter l'accès des professionnels de santé aux documents médicaux. En effet, le volume de données et de connaissances textuelles

⋆. Ce travail a été financé par l'ANR (Agence Nationale de la Recherche) au travers des projets de recherche L3IM (ANR-08-TECS-007) et SiFaDo (ANR-11-TECS-0014).

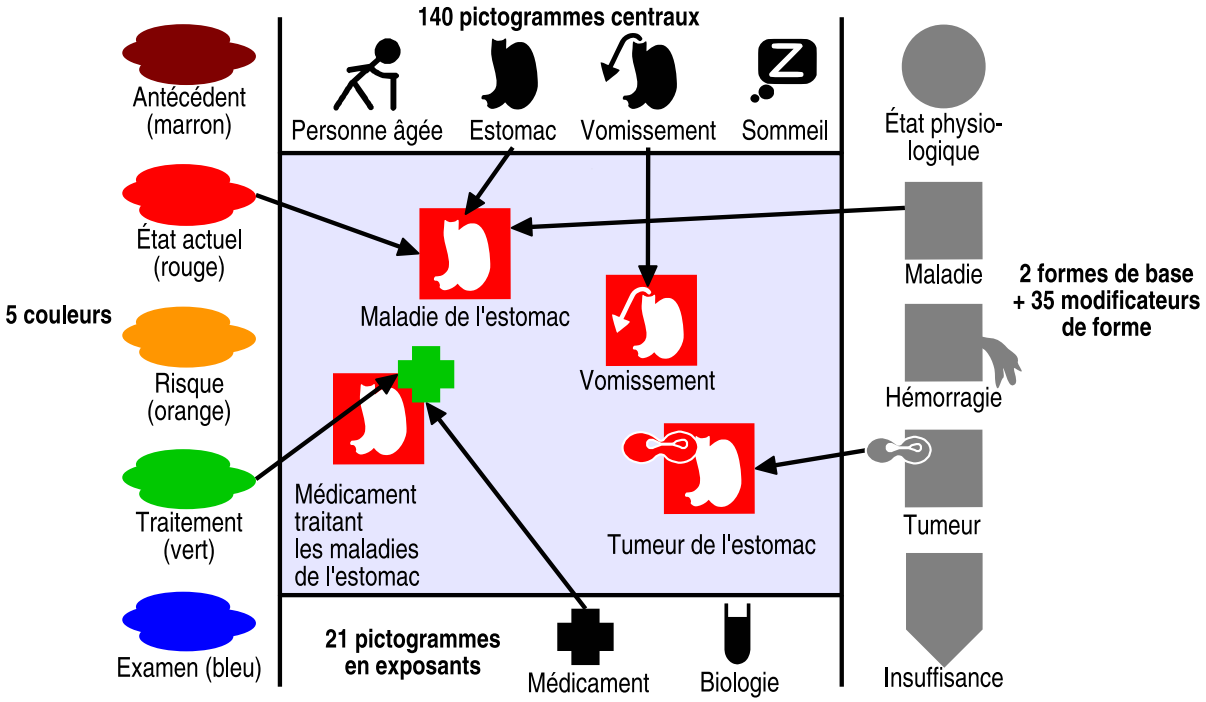


FIGURE 1 – Exemples d'icônes VCM générées par la combinaison de plusieurs éléments.

rend difficile la lecture des dossiers patients, des Résumés des Caractéristiques Produits (RCP) des médicaments ou des guides de bonnes pratiques cliniques (Ely *et al.*, 2002).

Cependant, dans un langage iconique, une icône syntaxiquement correcte peut malgré tout être sémantiquement fausse : par exemple un panneau routier avec le triangle rouge *attention danger* et le pictogramme *pneu neige*. L'absence de sémantique validée rend aussi plus difficile l'alignement des icônes avec les ressources ou terminologies existantes du domaine : les alignements doivent être réalisés manuellement par un expert qui interprète les icônes, mais comme toute interprétation humaine, celle-ci risque d'être subjective.

Dans cet article, nous proposons une approche consistant à valider la sémantique d'un langage iconique à l'aide d'une ontologie formelle. Nous présenterons le langage VCM qui servira d'exemple, puis la construction de l'ontologie des icônes VCM. Ensuite, nous décrirons trois applications de cette ontologie : la vérification de la consistance des icônes constituées, l'alignement semi-automatique avec une terminologie médicale de référence, et la génération d'un lexique des pictogrammes. Enfin, nous discuterons des avantages et des limites de cette approche fondée sur une ontologie formelle. Les ontologies, alignements et programmes réalisés au cours de ce travail ont été diffusés en logiciel libre dans PyMedTermino (https://pypi.python.org/pypi/PyMedTermino), un ensemble de modules pour accéder aux terminologies médicales en langage Python.

2 Le langage VCM

Le langage iconique VCM (Lamy *et al.*, 2008, 2009) propose des icônes pour représenter les principales conditions cliniques d'un patient, dont les symptômes, les maladies, les états physiologiques (tels que la grossesse ou les classes d'âge), les risques et les antécédents de maladie, les traitements médicamenteux ou non, les examens de biologie et les procédures de suivi. VCM comprend un ensemble de primitives graphiques (pictogrammes, formes et couleurs) pouvant être combinées selon une grammaire pour générer un grand nombre d'icônes. Ces icônes n'ont pas pour objectif d'être aussi précises que les textes médicaux, mais visent au contraire à les compléter de manière synthétique. Un didacticiel est disponible sur le site Internet dédié à VCM : http://vcm.univ-paris13.fr/.

La Figure 1 montre différents exemples de combinaisons. Une icône VCM comprend une couleur, une forme de base et un ensemble de modificateurs de forme, un pictogramme central, ainsi qu'éventuellement un exposant associant un pictogramme et une couleur, et un second exposant. Une icône simple combine (1) une couleur indiquant l'aspect temporel de l'icône : rouge pour un état actuel du patient, marron pour un antécédent, orange pour un risque futur ; (2) une forme de base : un cercle pour un état physiologique ou un carré pour un état pathologique ; et (3) un pictogramme central indiquant la localisation anatomo-fonctionnelle (par exemple *cardiaque*, *pulmonaire*,...) ou la caractéristique patient (par exemple *grossesse*) impliquée. Les structures anatomiques et les fonctions qu'elles réalisent sont représentées par le même pictogramme (par exemple *poumon* et *respiration*).

Les icônes de maladie ou de symptôme peuvent ensuite être précisées, selon deux approches différentes : (1) pour les maladies ou symptômes spécifiques à un système anatomo-fonctionnel (par exemple les *vomissements*, qui sont spécifiques à l'*estomac*), le pictogramme central est modifié ; (2) pour les maladies ou symptômes génériques pouvant être décrits comme une morphologie associée à un système anatomo-fonctionnel (par exemple les *tumeurs*, les *infections* ou les *insuffisances* d'une fonction), un modificateur de forme est ajouté à la forme de base (par exemple, les tumeurs sont représentées par deux cellules en division). Ces deux approches peuvent être combinées, et plusieurs modificateurs de forme peuvent être présents du moment qu'ils ne se recouvrent pas spatialement.

Les icônes de traitements et d'examens sont construites à partir des icônes de la maladie traitée ou du risque surveillé, en ajoutant un pictogramme en exposant, de couleur verte pour les traitements ou bleue pour les examens. Ce pictogramme en exposant indique le type de traitement (par exemple *chirurgical*) ou d'examen (par exemple *imagerie*). Un second exposant peut être ajouté pour représenter un professionnel de santé ou un document en relation avec une maladie, par exemple l'icône *cardiologue* sera construite en ajoutant l'exposant *professionnel de santé* à l'icône des *maladies cardiaques*.

3 L'ontologie des icônes VCM

L'ontologie des icônes VCM (Lamy *et al.*, 2013a) a pour objectif d'aider à la validation de la sémantique du langage VCM. Nous présenterons tout d'abord les principes généraux suivis lors de la construction de cette ontologie, puis nous la décrirons plus en détails.

3.1 Principes généraux pour la construction de l'ontologie

Le premier principe que nous avons appliqué lors de la construction de l'ontologie a été de distinguer les primitives de VCM et leur signification : en effet le *pictogramme en forme de poumon* est distinct de l'organe *poumon*. Un second principe a été d'utiliser au maximum les relations de subsomption *est-un* plutôt que d'autres relations comme les relations méréologiques *partie-de*, puisque la plupart des outils d'édition et des moteurs d'inférences s'appuient sur la subsomption. En particulier, les objets anatomiques ont été désignés comme "structure + adjectif" plutôt que par leur nom d'organe ; nous dirons par exemple "une *structure bronchique* est une *structure pulmonaire*" plutôt que "les *bronches* sont une partie des *poumons*". Les organes peuvent ensuite être ajoutés comme fils des structures anatomiques : le *poumon* est une *structure pulmonaire*. Bien que contraire au principe du "biais minimal d'encodage", on retrouve cette approche dans plusieurs terminologies médicales, dont la SNOMED CT (*Systematized Nomenclature of Medicine - Clinical Terms* Cornet 2009) et nous l'avons donc adoptée. Un troisième principe est que toute icône décrit un état du patient, y compris pour les icônes de traitement ou d'examen. Par exemple l'icône *anti-asthmatique* a été modélisée comme *patient traité par un anti-asthmatique*. Cela correspond à la manière dont VCM représente les traitements et les examens, en reprenant la pathologie qui est traitée ou le risque qui est surveillé.

L'ontologie intègre deux types de contraintes : *(i) des contraintes graphiques* portant sur les primitives de VCM : par exemple les modificateurs de forme *tumeur* et *virus* occupent le même emplacement sur les icônes, et ne peuvent donc pas être conjointement associés ; et *(ii) des contraintes médicales* portant sur les structures anatomiques, les morphologies,... représentées par les primitives : par exemple une *tumeur* est une morphologie qui ne peut s'appliquer qu'à une structure anatomique, mais pas à une fonction biologique.

Lors de la construction de l'ontologie, un jeu d'une centaine d'icônes tests a été mis au point de manière itérative, et un raisonneur a été utilisé pour tester la consistance de l'ontologie et vérifier les résultats obtenus sur les icônes tests.

3.2 Structure de l'ontologie des icônes VCM

L'ontologie des icônes VCM (Figure 2, voir Lamy *et al.* 2013a pour une version plus détaillée) a été découpée en trois modules. Le premier (240 concepts, 21 relations et 2 597 axiomes) décrit les primitives et les icônes VCM. Il a été généré à partir de la liste des primitives du langage VCM. Le fichier OWL obtenu a ensuite été importé puis édité manuellement dans Protégé pour y ajouter les règles de composition des icônes et les contraintes graphiques. Ces règles restreignent les composantes d'une icône (par exemple, au plus un pictogramme central) et prennent en compte les contraintes spatiales.

Le second module (369 concepts, 18 relations et 828 axiomes) décrit les concepts médicaux représentés par les primitives du langage VCM : structures anatomiques, fonctions biologiques, morphologies processus pathologiques, caractéristiques patient (telles que les classes d'âge) et types de traitements et de surveillances. L'ontologie contient les concepts médicaux de base et les règles de combinaison, mais pas l'ensemble des maladies, traitements,... qui peuvent être générées par combinaison (il s'agit donc de *post-coordination*). Ce second module a été modélisé avec Protégé, en s'appuyant sur les terminologies médicales (SNOMED CT en particulier) et le réseau sémantique de l'UMLS (Kashyap &

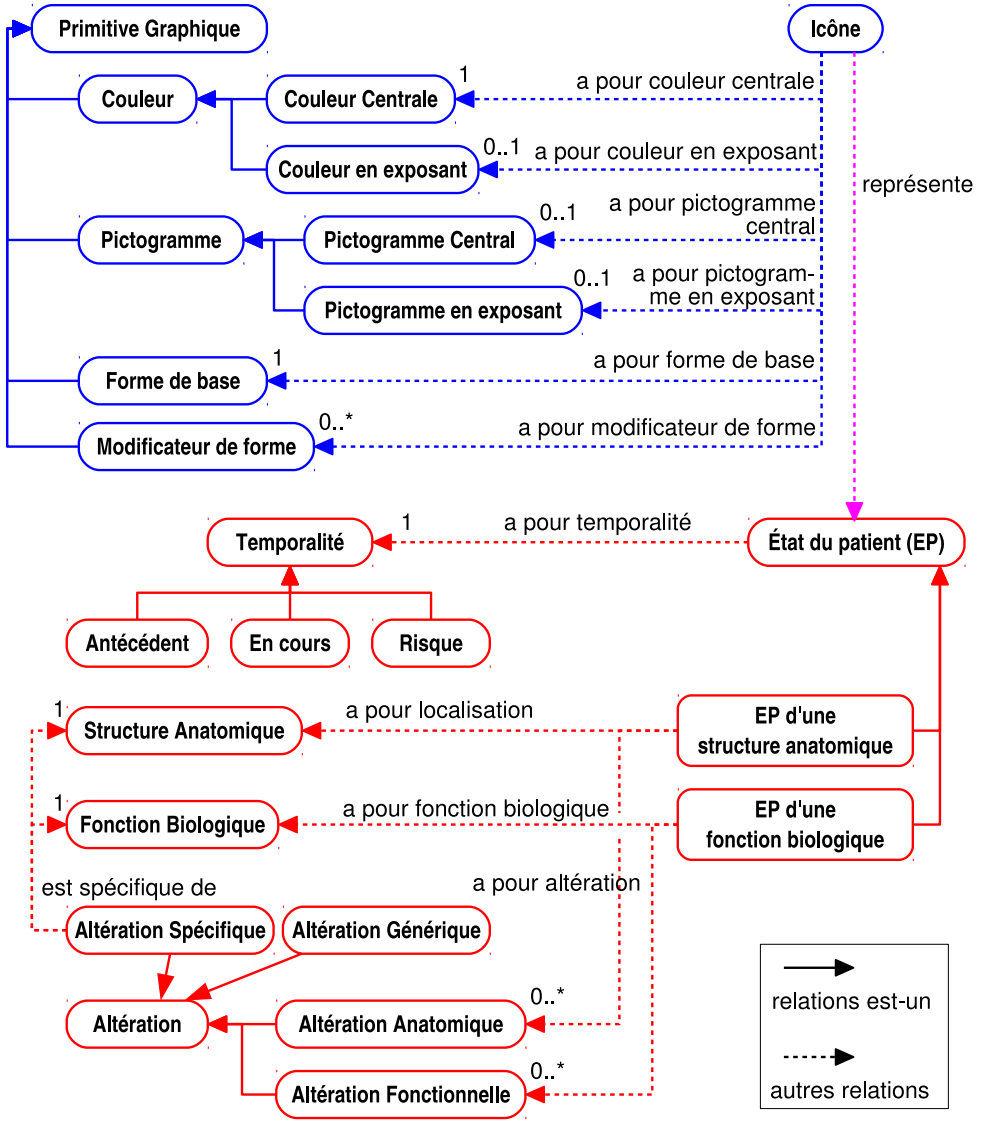


FIGURE 2 – Principaux concepts et relations dans l'ontologie des icônes VCM. Les couleurs correspondent aux modules de l'ontologie.

Borgida, 2003), en se limitant au niveau de granularité élevé qui est celui de VCM.

Les concepts de ces deux modules sont reliés par des relations *représente* (509 axiomes). Par exemple le pictogramme central *poumon* est seulement présent sur des icônes qui représentent des concepts médicaux liés à une *structure pulmonaire* ou à la *fonction respiratoire*. Ces relations ont été générées automatiquement à partir d'un fichier texte faisant correspondre à chaque primitive le ou les concepts médicaux correspondants.

L'ontologie comprend en tout 609 concepts, 41 relations et 3 934 axiomes, définis à l'aide du langage OWL-DL. Elle appartient à la famille *ALCRIQ* des logiques de description (Attribute Language, Complex concept negation, Role, Inverse property, Qualified cardinality restriction), qui est décidable (Horrocks & Sattler, 2003).

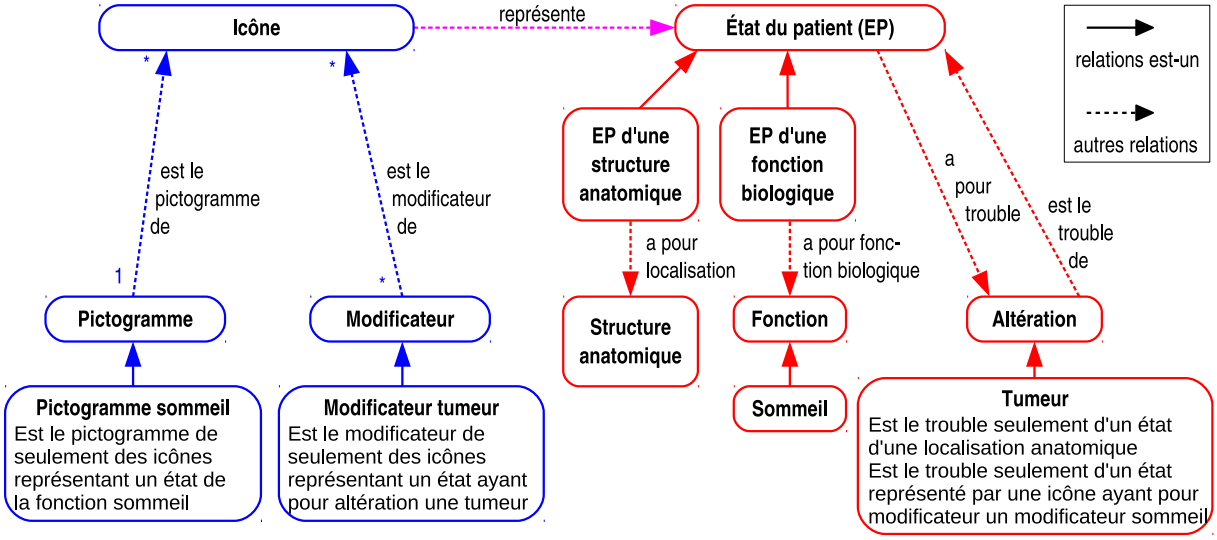


FIGURE 3 – Représentation de l'icône inconsistante *tumeur du sommeil* dans l'ontologie.

4 Vérification de la consistance des icônes

La première application de l'ontologie des icônes VCM a été la vérification de la consistance des icônes (Lamy *et al.*, 2013a). En effet, certaines combinaisons du langage VCM sont inconsistantes : par exemple une icône associant le modificateur de forme *tumeur* et le pictogramme central *sommeil* signifie "tumeur du sommeil", ce qui n'a aucun sens du point de vue médical. Cependant, ces icônes inconsistantes peuvent poser problème, notamment lorsque des utilisateurs sont amenés à construire eux-mêmes une icône en sélectionnant plusieurs primitives. Nous allons donc voir comment les contraintes exprimées dans l'ontologie permettent de vérifier la consistance des icônes.

La Figure 3 montre un exemple de représentation dans l'ontologie d'une icône inconsistante. L'inconsistance peut être déduite à partir des contraintes modélisées dans l'ontologie : (a) l'icône a le pictogramme *sommeil*, et donc elle représente un état de la fonction *sommeil*, (b) l'icône a le modificateur de forme *tumeur*, et donc elle représente un état qui a pour altération la morphologie *tumeur*, (c) la morphologie *tumeur* est une altération anatomique, qui ne peut s'appliquer qu'à des états d'une structure anatomique, (d) les structures anatomiques et les fonctions biologiques sont disjointes.

Ce raisonnement a pu être reproduit avec le raisonneur HermiT (Motik *et al.*, 2009), soit pour une seule icône, soit pour une classe d'icônes, en créant dans l'ontologie le concept des icônes partageant telle ou telle composante. La détermination des inconsistances a ensuite été évaluée (Lamy *et al.*, 2013a).

5 Alignement de VCM avec les terminologies médicales : la SNOMED CT

L'utilisation de VCM dans les applications médicales nécessite son alignement avec des référentiels existants, afin notamment de pouvoir associer automatiquement des icônes aux ressources déjà indexées par ces référentiels, comme par exemple les Dossiers Pa-

tient Informatisés. Dans cette section, nous étudierons la SNOMED CT, une terminologie couvrant les différents concepts médicaux, dont l'anatomie, les conditions cliniques, les procédures,... Elle inclut de nombreuses relations entre ces concepts : des relations de subsomption *est-un* mais aussi des relations entre conditions cliniques et structures anatomiques ou morphologies associées (par exemple l'hépatite *est localisée dans* le foie et *a pour morphologie* l'inflammation).

La méthode que nous proposons pour l'alignement s'appuie sur la nature compositionnelle similaire de la SNOMED CT et de l'ontologie des icônes VCM. L'alignement s'est fait en deux étapes : (1) alignement manuel des concepts médicaux de l'ontologie des icônes VCM (n=369) aux termes SNOMED CT correspondants (structures anatomiques, étiologies, etc), et (2) alignement automatique des termes SNOMED CT de conditions cliniques aux icônes VCM, en décomposant les termes SNOMED CT et en traduisant chacun des termes obtenus avec l'alignement manuel précédent. Cette méthode a initialement été appliquée à un sous-ensemble de la SNOMED CT, la CORE problem list (Lamy *et al.*, 2013b) ; nous présenterons ici les résultats sur l'ensemble de la SNOMED CT.

5.1 Alignement manuel entre concepts médicaux de l'ontologie VCM et la SNOMED CT

Les 369 concepts de l'ontologie des icônes VCM ont été manuellement alignés sur la SNOMED CT. La SNOMED CT utilisant l'héritage multiple, certaines structures anatomiques sont classées dans plusieurs branches. Cela conduit à associer plusieurs pictogrammes à ces structures, et donc plusieurs icônes aux maladies correspondantes. Par exemple les *osselets de l'oreille* sont à la fois une *structure osseuse* et une *structure auditive*, et donc une *maladie des osselets* aura les deux icônes *maladie de l'os* et *maladie de l'oreille*. Cependant, si cela est juste d'un point de vue ontologique, cela ne correspond pas à ce qu'attend un clinicien : en effet, les maladies sont classées par spécialité notamment durant les études médicales, chaque maladie étant attribuée de manière consensuelle à une et une seule spécialité (par exemple les maladies des osselets relèvent de l'ORL, oto-rhino-laryngologie, ce qui fait qu'un médecin voyant l'icône *maladie de l'os* ne pensera pas à une maladie des osselets). C'est pourquoi nous avons choisi d'associer chaque structure anatomique à au plus un pictogramme (l'oreille dans notre exemple).

Pour cela, une première version de l'alignement manuel a été réalisée, puis nous avons recherché tous les termes SNOMED CT de structures anatomiques qui conduisaient *via* l'ontologie à plusieurs pictogrammes (n=181). Chacun de ces termes a ensuite été associé à un et un seul concept de l'ontologie, ce qui a conduit à la création de nouveaux concepts médicaux dans l'ontologie (n=97, inférieur à 181 car certains concepts proches ont été regroupés). Ces concepts ont ensuite été associés à un seul pictogramme, choisi selon (1) la spécialité médicale associée à la structure anatomique et aux maladies correspondantes, et (2) la position de ces maladies dans les terminologies médicales monoaxiales et notamment la CIM10 (Classification Internationale des Maladies, version 10).

Au total, l'alignement manuel a fait intervenir 1 753 termes SNOMED CT.

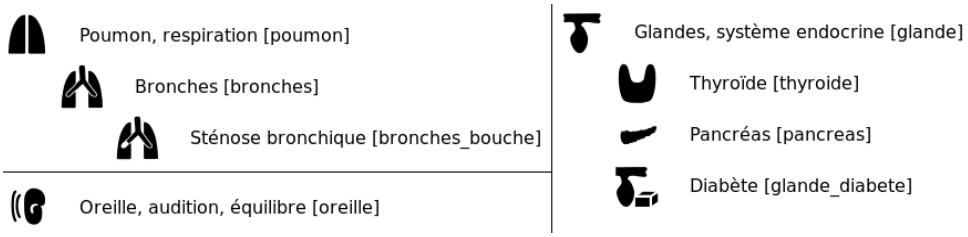


FIGURE 4 – Trois extraits du lexique des primitives VCM rédigé manuellement. Les identifiants des primitives figurent entre crochets.

5.2 Alignement automatique entre SNOMED CT et VCM

Les termes de conditions cliniques de la SNOMED CT ont ensuite été alignés automatiquement aux icônes VCM, en procédant par décomposition. Chaque terme est décomposé à l'aide des relations de la SNOMED CT, puis les termes SNOMED CT obtenus sont traduits en concept de l'ontologie VCM à l'aide de l'alignement précédent, et en prenant en compte les relations *est-un* et *partie-de* existant dans la SNOMED CT. Ces concepts sont ensuite traduits en primitives VCM à l'aide des relations de l'ontologie VCM, puis les primitives obtenues sont assemblées pour former une icône. Par exemple, le terme SNOMED CT *Uveitis* est décomposé en une structure anatomique, *Uveal tract*, et une morphologie, *Inflammation*. *Uveal tract* est une partie de *Entire eye* qui est une *Structure of visual system*, laquelle est traduite en *Structure visuelle* dans l'ontologie VCM. *Structure visuelle* est alors traduit en *pictogramme œil* dans les primitives VCM. *Inflammation* conduit au modificateur de forme *carré avec une flamme*, et ce modificateur est combiné au pictogramme *œil* pour donner l'icône adéquate.

L'alignement obtenu porte sur l'ensemble des 99 626 conditions cliniques (*clinical findings*) de la SNOMED CT, et a fait appel à 1 957 icônes VCM. 77 754 (78,0%) ont été alignés avec une seule icône, 7 573 (7,6%) ont été alignés avec 2 icônes et 517 (0,5%) avec 3 icônes ou plus ; 13 782 (13,8%) termes n'ont pas pu être alignés avec VCM (la méthode aboutissant à aucune icône ou à une icône "vide" sans pictogramme ni modificateur de forme). L'analyse manuelle des termes non-alignés montre qu'il s'agit principalement de termes qui ne sont pas des conditions cliniques au sens de VCM (par exemple *Drug therapy finding*), de termes servant à qualifier des conditions cliniques (*Clinical stage finding*), de termes très généraux (*Alive*) ou de symptômes non-associés à une localisation précise (*Erythema*, à distinguer de *Erythema of skin*).

6 Génération du lexique des primitives du langage VCM

La documentation du langage VCM comprend un lexique des primitives du langage (pictogrammes, modificateurs de forme et couleurs). Ce lexique sert à l'apprentissage de VCM, mais aussi de référence pour les experts. Le lexique se présente sous la forme d'une liste arborescente, associant sur chaque ligne une primitive et le ou les libellés associés, ainsi que l'identifiant de la primitive.

La version originelle du lexique (figure 4) a été rédigée manuellement. Cette rédaction manuelle pose plusieurs problèmes : le lexique doit être remis à jour systématiquement à

chaque modification du langage iconique, et il peut subsister des ambiguïtés (par exemple deux pictogrammes pouvant correspondre au même organe) et des zones d'ombre (par exemple un organe pour lequel aucun pictogramme n'est indiqué).

Dans cette section, nous proposons une méthode pour générer automatiquement le lexique à partir de l'ontologie des icônes VCM.

6.1 Méthode de génération du lexique

La construction du lexique s'est faite en quatre étapes : (1) l'extraction de l'ensemble des primitives VCM de l'ontologie , (2) pour chaque primitive, récupération *via* les relations de l'ontologie de la liste des concepts médicaux que la primitive peut représenter, (3) ordonnancement de ces concepts et (4) obtention du ou des libellés associés à chaque concept, et génération du lexique.

Lors de l'étape 2, une primitive correspond souvent à plusieurs concepts. En effet, VCM utilise fréquemment le même pictogramme pour représenter un organe et sa fonction. La liste de concepts inclut aussi tous les concepts descendants (fils, petit-fils,... par exemple *structure de la plèvre* est un descendant de *structure pulmonaire*), à l'exception de ceux qui sont reliés à une primitive plus spécifique (par exemple *structure bronchique* qui est relié au pictogramme *bronche*).

Lors de l'étape 3, lorsque plusieurs concepts sont associés à une même primitive, se pose alors la question de l'ordre dans lequel ils apparaîtront dans le lexique. Pour cela, nous avons mis au point des règles combinant la nature des concepts (anatomique ou fonctionnelle), le niveau d'échelle (macroscopique ou microscopique), la spécificité (générale ou spécifique) et la lisibilité des libellées correspondants (les noms d'organe sont généralement plus lisibles que les noms de structure correspondants, par exemple *poumon* vs. *structure pulmonaire*). Les règles appliquées pour ordonner les concepts sont les suivantes (par ordre décroissant de priorité) :

1. Lorsqu'un organe est présent parmi les concepts, celui-ci vient en premier.
2. Les concepts appartenant à plusieurs hiérarchies anatomiques (par exemple les *osselets de l'oreille interne* qui sont à la fois dans la hiérarchie des *structures auditives* et dans la hiérarchie des *structures osseuses*) sont placés en dernier, et écrits en gris.
3. Les concepts de structures anatomiques sont placés avant les concepts de fonctions biologiques.
4. Les concepts de structures anatomiques sont ordonnés entre eux en allant du macroscopique au microscopique. Pour cela, nous avons utilisé les quatre niveaux d'échelle (région anatomique, tissus, cellule, liquide) que distingue l'ontologie.
5. Les concepts généraux sont placés avant les concepts plus spécifiques (en s'appuyant sur les relations *est-un*, par exemple *diabète* est placé avant *diabète de type 2*).

Lors de l'étape 4, nous avons récupéré les libellés associés aux concepts. Pour chaque concept, l'ontologie contient un libellé principal, et éventuellement un ou plusieurs synonymes ou hyponymes. Le libellé principal a été placé en premier, suivi de la mention "inclut" et des autres libellés. L'ordre dans lequel les primitives apparaissent dans le lexique a été déterminé à la main, en reprenant l'ordre du lexique manuel, qui suivait une logique anatomique (par exemple en regroupant le système cardiovasculaire, pulmonaire,...).

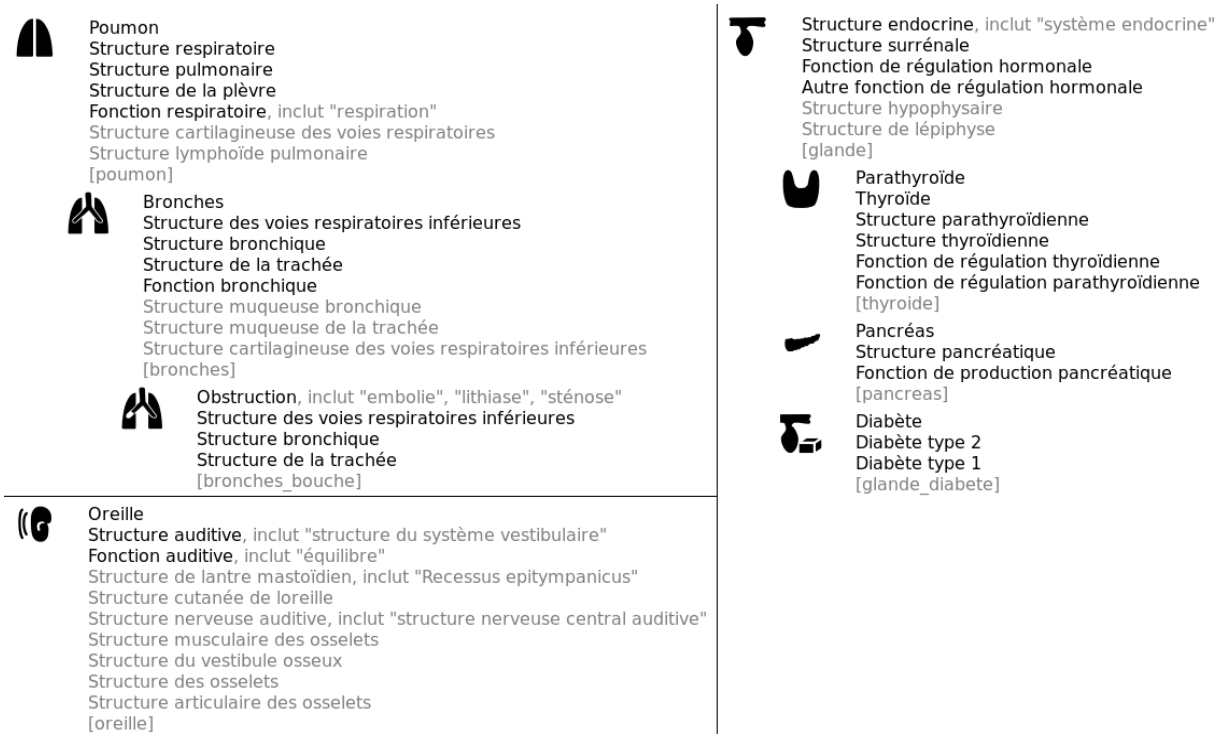


FIGURE 5 – Trois extraits du lexique des primitives VCM produit à partir de l'ontologie.

6.2 Le lexique produit à partir de l'ontologie

La Figure 5 montre des extraits du lexique produit à partir de l'ontologie. En comparaison avec l'ancien lexique (Figure 4), le nouveau lexique est plus riche.

L'alignement des concepts médicaux de l'ontologie des icônes VCM avec la SNOMED CT garantit la couverture de l'ensemble du domaine, notamment en ce qui concerne l'anatomie. Le lexique généré ne contient donc pas de zone d'ombre : toutes les structures anatomiques sont présentes dans le lexique, soit directement, soit via une structure parente plus générale (par exemple *structure de la plèvre* a été rattachée au pictogramme poumon et apparaît sur le lexique, alors qu'elle était absente de l'ancien lexique). De plus, lors de l'alignement avec la SNOMED CT, nous avons repéré les 181 concepts médicaux ambigus qui étaient associés à plusieurs primitives. Tous ces concepts ambigus ont été ajoutés dans l'ontologie et apparaissent donc dans le lexique, clairement associés à un seul pictogramme (par exemple *structure des osselets*, qui apparaît en face du pictogramme *oreille*).

7 Discussion et conclusion

Nous avons vu comment valider la sémantique d'un langage iconique en représentant les concepts relatifs aux icônes à l'aide d'une ontologie formelle, puis nous avons proposé trois applications d'une telle ontologie : la vérification de la consistance des icônes, l'alignement des icônes avec des terminologies existantes et la génération d'un lexique des pictogrammes. Les langages iconiques sont traditionnellement définis par la décomposition

graphique des composantes des icônes (Meunier, 1998), plus récemment X. Ma a proposé de structurer selon le modèle Hyppertopic, issu du Web socio-sémantique, des icônes destinées au taguage de documents (Ma & Cahier, 2012). La validation ontologique que nous proposons ici va plus loin en séparant les éléments graphiques (le signifiant) de ce qu'il représente (le signifié), et a permis d'automatiser les différentes applications. Les trois applications présentées auraient pu être accomplies avec des représentations des connaissances spécifiques à chacune d'elles (par exemple des règles de grammaire pour vérifier la syntaxe des icônes), cependant l'ontologie a permis de toutes les réaliser.

Ces méthodes ont été appliquées avec succès à VCM, un langage iconique représentant les principaux concepts médicaux. Vu la généricité de l'approche proposée, elles pourraient cependant être appliquées à d'autres langages iconiques, tels que les panneaux routiers.

La première application de l'ontologie a porté sur la vérification de la consistance des icônes (Lamy *et al.*, 2013a). Ensuite, nous avons aligné les icônes VCM avec une terminologie de référence en médecine, la SNOMED CT (Lamy *et al.*, 2013b). Cet alignement a pu être réalisé de manière semi-automatique grâce aux relations présentes dans l'ontologie et dans la SNOMED CT, cependant la méthode utilisée sera difficilement reproductible avec des ressources terminologiques moins structurées ou formalisées, telle que la CIM10. Enfin, nous avons produit un lexique des primitives graphiques à partir de l'ontologie d'un langage iconique. Le lexique ainsi construit est plus riche qu'un lexique rédigé à la main, et l'alignement de l'ontologie avec les terminologies du domaine médical permet de s'assurer de l'absence de zones d'ombre ou d'ambiguïté dans le lexique. Ce lexique peut être mis à jour automatiquement en cas de modification de l'ontologie.

Les principales difficultés rencontrées lors de la génération du lexique étaient liées à l'ordre des éléments présents dans le lexique. En effet, l'ontologie ne définit pas d'ordre entre les concepts partageant une relation (par exemple si l'ontologie définit la *bouche*, l'*œsophage* et l'*estomac* comme étant des *structures digestives*, elle ne donne pas l'ordre dans lequel présenter ces trois organes). Nous avons proposé des règles pour ordonner les différents concepts listés en face d'une entrée du lexique, en prenant en compte la nature, le niveau d'échelle, la spécificité, et la lisibilité des libellés des concepts. En revanche, l'ordre dans lequel figure les primitives dans le lexique a été déterminé manuellement. Cet ordre est important car un utilisateur s'attend à trouver ensemble les primitives d'un même système (système digestif par exemple). De plus, pour certains systèmes comme le système digestif, il existe un ordre logique dans lequel présenter les primitives, en suivant le trajet du bol alimentaire (bouche, œsophage, estomac, intestin grêle, côlon, anus). Des ontologies plus complètes sur l'anatomie, telles que la FMA (Rosse & Mejino JL, 2003), contiennent des relations de connexion entre organes (par exemple la bouche est connectée à l'œsophage), à partir desquelles nous pourrions inférer l'ordre des organes du tube digestif. Cependant, même ainsi, il n'est pas possible de déterminer le sens de présentation des organes (de la bouche à l'anus, ou de l'anus à la bouche ?).

Dans la littérature, la plupart des travaux concernant les ontologies et les lexiques visent à construire une ontologie en s'appuyant sur un lexique existant, ce qui est le contraire de ce que nous proposons ici. Cependant, la production de lexique textuel à partir d'ontologie a déjà été envisagée (Hirst, 2009), notamment dans des domaines techniques bien définis. Le problème que nous avons rencontré pour ordonner les éléments du lexique et les libellés dans chaque entrée est aussi rencontré par les outils générant des descriptions

en langage naturel à partir d'une ontologie, tel que ELEON/NATURALOWL (Konstantopoulos *et al.*, 2011). Ces outils génèrent une définition textuelle d'un concept d'une ontologie, à partir des relations qui ont été définies. L'ordre dans lequel les relations sont prises en compte et apparaissent dans le texte doit être configuré par l'utilisateur.

De futurs travaux porteront sur l'alignement de VCM avec des terminologies moins structurées que la SNOMED CT, telles que la CIM10, en s'appuyant sur l'alignement déjà construit, ainsi que l'alignement de l'ontologie sur une ontologie de fondement et la complétion de VCM en fonction des manques identifiés lors des alignements.

Références

Cornet R. (2009). Definitions and qualifiers in SNOMED CT. *Methods Inf Med*, **48**(2), 178–183.

Dreyfuss H. (1984). *Symbol sourcebook : An Authoritative Guide to International Graphic Symbols*. John Wiley and sons.

Ely J. W., Osheroff J. A., Ebell M. H., Chambliss M. L., Vinson D. C., Stevermer J. J. & Pifer E. A. (2002). Obstacles to answering doctors' questions about patient care with evidence : qualitative study. *BMJ*, **324**(7339), 710.

Hirst G. (2009). *Handbook on ontologies*, chapter Ontology and the Lexicon.

Horrocks I. & Sattler U. (2003). Decidability of SHIQ with complex role inclusion axioms. *Artificial Intelligence*, **160**, 2004.

Kashyap V. & Borgida A. (2003). Representing the UMLS Semantic Network in OWL. In *Proceedings of ISWC 2003 (International Semantic Web Conference)*, volume 1-16.

Konstantopoulos S., Karkaletsis V., Vogiatzis D. & Bilidas D. (2011). *Language Technology for Cultural Heritage*, volume 115-132, chapter Authoring semantic and linguistic knowledge for the dynamic generation of personalized descriptions.

Lamy J. B., Duclos C., Bar-Hen A., Ouvrard P. & Venot A. (2008). An iconic language for the graphical representation of medical concepts. *BMC Medical Informatics and Decision Making*, **8**, 16.

Lamy J. B., Duclos C. & Venot A. (2009). De l'analyse d'un corpus de texte à la conception d'une interface graphique facilitant l'accès aux connaissances sur le médicament. In *Actes des 20es Journées Francophones d'Ingénierie des Connaissances*, volume 265-276, Hammamet, Tunisie : PUG.

Lamy J. B., Soualmia L. F., Kerdelhué G., Venot A. & Duclos C. (2013a). Validating the semantics of a medical iconic language using ontological reasoning. *J Biomed Inform*, **46**(1), 56–67.

Lamy J. B., Tsopra R., Venot A. & Duclos C. (2013b). A Semi-automatic Semantic Method for Mapping SNOMED CT Concepts to VCM Icons. *Stud Health Technol Inform*, **192**, 42–6.

Ma X. & Cahier J. P. (2012). Vers un système de taguage iconique basé sur Hypertopic. In *Actes des 23e journées francophones d'ingénierie des connaissances (IC)*.

Meunier J. G. (1998). The categorial structure of iconic languages. *Theory&Psychology*, **8**(6), 805–825.

Motik B., Shearer R. & Horrocks I. (2009). Hypertableau reasoning for description logics. *Journal of Artificial Intelligence Research*, **36**, 165–228.

Rosse C. & Mejino JL V. (2003). A reference ontology for biomedical informatics : the Foundational Model of Anatomy. *J Biomed Inform*, **36**, 478–500.