
Formalisation de la sémantique des langages iconiques

Méthode à base d'ontologie et applications

Jean-Baptiste Lamy¹, **Lina F. Soualmia**^{1,2}, **Catherine Duclos**¹,
Alain Venot¹

1. LIMICS, Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny, France,
INSERM UMRS 1142, UPMC Université Paris 6, Sorbonne Universités, Paris,
France

jean-baptiste.lamy@univ-paris13.fr, catherine.duclos@avc.ap-hop-paris.fr, alain.venot@univ-paris13.fr

2. Normandie Université, LITIS EA 4108, NormaSTIC CNRS FRE 3638, Université
de 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 utilisateur. Cependant, ces langages n'associent pas de sémantique logique à leurs icônes, ce qui peut poser divers problèmes : des combinaisons inconsistantes de primitives graphiques, des interprétations différentes d'une même icône par deux personnes, des difficultés à mettre en correspondance les icônes avec des concepts de ressources termino-ontologiques existantes...*

Dans cet article, nous proposons une méthode de formalisation de la sémantique d'un langage iconique à l'aide d'une ontologie. Cette méthode a initialement été développée pour le langage iconique VCM (Visualisation des Concepts en Médecine), qui permet de représenter par des icônes les principaux concepts médicaux (antécédents, maladies, traitements...). Nous montrons que cette méthode est généralisable à d'autres langages iconiques en l'appliquant à la signalisation routière. Nous décrivons quatre applications de la formalisation du langage : la vérification de la consistance des icônes constituées, l'alignement semi-automatique des icônes avec une terminologie médicale, la génération d'un lexique des pictogrammes et la génération de libellés pour les icônes.

ABSTRACT. *Iconic languages can represent concepts by the combination of graphical primitives (such as colors or pictograms). There are numerous examples, from traffic signs to computer user interface icons. However, these languages do not associate semantics to their icons, which*

raises various problems: inconsistent combinations of graphical primitives, different interpretations of a given icon by two persons, difficulties to map the icons with the concepts of existing termino-ontological resources...

In this article, we propose a method for formalizing the semantics of an iconic language with an ontology. This method was initially developed for the VCM iconic language (Visualization of Concepts in Medicine), which enables to represent the main medical concepts (antecedents, disorders, treatments...) by icons. We show that this method is generalizable to other iconic languages by applying it to traffic signs. Four applications are described : the verification of icons consistency, the semi-automatic alignment with a medical terminology, the generation of a pictogram lexicon and the generation of icon labels.

MOTS-CLÉS : icônes, langage iconique, sémantique, ontologies, alignement, médecine.

KEYWORDS: icons, iconic language, semantics, ontologies, alignment, medicine.

DOI:10.3166/RIA.30.579-606 © 2016 Lavoisier

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 sont utilisés (Dreyfuss H, 1984) dans des domaines aussi variés que les interfaces homme-machine, la signalisation dans les lieux publics ou l'étiquetage des produits chimiques. Cependant, un être humain ne peut pas mémoriser une infinité d'icônes et, lorsque l'on veut représenter un grand nombre de concepts, il n'est pas possible d'apprendre une icône pour chaque concept. Une solution consiste à concevoir un *langage iconique* doté d'une syntaxe et d'une sémantique, dans lequel il est possible de générer un grand nombre d'icônes par la combinaison d'un nombre restreint de primitives telles que des couleurs ou des pictogrammes.

Un exemple quotidien est la signalisation routière dans laquelle des panneaux sont utilisés pour signifier des ordres (par exemple une limitation de vitesse), des dangers (attention aux virages) ou des informations (directions, commodités) aux conducteurs. Elle recourt à un langage iconique assez simple : les panneaux sont construits par composition de plusieurs éléments (cercle rouge, pictogramme...). Ce langage répond à deux objectifs : les panneaux doivent être reconnus rapidement et doivent être indépendants de la langue (pour les touristes étrangers). Il fait l'objet d'un apprentissage lors du Code de la route, au cours duquel les symboles arbitraires du langage doivent être mémorisés. Un autre exemple est le langage iconique VCM (Visualisation des Concepts en Médecine, (Lamy, Duclos *et al.*, 2008 ; Lamy *et al.*, 2009)) qui permet 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 tex-

tuelles 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, la grammaire et la sémantique des langages iconiques restent souvent informelles, ou au mieux décrites de manière textuelle. Cela conduit à un certain flou. Ce flou complique l'apprentissage du langage d'une part, et d'autre part il oblige à réaliser manuellement toutes les opérations et les travaux relatifs à la sémantique du langage, tels que l'écriture d'un lexique des pictogrammes ou de libellés textuels pour les icônes. De plus, dans un langage iconique, une icône syntaxiquement correcte peut malgré tout être sémantiquement absurde : par exemple un panneau routier avec le triangle rouge « attention danger » et le pictogramme « pneu neige » est absurde. L'absence de sémantique validée logiquement rend aussi plus difficile la mise en correspondance des icônes avec les concepts de ressources termino-ontologiques 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. La formalisation des langages graphiques présente quatre intérêts (Erwig M, 1997) : (a) améliorer la compréhension et l'apprentissage du langage, (b) faciliter son implémentation, (c) effectuer des raisonnements et prouver des choses sur le langage, et (d) intégrer le langage dans d'autres environnements.

Nous proposons ici une méthode consistant à formaliser la sémantique d'un langage iconique à l'aide d'une ontologie formelle, modularisée en deux parties : *l'ontologie des icônes* et *l'ontologie du domaine*. L'objectif de cet article est de généraliser les méthodes qui ont été initialement mises au point sur le langage iconique VCM en montrant qu'elles sont applicables à d'autres langages iconiques. Cet article est organisé comme suit. Après un état de l'art relatif à la formalisation des langages graphiques en section 2, nous présentons en section 3 la méthode que nous proposons. Nous l'appliquons ensuite à la signalisation routière (section 4) et à VCM (section 5). Nous présentons différentes applications rendues possibles par l'expression de la sémantique du langage iconique : la vérification de la consistance des icônes constituées, l'alignement semi-automatique avec des concepts d'une terminologie médicale de référence, la génération d'un lexique des pictogrammes, et la génération de libellés multilingues pour les icônes. Enfin, nous discutons des avantages et des limites de cette approche fondée sur une ontologie formelle (section 6).

2. État de l'art sur la formalisation des langages graphiques

Historiquement, les langages graphiques ont d'abord fait l'objet d'études sémiologiques (Meunier JG, 1988 ; 1998) qui consistaient généralement à décomposer les signes graphiques tels que les panneaux routiers (voir figure 1), afin d'associer à chaque élément (couleur, pictogramme...) une partie du sens du signe. Bien que faciles à comprendre pour un humain, ces travaux ne permettaient pas une véritable formalisation des langages graphiques.

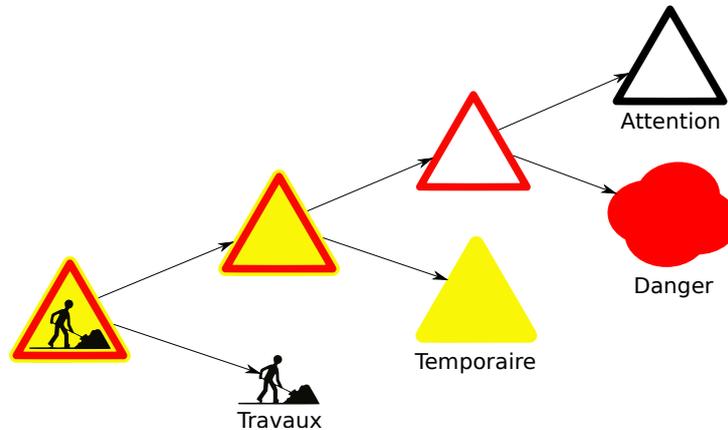


Figure 1. Structure sémiotique du panneau routier « attention travaux », d'après Meunier JG

Ensuite, les travaux se sont orientés vers la conception de grammaires pour les langages visuels. Plusieurs approches ont été proposées. On peut citer les grammaires de multienssembles à contraintes (*constraint multiset grammars*) (Marriott K, 1994) qui s'inspirent des grammaires de Chomsky pour définir la grammaire des langages visuels. Les grammaires positionnelles (*positional grammars*) (Costagliola *et al.*, 1997) représentent une autre approche qui s'appuie sur une grammaire textuelle classique complétée par un évaluateur de position (PE) chargé de traduire les phrases de manière graphique. Ces approches grammaticales ont notamment été appliquées aux langages visuels schématiques (*diagrammatic visual languages*) tels que les langages graphiques de programmation ou de requêtage dans lesquels un programme ou une requête est exprimé sous la forme d'un schéma. Ces grammaires ont permis la réalisation d'analyseurs (*parsers*) et de compilateurs pour ces langages.

Une autre approche consiste à transformer les « textes » d'un langage graphique en graphes orientés étiquetés, sans nécessairement passer par une grammaire, grâce à l'utilisation d'une syntaxe abstraite visuelle (*abstract visual syntax*) (Erwig M, 1997; Ellner, Taha, 2007). Le graphe constitué permet ensuite d'exprimer de la sémantique et d'effectuer des démonstrations. Des syntaxes abstraites visuelles ont notamment été définies pour des langages graphiques dérivant du lambda calcul.

Enfin, deux approches ont cherché à décrire les langages graphiques de manière lexicale et syntaxique. L'utilisation de logiques de description (DL) est une première approche de ce type (Haarslev V, 1998) (voir exemple en figure 2). Plus récemment, une approche basée sur le contexte local a été proposée (Costagliola *et al.*, 2015) pour les langages graphiques schématiques. Elle permet une définition lexicale et syntaxique, en définissant les éléments du langage (boîtes, flèches...) et la manière de les relier entre eux (cardinalités, *etc.*). Ces deux approches permettent de décrire les langages graphiques mais restent limitées dans l'expression de la sémantique.



Figure 2. Utilisation de logique de description pour définir la syntaxe et la sémantique d'une relation d'héritage en UML comme étant « une ligne qui touche deux rectangles et qui comporte un triangle à l'une de ses extrémités », d'après Haarslev V

En conclusion de cet état de l'art, diverses approches ont été proposées pour formaliser les langages graphiques. Un certain nombre d'entre elles sont de nature grammaticale et syntaxique, y compris lorsqu'il s'agit de représenter la sémantique du langage (nous reviendrons sur ce point en discussion). Néanmoins, la plupart des approches se sont intéressées aux langages schématiques tels qu'UML (*Unified Modeling Language*) plutôt qu'aux icônes et aux langages iconiques, ce qui est d'autant plus surprenant que les icônes et les pictogrammes sont omniprésents aujourd'hui dans notre vie quotidienne et les interfaces utilisateur. En ce qui concerne les icônes, différentes taxonomies ont été proposées, par exemple en fonction des stratégies de représentation iconique (Nakamura, Zeng-Treitler, 2012) ou selon l'objectif dans lequel les icônes sont employées (Ma *et al.*, 2015), sans aller jusqu'à la formalisation de la sémantique des icônes.

3. Méthode de formalisation de la sémantique d'un langage iconique

La méthode que nous proposons ici s'inspire de plusieurs approches présentées dans l'état de l'art (section 2) : (a) des études sémiotiques de Meunier JG (1988 ; 1998) pour la décomposition des icônes en éléments (pictogrammes, formes géométriques, couleurs), (b) des syntaxes abstraites visuelles (*abstract visual syntax*) (Erwig M, 1997 ; Ellner, Taha, 2007) pour l'emploi d'une syntaxe non graphique plus abstraite et formelle que les langages visuels et (c) des travaux de Haarslev V (1998) pour l'emploi d'une ontologie formelle en logiques de description (LD). Cependant, contrairement à Haarslev, nous avons cherché à représenter sous forme d'ontologie la syntaxe du langage graphique mais aussi sa sémantique.

Un second point important est que nous avons voulu séparer les aspects syntaxiques et les aspects sémantiques, afin d'éviter toute confusion entre les objets graphiques (les icônes et leurs composants: pictogrammes, couleurs, *etc.*, par exemple le pictogramme « ♥ ») et les objets du domaine qu'ils représentent (par exemple l'organe cœur). Pour cela, nous avons fait le choix de représenter dans deux ontologies séparées la syntaxe du langage (*l'ontologie des icônes*) et le sens associé à chaque objet graphique (*l'ontologie du domaine*), puis de les relier entre elles.

La méthode que nous proposons pour formaliser la sémantique d'un langage iconique comporte donc les étapes suivantes : (1) construire l'*ontologie des icônes* décrivant la syntaxe des icônes du langage (similaire à V. Haarslev), (2) construire l'*ontologie du domaine* (ou réutiliser une ontologie existante) décrivant les objets représentés par le langage iconique, (3) établir des liens entre les concepts des deux ontologies à l'aide des relations « icône - représente - objet du domaine » et « objet du domaine - est représenté par - icône ». Il est possible d'utiliser des relations filles de « représente / est représenté par » pour distinguer les différents types de représentation possibles. La plupart des auteurs en distinguent trois types : (a) les « icônes vraies » qui ressemblent à l'objet représenté (analogie, par exemple le pictogramme «  » ressemble à la forme en haricot caractéristique d'un rein), (b) les indices qui sont des choses associées à l'objet (par exemple une disquette pour l'action d'enregistrer) et (c) les symboles qui sont des associations arbitraires (convention, par exemple « H » pour « hôpital »).

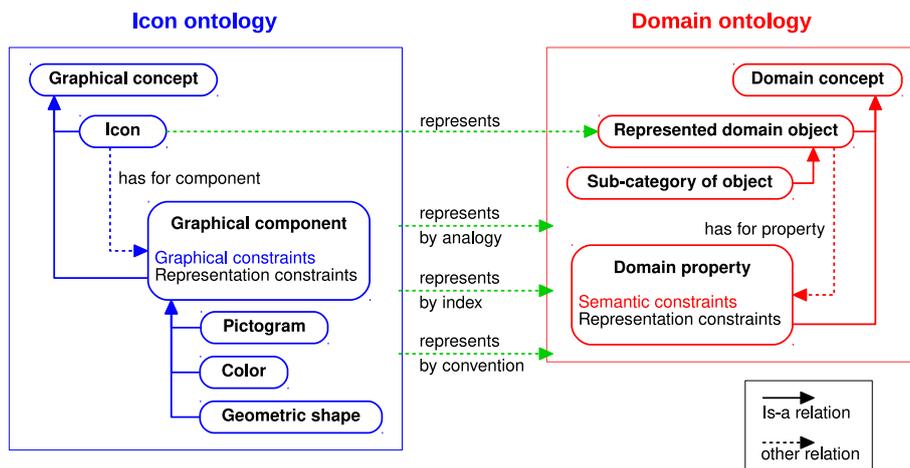


Figure 3. Structure générale d'une ontologie des icônes (en bleu) et d'une ontologie du domaine (en rouge). Les relations « représente » (en vert) relient des concepts de l'ontologie des icônes à ceux de l'ontologie du domaine

Différentes contraintes peuvent être formulées à plusieurs niveaux : (1) *des contraintes de nature graphique* dans l'ontologie des icônes (par exemple « un panneau routier contient au plus un pictogramme »), (2) *des contraintes sémantiques* dans l'ontologie du domaine (par exemple « les travaux peuvent être associés à une consigne de signalisation routière annonçant un danger, mais pas à des consignes d'ordre ou d'interdiction », c'est-à-dire donnant l'ordre de faire des travaux ou interdisant d'en faire, car les automobilistes ne font pas de travaux) et (3) *des contraintes de représentations* (par exemple « le pictogramme 'travaux' est présent uniquement sur des panneaux représentant des consignes associées aux travaux » et « les travaux sont uniquement associés à des consignes représentées par des panneaux incluant le pictogramme 'travaux' »). Les relations et contraintes de représentation permettent ainsi de

« propager » les contraintes de chaque ontologie vers l'autre : en effet, les contraintes graphiques peuvent avoir un impact sur la sémantique du langage et limiter celle-ci, de même que les contraintes sémantiques peuvent avoir des implications graphiques et rendre certaines icônes inconsistantes. La figure 3 montre la structure générale de l'ontologie des icônes et de l'ontologie du domaine.

La formalisation que nous proposons ici permet non seulement de décrire la syntaxe et la sémantique des langages iconiques, mais rend également possibles plusieurs applications : déterminer automatiquement les icônes inconsistantes, générer des correspondances avec des concepts issus d'autres ressources termino-ontologiques, automatiser la production du lexique du langage iconique et des libellés des icônes. Dans la section suivante, nous appliquons cette formalisation à un exemple simple : les panneaux de la signalisation routière. Dans la section d'après, nous l'appliquons au langage VCM et nous montrons des exemples des différentes applications possibles.

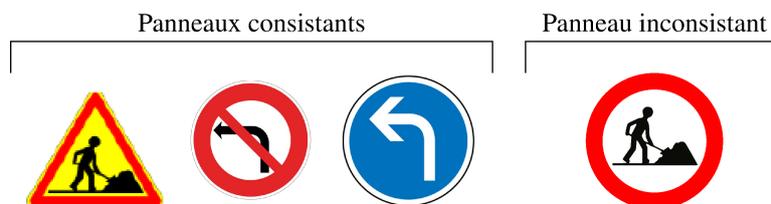


Figure 4. Exemples de panneaux qu'il est possible de former en utilisant la signalisation routière comme un langage iconique. De gauche à droite : attention travaux, interdiction de tourner à gauche, virage à gauche obligatoire, interdiction de travaux. Ce dernier panneau est un exemple de panneau inconsistant (et inexistant), car les automobilistes ne font pas de travaux

4. Application à la signalisation routière

4.1. Les panneaux de signalisation routière

Les panneaux de signalisation routière reposent sur un langage iconique simple et bien connu qui permet de donner des consignes visuelles aux automobilistes. Les icônes de ce langage sont les panneaux routiers. Quatre principaux types de panneaux peuvent être distingués : des panneaux d'interdiction (avec un rond rouge, parfois barré), des panneaux d'obligation (rond bleu), des panneaux informant d'un danger à venir (triangle rouge) et des panneaux apportant d'autres types d'informations (carré). Les panneaux temporaires ont un fond jaune et les permanents un fond blanc. Enfin un pictogramme placé au centre vient préciser l'objet du panneau.

Une grammaire combinatoire permet de combiner ces trois éléments entre eux (forme, fond et pictogramme, voir exemple figure 4). Cependant, certaines combinaisons n'ont pas de sens et des règles sémantiques sont nécessaires en plus des règles purement grammaticales pour empêcher la construction de panneaux inconsistants. Par

exemple, le pictogramme « virage à gauche » peut figurer à la fois sur des panneaux d’ordre (virage à gauche obligatoire), d’interdiction (interdit de tourner à gauche) ou d’information de danger (attention virage à gauche dangereux). Au contraire, le pictogramme « travaux » peut figurer sur des panneaux d’information de danger (attention travaux), mais pas sur des panneaux d’ordre ou d’interdiction (voir quatrième panneau figure 4).

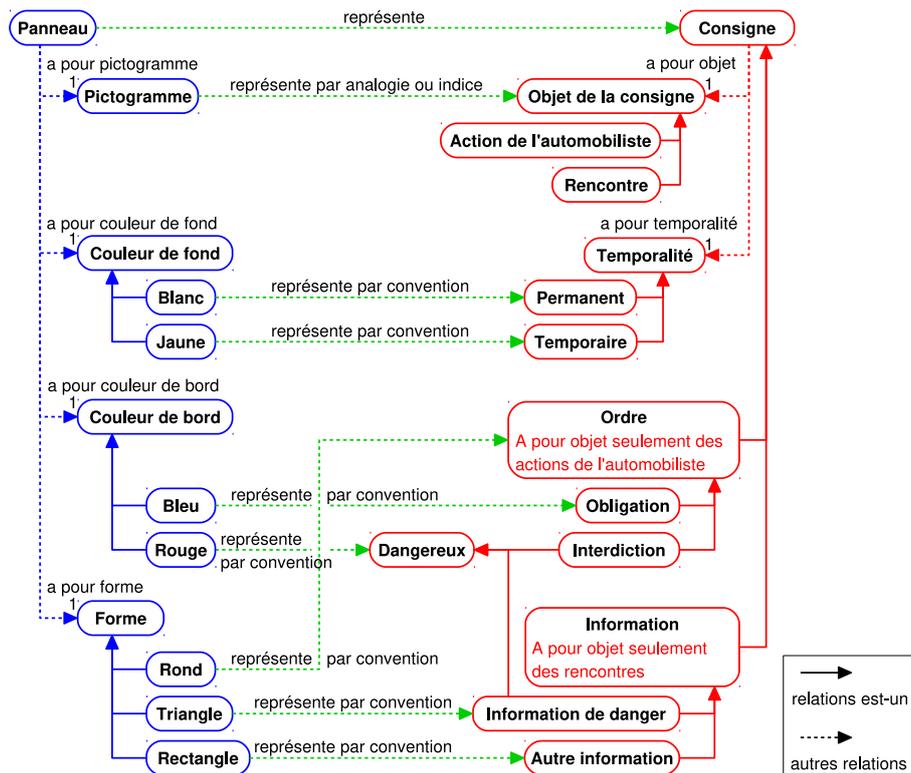


Figure 5. Ontologie des panneaux routiers (en bleu) et des consignes associées (en rouge)

4.2. Formalisation de la sémantique des panneaux

Les panneaux routiers peuvent facilement être décrits à l’aide d’une ontologie (ontologie des icônes) ainsi que les consignes associées (ontologie du domaine), ce qui conduit à l’ontologie de la figure 5. Dans l’ontologie du domaine, nous distinguons deux grandes catégories de consignes : les ordres et les informations. Les ordres se subdivisent en obligations et interdictions, et les informations en informations de danger et autres informations. Chaque consigne est associée à une temporalité (permanente ou temporaire) et un objet, qui peut être (a) une *action de l’automobiliste* (par exemple tourner à gauche, rouler à 50 km/h, dépasser un autre véhicule, conduire un

tracteur, etc.) ou (b) une *rencontre* portant sur un élément extérieur à l'automobiliste (par exemple un virage à gauche, un hôpital, des animaux sauvages, etc.). De manière évidente, les ordres ne peuvent porter que sur des actions de l'automobiliste (on ne peut pas ordonner à l'automobiliste de rencontrer quelque chose) et les informations ne peuvent porter que sur des rencontres (la signalisation routière n'est pas là pour informer l'automobiliste de ses propres actions). Nous traduisons cela par deux contraintes sémantiques dans l'ontologie du domaine.

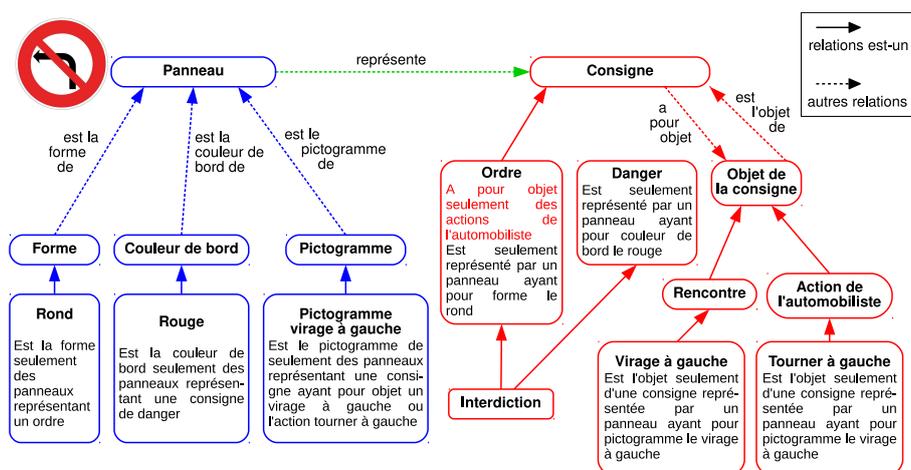


Figure 6. Représentation du panneau « interdiction de tourner à gauche » dans l'ontologie et la sémantique associée

4.3. Exemples et applications pratiques

L'ontologie présentée permet de formaliser la sémantique des panneaux routiers. En associant les différents éléments constitutifs des panneaux routiers, il est possible de construire les panneaux existants ou bien d'en inventer de nouveaux. Une première application consiste à déterminer la consistance et la sémantique de ces panneaux.

La figure 6 montre la description du panneau « interdiction de tourner à gauche ». On notera qu'un même élément graphique (ici le pictogramme « virage à gauche ») peut représenter plusieurs choses dans l'ontologie du domaine (ici la rencontre d'un virage à gauche et l'action de tourner à gauche). Ce type d'ambiguïté est fréquent dans les langages iconiques, car les icônes sont souvent moins précises que le texte (ou les ontologies).

Lorsque l'on essaie de décrire un panneau inconsistant comme le panneau « interdiction de travaux » (voir figure 7), les contraintes présentes dans l'ontologie permettent de déduire l'inconsistance. La déduction peut être automatisée à l'aide d'un raisonneur, comme nous le détaillons avec VCM (section 5).

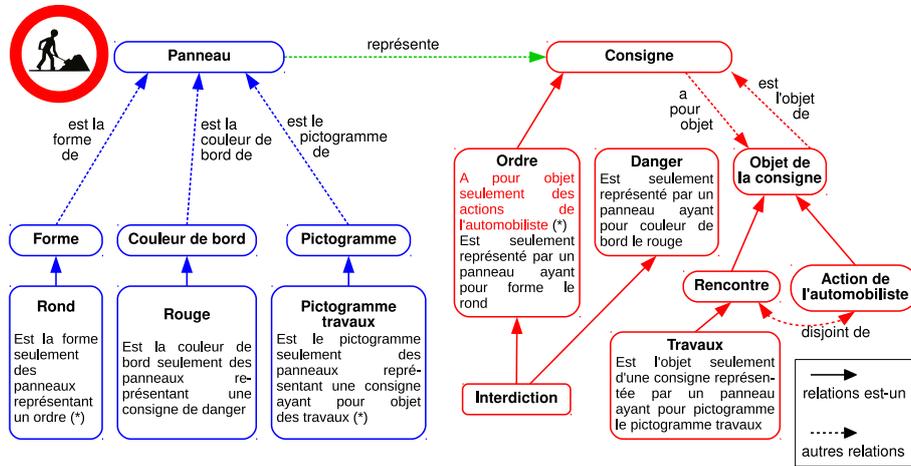


Figure 7. Représentation du panneau inconsistant « interdiction de travaux » dans l’ontologie et la sémantique associée. Les trois contraintes marquées d’une étoile (*) sont celles qui permettent d’inférer l’inconsistance sémantique de ce panneau. Deux contraintes de représentation permettent de déduire que ce panneau représente un ordre ayant pour objet les travaux, or une contrainte sémantique indique que les ordres ne peuvent avoir pour objet que des actions de l’automobiliste et non des rencontres, alors que les travaux sont une rencontre

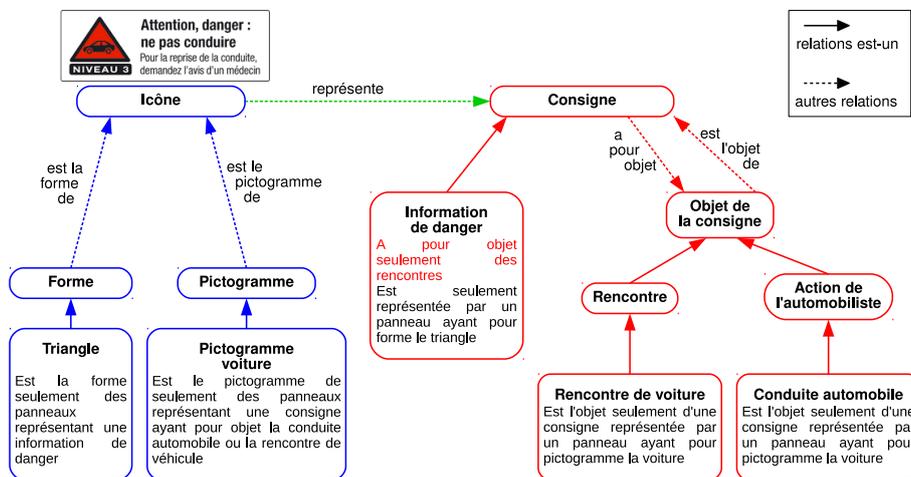


Figure 8. Icône « attention danger : ne pas conduire » qui figure sur certains médicaments et sa représentation dans l’ontologie des panneaux routiers. Cette représentation conduit à un sens différent pour cette icône : « attention : risque de rencontrer des voitures », car la forme triangle est normalement associée à une information de danger et à une rencontre, mais pas à une interdiction

Les panneaux routiers ont souvent été utilisés comme source d'inspiration pour la construction d'icônes hors du domaine de la signalisation routière. Un exemple est l'icône « attention danger : ne pas conduire » qui figure sur certains médicaments (figure 8). Une seconde application de notre ontologie consiste à vérifier que ces langages iconiques dérivés respectent bien la sémantique originelle des panneaux routiers.

Lorsque l'on décrit l'icône « attention danger : ne pas conduire » dans l'ontologie des panneaux routiers, le sens obtenu n'est pas celui attendu. En effet, dans la signalisation routière, la forme « triangle » doit être associée à une information de danger et à une rencontre. Le sens devient donc « attention : risque de rencontrer des voitures » !

Une troisième application consiste à améliorer la cohérence des panneaux routiers. Par exemple certains panneaux d'interdiction sont barrés et d'autres ne le sont pas. De même, certains panneaux partagent un même élément sémantique qui est représenté visuellement de manière similaire mais pas identique ; par exemple la flèche à gauche est différente sur les panneaux « interdiction de tourner à gauche » et « virage à gauche obligatoire » (voir deuxième et troisième panneaux en figure 4).

5. Application au langage VCM

5.1. Le langage VCM

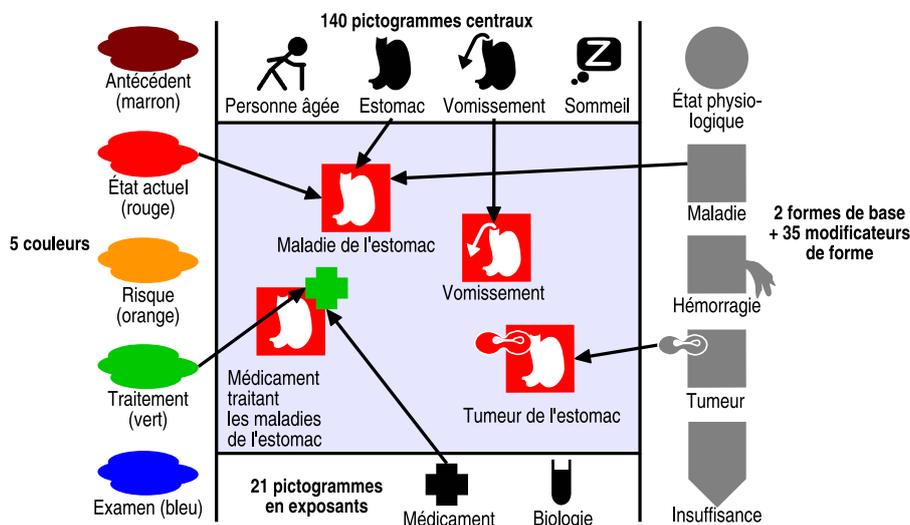


Figure 9. Création d'icônes VCM par la combinaison de plusieurs éléments

Le langage iconique VCM (Lamy, Duclos *et al.*, 2008 ; Lamy *et al.*, 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

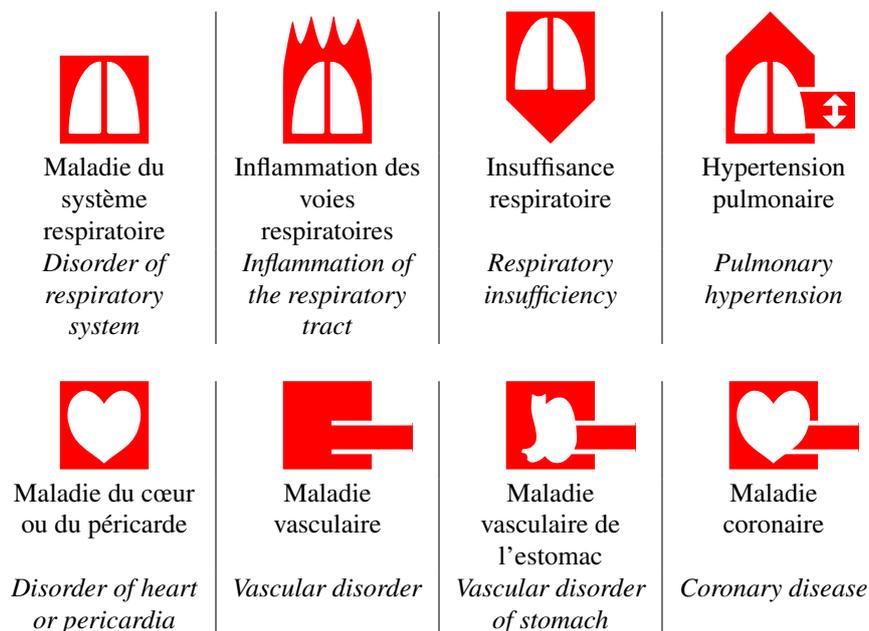


Figure 10. Exemples d'icônes VCM avec leur libellé en français et en anglais

d'âge), les risques et les antécédents de maladie, les traitements médicamenteux ou non, les examens de biologie et les procédures de suivi. Le langage 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 9 montre l'assemblage des icônes VCM et la figure 10 des exemples d'icônes. Une icône VCM se compose d'une couleur, d'une forme de base et d'un ensemble de modificateurs de forme, d'un pictogramme central, ainsi qu'éventuellement d'un exposant associant un pictogramme et une couleur, et d'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; (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 anatomofonctionnelle (par exemple cardiaque, pulmonaire...) ou la caractéristique patient (par exemple grossesse) impliquée. Les structures anatomiques et les fonctions qu'elles effectuent 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 « maladies cardiaques ».

Les icônes VCM ont initialement été développées pour présenter les connaissances sur les médicaments (contre-indications, interactions médicamenteuses, précautions d'emploi, effets indésirables, *etc.*). Dans cette application, nous avons comparé une interface interactive utilisant VCM par rapport à une interface « classique » à base de texte. Les résultats montrent que l'interface iconique permet un accès deux fois plus rapide à l'information, tout en conduisant à une diminution significative des erreurs (Lamy, Venot *et al.*, 2008).

Par la suite, le langage iconique a été étendu et d'autres applications ont été développées : (a) l'association d'icônes dans les dossiers patients, pour en faciliter la lecture (Simon *et al.*, 2014), (b) l'aide à la consultation des guides de bonnes pratiques critiques, qui contiennent souvent de nombreux cas particuliers que les icônes permettent d'identifier plus facilement (Pereira *et al.*, 2014), et (c) l'utilisation dans les moteurs de recherche médicaux, pour faciliter la lecture des listes de résultats parfois longues et fastidieuses (Griffon *et al.*, 2014).

5.2. L'ontologie VCM

L'ontologie VCM (Lamy, Soualmia *et al.*, 2013) a pour objectif d'aider à la validation de la sémantique du langage VCM. Nous présentons tout d'abord les principes généraux suivis lors de la construction de cette ontologie, puis nous la décrivons plus en détail. L'ontologie, ainsi que les alignements et les programmes que nous présentons ensuite, sont diffusés en logiciel libre et intégrés dans PyMedTermino¹ (Lamy *et al.*, 2015), un ensemble de modules permettant d'accéder aux terminologies médicales en langage Python.

1. <https://pypi.python.org/pypi/PyMedTermino>

5.2.1. 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 d'un côté les icônes et les primitives graphiques de VCM (*ontologie des icônes*) et de l'autre les concepts médicaux correspondants (*ontologie du domaine*), comme indiqué dans la méthode décrite en Section 3. En effet, le pictogramme « poumon » est distinct de l'organe « poumon ».

Un second principe a été d'employer au maximum les relations de subsomption « est-un » plutôt que d'autres types de 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 stomacale* est une *structure digestive* » plutôt que « l'*estomac* est une partie du *tube digestif* ». Les organes peuvent ensuite être ajoutés comme fils (relation *est-un*) 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 R, 2009)) et nous l'avons adoptée.

Un troisième principe a été que toute icône décrit un état du patient, y compris les icônes de traitement ou d'examen. Par exemple l'icône « antiasthmatique » a été modélisée comme « patient traité par un antiasthmatique ». 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.

Un quatrième principe a été d'utiliser uniquement des classes et non des individus. Dans l'ontologie du domaine, seules des connaissances générales sont représentées, et non des informations spécifiques à un patient : par exemple l'organe « poumon » est une classe que l'on pourrait instancier pour obtenir le poumon d'un patient donné (cependant cette instanciation n'est pas nécessaire pour l'expression de la sémantique du langage). Dans un article récent (Lamy JB, 2016), nous avons développé plus en détail la difficulté à employer des instances dans le domaine médical.

Nous retrouvons les trois types de contraintes dans l'ontologie : (1) *des contraintes graphiques* dans l'ontologie des icônes, qui portent 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 ; (2) *des contraintes médicales* dans l'ontologie du domaine, qui portent 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 ; (3) *des contraintes de représentation* qui relient les éléments de l'ontologie des icônes à ceux de l'ontologie du domaine (par exemple pour relier le pictogramme « poumon » à l'organe « poumon » et à la fonction « respiration »).

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.

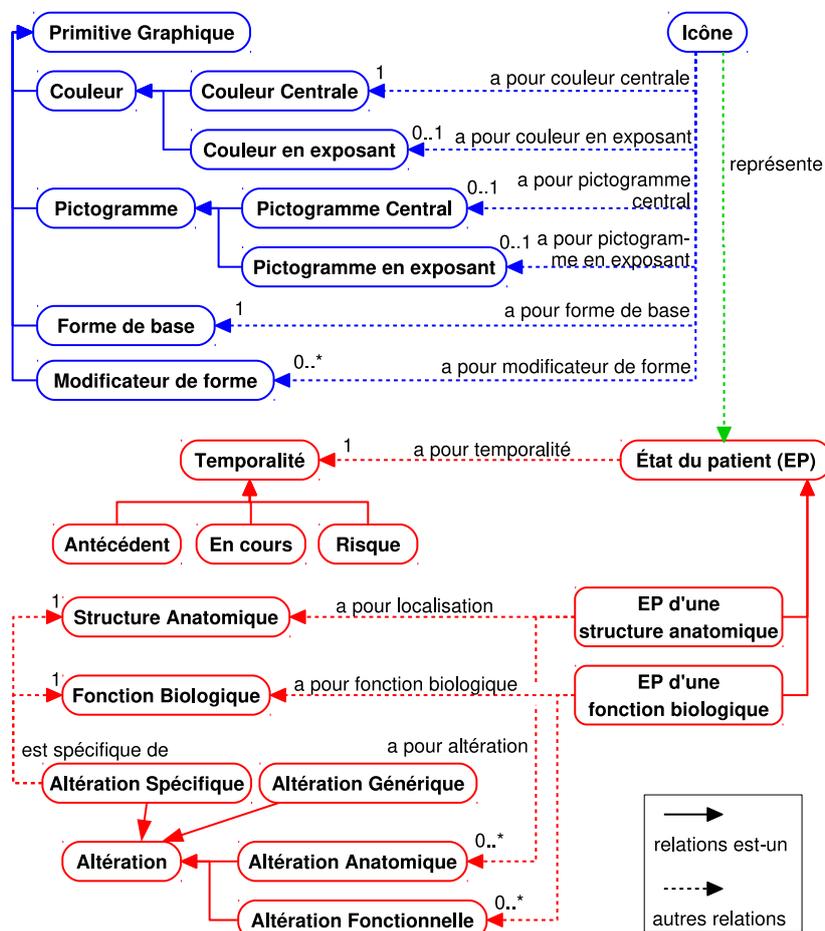


Figure 11. Principaux concepts et relations dans l'ontologie VCM. Les couleurs correspondent aux trois modules de l'ontologie

5.2.2. Structure de l'ontologie VCM

L'ontologie VCM (figure 11, voir (Lamy, Soualmia *et al.*, 2013) pour une version plus détaillée) a été découpée en trois modules. Le premier (ontologie des icônes, comprenant 240 concepts, 21 relations et 2 597 axiomes) décrit les primitives et les icônes VCM. Il a été généré automatiquement à partir de la liste des primitives du langage VCM et de scripts Python. 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 (ontologie du domaine, 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 ne comporte pas l'ensemble des maladies, traitements... qui peuvent être générés 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 (*Unified Medical Language System*) (Kashyap, Borgida, 2003), et en se limitant à un niveau de granularité élevé qui est celui de VCM.

Les concepts de ces deux ontologies, des icônes et du domaine, sont reliés par des relations « représente » et « est représenté par » (509 axiomes) qui constituent le troisième module. 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 graphique le ou les concepts médicaux correspondants, à l'aide de scripts Python.

L'ontologie dans sa totalité comprend 609 concepts, 41 relations et 3 934 axiomes, définis à l'aide du langage OWL-DL. Elle appartient à la famille *ALCRIO* des logiques de description (*Attribute Language, Complex concept negation, Role, Inverse property, Qualified cardinality restriction*), pour laquelle il est démontré que le raisonnement est décidable (Horrocks, Sattler, 2003).

5.3. Vérification de la consistance des icônes

La première application de l'ontologie VCM a été la vérification de la consistance des icônes (Lamy, Soualmia *et al.*, 2013). 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 est absurde du point de vue médical. Ces icônes inconsistantes peuvent poser problème, notamment lorsque des utilisateurs sont amenés à construire eux-mêmes une icône en sélectionnant plusieurs primitives. De la même manière que pour les panneaux routiers, l'ontologie VCM permet de vérifier la consistance des icônes.

La figure 12 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.

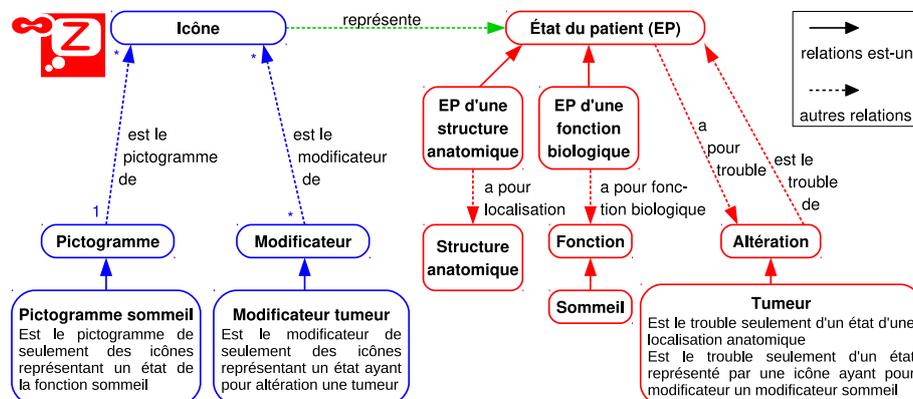


Figure 12. Représentation de l'icône inconsistante « tumeur du sommeil » dans l'ontologie

Ce raisonnement a pu être mis en œuvre avec le raisonneur Hermit (Motik *et al.*, 2009), soit pour une seule icône (un individu), 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, Soualmia *et al.*, 2013).

5.4. Alignement de VCM avec les terminologies médicales : la SNOMED CT

L'utilisation de VCM dans les applications médicales nécessite son alignement avec les référentiels existants, afin notamment de pouvoir associer automatiquement des icônes aux ressources déjà indexées par les concepts de ces référentiels, comme par exemple les Dossiers Patient Informatisés. Les premiers alignements entre VCM et les terminologies médicales ont été réalisés de manière manuelle (Griffon *et al.*, 2014). Dans cette section, nous étudions l'alignement semi-automatique avec 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 subsumption *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 VCM. L'alignement s'est fait en deux étapes : (1) l'alignement manuel des concepts médicaux de l'ontologie VCM ($n=369$) aux termes SNOMED CT correspondants (structures anatomiques, étiologies, *etc.*), puis (2) l'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, Tsopra *et al.*, 2013) ; nous présentons ici les résultats sur l'ensemble de la SNOMED CT.

5.4.1. *Alignement manuel entre concepts médicaux de l'ontologie VCM et la SNOMED CT*

Les concepts de l'ontologie 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 par extension, 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 sera représentée par 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 fait le choix d'associer chaque structure anatomique à au plus un pictogramme (l'oreille dans notre exemple). De la sorte, seules les maladies touchant plusieurs organes se voient associer plusieurs icônes VCM.

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 CIM-10 (Classification Internationale des Maladies, version 10).

Au total, l'alignement manuel a fait intervenir 1 753 termes SNOMED CT et 369 concepts de l'ontologie VCM.

5.4.2. *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 concepts de l'ontologie VCM à l'aide de l'alignement précédent, et en prenant en compte les relations « est-un » et « partie-de » existantes 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* » (uvéite) est décomposé en une structure anatomique, « *uveal tract* », et une morphologie, « inflamm-

tion ». « *Uveal tract* » est une partie de « *entire eye* » qui est une « *structure of visual system* », laquelle est traduite en « structure visuelle » dans l'ontologie du domaine. « Structure visuelle » est alors traduit en pictogramme « œil » dans l'ontologie des icônes. « 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 n'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*).

5.5. 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 13) a été rédigée manuellement. Cette rédaction manuelle pose deux problèmes : (1) le lexique doit être remis à jour systématiquement à chaque modification de VCM, et (2) 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 VCM.

	Poumon, respiration [poumon]		Glandes, système endocrine [glande]
	Bronches [bronches]		Thyroïde [thyroïde]
	Sténose bronchique [bronches_bouche]		Pancréas [pancreas]
	Oreille, audition, équilibre [oreille]		Diabète [glande_diabete]

Figure 13. Trois extraits du lexique des primitives VCM rédigé manuellement. Les identifiants des primitives figurent entre crochets

5.5.1. 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, la récupération *via* les relations de l'ontologie de la liste des concepts médicaux que la primitive peut représenter, (3) le tri de ces concepts et (4) l'obtention du ou des libellés associés à chaque concept, puis la 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 aussi bien un organe que sa fonction. La liste de concepts inclut également tous les concepts descendants (fils, petit-fils... par exemple pour le pictogramme « poumon », « 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 libellés correspondants (les noms d'organe sont généralement plus lisibles que les noms de structure correspondants, par exemple « poumon » est plus lisible que « 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 »).

À 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...).

5.5.2. *Le lexique produit à partir de l'ontologie*

 <ul style="list-style-type: none"> Poumon Structure respiratoire Structure pulmonaire Structure de la plèvre Fonction respiratoire, inclut "respiration" Structure cartilagineuse des voies respiratoires Structure lymphoïde pulmonaire [poumon]  <ul style="list-style-type: none"> Bronches Structure des voies respiratoires inférieures Structure bronchique Structure de la trachée Fonction bronchique Structure muqueuse bronchique Structure muqueuse de la trachée Structure cartilagineuse des voies respiratoires inférieures [bronches]  <ul style="list-style-type: none"> Obstruction, inclut "embolie", "lithiase", "sténose" Structure des voies respiratoires inférieures Structure bronchique Structure de la trachée [bronches_bouche] 	 <ul style="list-style-type: none"> Structure endocrine, inclut "système endocrine" Structure surrénale Fonction de régulation hormonale Autre fonction de régulation hormonale Structure hypophysaire Structure de l'épiphyse [glande]  <ul style="list-style-type: none"> Parathyroïde Thyroïde Structure parathyroïdienne Structure thyroïdienne Fonction de régulation thyroïdienne Fonction de régulation parathyroïdienne [thyroïde]  <ul style="list-style-type: none"> Pancréas Structure pancréatique Fonction de production pancréatique [pancreas]  <ul style="list-style-type: none"> Diabète Diabète type 2 Diabète type 1 [glande_diabete]
 <ul style="list-style-type: none"> Oreille Structure auditive, inclut "structure du système vestibulaire" Fonction auditive, inclut "équilibre" Structure de l'antre mastoïdien, inclut "Recessus epitympanicus" Structure cutanée de l'oreille Structure nerveuse auditive, inclut "structure nerveuse central auditive" Structure musculaire des osselets Structure du vestibule osseux Structure des osselets Structure articulaire des osselets [oreille] 	

Figure 14. Trois extraits du lexique des primitives VCM produit à partir de l'ontologie

La figure 14 montre des extraits du lexique produit à partir de l'ontologie. En comparaison avec le lexique d'origine (figure 13), le nouveau lexique est incontestablement plus riche.

L'alignement des concepts médicaux de l'ontologie 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 se trouvent 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 la « structure des osselets » apparaît en face du pictogramme « oreille » mais pas en face du pictogramme « os »).

5.6. *Génération de libellés multilingues pour les icônes VCM*

Afin de faciliter l'apprentissage du langage VCM par les utilisateurs, il est pratique de disposer d'un libellé textuel pour chaque icône, comme ceux de la figure 10. En utilisant l'ontologie VCM, il est possible de générer automatiquement ces libellés.

{concepts médicaux,...} ⇒ "libellé français / libellé anglais" (segment)

- {altération pathologique} ⇒ "maladie / *disorder*" (base)
- {altération pathologique, hypofonction} ⇒ "insuffisance / *insufficiency*" (base)
- {altération pathologique, inflammation} ⇒ "inflammation / *inflammation*" (base)

- {structure respiratoire} ⇒ "des voies respiratoires / *of respiratory tract*" (comp.)
- {fonction respiratoire} ⇒ "respiratoire / *respiratory*" (adj.)
- {structure respiratoire, fonction respiratoire}
⇒ "du système respiratoire / *of respiratory system*" (comp.)
- {altération pathologique, structure respiratoire, hyperfonction, régulation de la pression artérielle} ⇒ "hypertension pulmonaire / *pulmonary hypertension*" (base)

- {structure cardiaque} ⇒ "du cœur ou du péricarde / *of heart or pericardia*" (comp.)
- {structure cardiaque, fonction cardiaque}
⇒ "du cœur ou du péricarde / *of heart or pericardia*" (comp.)

- {structure gastrique} ⇒ "de l'estomac / *of stomach*" (comp.)

- {structure vasculaire} ⇒ "vasculaire / *vascular*" (adj.)
- {structure vasculaire, structure cardiaque} ⇒ "coronaire / *coronary*" (adj.)

Figure 15. Extrait du dictionnaire contenant les segments de libellés correspondant aux icônes de la figure 10

5.6.1. Méthode de génération des libellés

Les libellés sont segmentés en six parties : pré-préfixe, préfixe, base, adjectif (adj.), complément (comp.) et suffixe. Par exemple, le libellé « traitement des antécédents de maladie vasculaire des voies respiratoires avec obstruction » peut être découpé ainsi :

- pré-préfixe : médicament de
- préfixe : antécédents de
- base : maladie
- adjectif : vasculaire
- complément : des voies respiratoires
- suffixe : avec obstruction

Pour pouvoir générer automatiquement les libellés, nous avons manuellement créé un dictionnaire faisant correspondre des sous-ensembles de concepts médicaux à des segments de libellés bilingues (voir extrait figure 15). Ce dictionnaire contient des segments de libellés de différents niveaux de granularité, à la fois très général (par exemple « maladie ») et très spécifique (« hypertension pulmonaire »). Les entrées du dictionnaire contenant des sous-ensembles de concepts plus spécifiques (nombre de concepts supérieur) sont prioritaires sur celles qui ont des sous-ensembles plus

petits. Par exemple, l'entrée {altération pathologique, hypofonction} est prioritaire sur l'entrée {altération pathologique}. De plus, des règles de priorité ont été définies à la main entre certaines entrées de même taille, lorsque cela s'avère nécessaire.

La génération d'un libellé à partir d'une icône se fait en 5 étapes : (1) obtenir les primitives graphiques de l'icône issues de l'ontologie des icônes et construire l'icône dans l'ontologie VCM, (2) obtenir l'ensemble des concepts médicaux issus de l'ontologie du domaine correspondant à l'icône *via* l'ontologie VCM, (3) rechercher au sein de cet ensemble les sous-ensembles entrées du dictionnaire, en prenant en compte l'ordre de priorité défini précédemment, (4) obtenir les segments de libellé correspondant à ces sous-ensembles dans le dictionnaire, et (5) assembler les segments de libellés pour construire le libellé final. En français, l'ordre d'assemblage des six segments est le suivant : pré-préfixe, préfixe, base, adjectif, complément, suffixe. En anglais, l'ordre défini est : pré-préfixe, préfixe, adjectif, base, complément, suffixe (l'adjectif est placé avant le nom). Lorsque plusieurs segments du même type sont présents, par exemple deux adjectifs, l'ordre est arbitraire.

Par exemple, pour l'icône « insuffisance respiratoire », (1) les primitives graphiques sont : pictogramme « poumon », modificateur de forme « flèche vers le bas » et couleur « rouge ». (2) L'ensemble des concepts médicaux associé à l'icône est {altération pathologique, en cours, fonction respiratoire, hypofonction}. (3) Les sous-ensembles suivants correspondent dans le dictionnaire (figure 15) : {altération pathologique, hypofonction} et {fonction respiratoire} (le sous-ensemble {altération pathologique} est aussi présent, mais {altération pathologique, hypofonction} est prioritaire et « consomme » le concept « altération pathologique » qui n'est ainsi plus disponible). (4) À l'aide du dictionnaire, on obtient les segments de libellé suivants (en français) : « insuffisance » (base) et « respiratoire » (adj.). (5) Leur assemblage conduit au libellé final : « insuffisance respiratoire ».

5.6.2. Libellés résultants

Le système de génération des libellés permet de produire des libellés bilingues français et anglais pour l'ensemble des icônes VCM. À titre d'exemple, les libellés de la figure 10 ont été générés de cette manière. Le dictionnaire complet contient 460 segments de libellés bilingues (français - anglais) dont 257 ont dû être classés par ordre de priorité.

6. Discussion

Dans cet article, nous avons proposé une méthode pour formaliser la sémantique d'un langage iconique, à l'aide d'une ontologie formelle qui décrit les *icônes* du langage et leur syntaxe d'une part, les *objets du domaine* représentés par les icônes et les primitives graphiques d'autre part, et les relations « représente » qui existent entre les deux. Nous avons montré comment cette méthode pouvait être appliquée à deux langages iconiques : la signalisation routière et le langage VCM (Visualisation des Connaissances en Médecine). Nous avons aussi montré l'intérêt de cette formalisation

à travers différentes applications : la vérification de la consistance des icônes, l'alignement des icônes avec des ressources existantes, la génération d'un lexique des pictogrammes et la génération de libellés multilingues pour les icônes. Ces applications 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 développée a permis de les réaliser toutes à partir d'une même source de connaissance. Il pourrait aussi être intéressant d'évaluer ces applications auprès d'utilisateurs finaux et de les comparer à des applications similaires mais construites à partir de représentations des connaissances spécifiques à chacune.

6.1. Grammaire ou sémantique ?

Dans l'état de l'art (Section 2), nous avons vu que plusieurs approches proposées pour formaliser la sémantique des langages graphiques sont en fait de nature grammaticale voire syntaxique. Par exemple, nous aurions pu déterminer la consistance des panneaux routiers ou des icônes VCM à l'aide d'une grammaire formelle, telle que la suivante :

```

<panneau> ::= <panneau d'interdiction> | <panneau d'obligation>
              | <panneau d'information de danger>
              | <autre panneau d'information>
<panneau d'information de danger> ::= triangle rouge avec <rencontre> au milieu
<panneau d'interdiction> ::= rond rouge avec <action> au milieu
<rencontre> ::= virage à gauche, travaux
<action> ::= virage à gauche, rouler à plus de 50 km/h

```

Cependant, cette formalisation intègre sans différenciation des éléments hétérogènes : des éléments graphiques (rond rouge) et des éléments non graphiques (travaux). Elle ne sépare pas l'objet représenté de sa représentation, ni la syntaxe de la sémantique. Cependant, le panneau « interdiction de travaux » ou l'icône VCM « cancer du sommeil » sont parfaitement corrects d'un point de vue syntaxique, même si tous deux sont absurdes d'un point de vue sémantique.

L'utilisation d'une grammaire pour exprimer la sémantique nous semble peu appropriée car la sémantique fait souvent appel à des relations d'héritage (est-un), pour lesquelles une ontologie est plus adaptée qu'une grammaire. Par ailleurs, les différentes applications que nous avons présentées n'auraient pas pu être réalisées si nous n'avions pas exprimé la sémantique indépendamment de la syntaxe. Par exemple, il aurait été difficile de générer un lexique des pictogrammes à partir d'une grammaire dans laquelle pictogrammes et concepts médicaux auraient été mélangés.

6.2. Discussion sur les applications

La première application de l'ontologie a porté sur la vérification de la consistance des icônes (Lamy, Soualmia *et al.*, 2013). Pour cette application, la plus grande difficulté est relative à la polysémie qui peut être très importante dans un langage iconique.

En effet, un langage comme VCM n'a pas pour objectif d'être aussi précis que le langage médical textuel, ce qui conduit à une polysémie importante. À titre d'exemple, le même pictogramme signifie à la fois « poumon » et « respiration ». L'utilisation d'ontologies s'avère une bonne méthode pour régler ce genre de problème (Pisanelli *et al.*, 2004).

Ensuite, nous avons aligné les icônes VCM avec une terminologie de référence en médecine, la SNOMED CT (Lamy, Tsopra *et al.*, 2013). Cet alignement a pu être réalisé de manière semi-automatique grâce aux relations présentes dans l'ontologie et dans la SNOMED CT, cependant nous n'avons pas encore réussi à reproduire cette méthode avec des ressources terminologiques moins structurées ou formalisées, telle que la CIM10. De telles ressources nécessiteront probablement un alignement manuel.

Nous avons également produit un lexique des primitives graphiques à partir de l'ontologie du langage iconique. Le lexique ainsi construit se révèle plus riche qu'un lexique rédigé manuellement, et l'alignement des concepts de l'ontologie avec les concepts des ressources terminologiques 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.

Une difficulté importante que nous avons rencontrée lors de la génération du lexique est relative au choix de l'ordre des éléments présents dans le lexique. En effet, l'ontologie ne définit pas d'ordre entre concepts (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 afin d'ordonner les différents concepts listés en face d'une entrée du lexique, en prenant en compte plusieurs éléments : la nature, le niveau d'échelle, la spécificité, et la lisibilité des libellés des concepts. En revanche, l'ordre dans lequel figurent 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 relatifs aux ontologies et aux lexiques visent à construire une ontologie en s'appuyant sur un lexique existant, ce qui est le chemin inverse de ce que nous présentons dans cet article. Cependant, la production de lexique textuel à partir d'ontologie a déjà été envisagée (Hirst G, 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 similaire à celui rencontré par les outils générant des descriptions en langage naturel à partir d'une ontologie, tels 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.

Enfin, nous avons présenté une méthode automatique pour générer des libellés multilingues associés aux icônes. Nous avons appliqué cette méthode au français et à l'anglais, cependant la génération de libellés dans des langues plus éloignées telles que l'arabe ou le chinois pourrait poser de nouveaux problèmes.

Au cours de travaux futurs, il serait intéressant d'essayer d'étendre notre approche dans des contextes où la réalité est moins formalisée, par exemple pour les icônes des interfaces utilisateurs dans les logiciels, et d'évaluer ce qu'une telle approche pourrait apporter.

7. Conclusion

En conclusion, nous avons proposé une méthode générique pour formaliser la sémantique des langages iconiques. La méthode a pu être appliquée à la signalisation routière et à VCM, un langage d'icônes médicales. Elle a permis non seulement de clarifier la grammaire et la sémantique du langage, mais aussi de réaliser différentes applications sémantiques. Cette méthode pourrait faciliter la création de nouveaux langages iconiques, améliorer la cohérence et la qualité de ces langages, faciliter leur apprentissage par la production automatique de lexiques et de libellés, et enfin faciliter leur utilisation en générant des alignements avec des ressources existantes.

Remerciements

Ce travail a été en partie 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).

Bibliographie

- Cornet R. (2009). Definitions and qualifiers in SNOMED CT. *Methods Inf Med*, vol. 48, n° 2, p. 178-183.
- Costagliola G., De Lucia A., Orefice S., Tortora G. (1997). A parsing methodology for the implementation of visual systems. In *Ieee transactions on software engineering*, vol. 23, p. 777-799.
- Costagliola G., De Rosa M., Fuccella V. (2015). Extending local context-based specifications of visual languages. *Journal of Visual Languages and Computing*, vol. 31, p. 184-195.
- Dreyfuss H. (1984). *Symbol sourcebook: An Authoritative Guide to International Graphic Symbols*. John Wiley and sons.
- Ellner S., Taha W. (2007). The Semantics of Graphical Languages. In *Proceedings of the acm sigplan symposium on partial evaluation and semantics-based program manipulation*.

- Ely J. W., Osheroff J. A., Ebell M. H., Chambliss M. L., Vinson D. C., Stevermer J. J. *et al.* (2002). Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ*, vol. 324, n° 7339, p. 710.
- Erwig M. (1997). *Semantics of Visual Languages* (vol. 300-307).
- Griffon N., Kerdelhué G., Soualmia L. F., Merabti T., Grosjean J., Lamy J. B. *et al.* (2014). Evaluating alignment quality between iconic language and reference terminologies using similarity metrics. *BMC medical informatics and decision making*, vol. 14, p. 17.
- Haarslev V. (1998). Visual Language Theory. In, vol. 261-292, chap. A fully formalized theory for describing visual notations. New York, Springer.
- Hirst G. (2009). Handbook on ontologies. In, chap. Ontology and the Lexicon. Springer.
- Horrocks I., Sattler U. (2003). Decidability of SHIQ with complex role inclusion axioms. *Artificial Intelligence*, vol. 160, p. 2004.
- Kashyap V., Borgida A. (2003). Representing the UMLS Semantic Network in OWL. In *Proceedings of iswc 2003 (international semantic web conference)*, vol. 1-16.
- Konstantopoulos S., Karkaletsis V., Vogiatzis D., Bilidas D. (2011). Language Technology for Cultural Heritage. In, vol. 115-132, chap. Authoring semantic and linguistic knowledge for the dynamic generation of personalized descriptions. Springer Berlin Heidelberg.
- 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*, vol. 8, p. 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*, vol. 265-276. Hammamet, Tunisie, PUG.
- Lamy J. B., Soualmia L. F., Kerdelhué G., Venot A., Duclos C. (2013). Validating the semantics of a medical iconic language using ontological reasoning. *J Biomed Inform*, vol. 46, n° 1, p. 56-67.
- Lamy J. B., Tsopra R., Venot A., Duclos C. (2013). A Semi-automatic Semantic Method for Mapping SNOMED CT Concepts to VCM Icons. *Stud Health Technol Inform*, vol. 192, p. 42-6.
- Lamy J. B., Venot A., Bar-Hen A., Ouvrard P., Duclos C. (2008). Design of a graphical and interactive interface for facilitating access to drug contraindications, cautions for use, interactions and adverse effects. *BMC Medical Informatics and Decision Making*, vol. 8, p. 21.
- Lamy J. B., Venot A., Duclos C. (2015). PyMedTermio: an open-source generic API for advanced terminology services. *Stud Health Technol Inform*, vol. 210, p. 924-928.
- Lamy JB. (2016). Ontology-Oriented Programming for Biomedical Informatics. *Stud Health Technol Inform*, vol. 221, p. 64-68.
- Ma X., Matta N., Cahier J. P., Qin C., Cheng Y. (2015). From action icon to knowledge icon: Objective-oriented icon taxonomy in computer science. *Displays*, vol. 39, p. 68-79.
- Marriott K. (1994). Constraint multiset grammars. In *Proceedings of ieee symposium on visual languages*.

- Meunier JG. (1988). La structure générique des systèmes sémiotiques. *Recherche sémiotique / Semiotic inquiries (RSSI)*, n° 8.
- Meunier JG. (1998). The categorial structure of iconic languages. *Theory&Psychology*, vol. 8, n° 6, p. 805-825.
- Motik B., Shearer R., Horrocks I. (2009). Hypertableau reasoning for description logics. *Journal of Artificial Intelligence Research*, vol. 36, p. 165-228.
- Nakamura C., Zeng-Treitler Q. (2012). A taxonomy of representation strategies in iconic communication. *Int. J. Human-Computer Studies*, vol. 70, n° 2012, p. 535-551.
- Pereira S., Hassler S., Hamek S., Boog C., Leroy N., Beuscart-Zéphir M. C. *et al.* (2014). Improving access to clinical practice guidelines with an interactive graphical interface using an iconic language. *BMC medical informatics and decision making*, vol. 14, n° 1, p. 77.
- Pisanelli D. M., Gangemi A., Battaglia M., Catenacci C. (2004). Coping with medical polysemy in the semantic web: the role of ontologies. *Stud Health Technol Inform*, vol. 107, p. 416-419.
- Rosse C., Mejino JL V. (2003). A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*, vol. 36, p. 478-500.
- Simon C., Hassler S., Beuscart-Zéphir M. C., Favre M., Venot A., Duclos C. *et al.* (2014). Using an iconic language to improve access to electronic medical records in general medicine. *Stud Health Technol Inform*, vol. 205, p. 333-7.