



**HAL**  
open science

## Intégration de multiples ontologies en anatomie pathologique

David Ouagne, Christel Daniel-Le Bozec, Eric Zapletal, Maxime Thieu,  
Marie-Christine Jaulent

► **To cite this version:**

David Ouagne, Christel Daniel-Le Bozec, Eric Zapletal, Maxime Thieu, Marie-Christine Jaulent. Intégration de multiples ontologies en anatomie pathologique. IC - 16èmes Journées francophones d'Ingénierie des Connaissances, May 2005, Nice, France. pp.193-204. hal-01024017

**HAL Id: hal-01024017**

**<https://inria.hal.science/hal-01024017>**

Submitted on 15 Jul 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.

# Intégration de multiples ontologies en anatomie pathologique

David Ouagne<sup>1</sup>, Christel Daniel-Le Bozec<sup>1</sup>, Eric Zapletal<sup>1</sup>, Maxime Thieu<sup>1</sup>,  
Marie-Christine Jaulent<sup>1</sup>

<sup>1</sup>INSERM, U729, Paris, F-75006 France,

{David.Ouagne, Christel.Lebozec, Eric.Zapletal, Maxime.Thieu,  
Marie-Christine.Jaulent}@spim.jussieu.fr

**Résumé** : La variabilité diagnostique en anatomie pathologique est en partie liée à l'utilisation de systèmes de classification différents, pouvant être considérés comme des points de vue différents, pour décrire des lésions. Notre objectif est de représenter ces points de vue et de proposer une solution pour permettre leur interopérabilité. L'approche hybride décrite par Wache nous permet de développer un système multi ontologique en trois étapes 1) la représentation des points de vue au sein d'ontologies locales, 2) la construction d'un vocabulaire partagé et 3) le développement d'un outil de traduction. L'évaluation du travail, conduite sur 33 cas, a consisté à évaluer les ontologies locales grâce à un outil de validation sémantique de cas et à évaluer l'outil de traduction. Nos résultats montrent que les pathologistes produisent des descriptions qui ne suivent pas toujours les règles d'interprétation des systèmes de classification auxquels ils se réfèrent. Si 62.5% à 100% des concepts des ontologies locales sont traduisibles, nous avons constaté que la validité des cas n'était pas toujours conservée après traduction.

**Mots-clés** : Représentation des connaissances, interopérabilité sémantique, alignement d'ontologies, imagerie médicale, cancer du sein

## 1 Introduction

L'examen anatomopathologique permet d'établir un diagnostic et de donner des indications pronostiques concernant des lésions observées au niveau d'images de tissus ou cellules issues de prélèvements. Les règles qui permettent d'établir une conclusion diagnostique et/ou pronostique à partir de caractéristiques morphologiques observées dans les images créées au cours de l'examen sont publiées dans le cadre de systèmes de classification. Il existe en anatomie pathologique une variabilité diagnostique largement rapportée dans la littérature (Fleming, 1996). Celle-ci est en partie liée au fait que les systèmes de classification sont nombreux et évolutifs. Ainsi, des cas identiques peuvent conduire à des conclusions diagnostiques différentes en fonction du système de classification sur lequel repose le point de vue du pathologiste (Wells, 2000).

Le système IDEM (Images et Diagnostics par l'Exemple en Médecine) a pour objectif d'assister les experts dans la constitution de descriptions consensuelles de

cas anatomopathologiques. Le projet est développé dans notre laboratoire depuis quelques années et propose un environnement dans lequel des descriptions de cas anatomopathologiques peuvent être comparées (Thieu 2004). La plateforme actuelle compare des cas exprimés selon un seul point de vue, c'est-à-dire conformément à une classification déterminée à l'avance. Dans l'optique de faire collaborer des experts distants, il est nécessaire aujourd'hui de comparer des cas exprimés selon différents points de vue.

Dans ce contexte particulier, notre objectif est de modéliser les connaissances traduisant les différents points de vue de pathologistes au sein d'ontologies locales afin de pouvoir comparer et rendre interopérables les conclusions fournies par des pathologistes se référant à des systèmes de classification différents.

Nous avons procédé en trois étapes consistant en 1) la représentation des différents points de vue sous forme d'ontologies locales, 2) la réalisation d'un vocabulaire partagé et le développement d'un environnement informatique permettant d'apparier les ontologies locales au vocabulaire partagé et 3) le développement d'un algorithme permettant de traduire la description d'un cas selon un point de vue dans un point de vue différent.

L'environnement informatique a été développé à partir d'un outil de fusion d'ontologies et trois points de vue différents ont été modélisés pour illustrer l'intérêt de cet environnement. Par la suite, une première étude a été réalisée pour l'évaluation de l'appariement entre deux de ces trois points de vue. Trente trois cas de pathologie tumorale mammaire ont été décrits par deux experts selon deux points de vue différents. Les descriptions ont été validées dans le sens où, pour chaque point de vue envisagé, leur conformité aux règles d'interprétation modélisées dans les ontologies locales a été vérifiée. En utilisant l'environnement informatique développé, chaque description a ensuite été traduite selon l'autre point de vue. Une première étude de l'interopérabilité a consisté à comparer la validité des descriptions originelles à celle des descriptions obtenues après traduction. Cette comparaison s'appuie sur la mesure du Kappa (Falissard 2001).

## 2 État de l'art

La prise en compte du point de vue dans les systèmes à base de connaissances permet d'indexer les connaissances afin de les rendre plus accessibles et réutilisables. Ribière analyse les modèles et les formalismes permettant la représentation de multiples points de vue (Ribière, 1999).

Parmi ces formalismes, les ontologies impliquent effectivement une certaine vue du monde par rapport à un domaine donné et sont conçues comme un ensemble des concepts d'un domaine – e.g. entités, attributs, processus –, leurs définitions et leurs interrelations (Bachimont, 2000) (Grüber, 1993) (Uschold, 1996). La définition d'une méthodologie de construction d'ontologies a fait l'objet de nombreux travaux. Uschold et Grüniger (Uschold, 1996) proposent une méthode en trois phases successives consistant à : identifier le domaine d'application et le contexte d'utilisation (la tâche) de l'ontologie, puis à construire effectivement l'ontologie,

c'est à dire à acquérir les concepts du domaine et à les coder dans un modèle conceptuel formel et enfin à valider l'ontologie construite.

En ce qui concerne l'intégration de connaissances hétérogènes (issues de points de vue différents), un certain nombre de travaux présentent des solutions qui s'appuient sur une représentation ontologique des connaissances permettant la comparaison des points de vue.

Selon Wache (Wache, 2001), on distingue généralement trois approches pour l'intégration de connaissances hétérogènes. L'approche mono ontologie vise à construire une seule ontologie exprimant l'ensemble des sources de connaissances (Noy, 2003) ; l'approche multi ontologies s'appuie sur des ontologies locales et définit des opérations d'appariement ou de transformation des ontologies locales (Mena, 2000) ; enfin l'approche hybride s'appuie sur des ontologies locales et définit un vocabulaire partagé et des relations – solutions d'appariement et de transformation – entre ce vocabulaire partagé et les ontologies locales (Buccella, 2003).

Des solutions techniques permettant de gérer de multiples ontologies se sont développées. Ces solutions peuvent être classées selon l'opération réalisée (appariement, transformation, gestion de version, etc.) ou selon les techniques de mise en correspondance utilisées (Noy, 2003). Une des difficultés des démarches multi ontologies et « hybride » est la définition de l'appariement (« mapping »). Il existe plusieurs environnements effectuant des opérations d'appariement en se basant sur des agents traducteurs (KRAFT (Preece, 1999)), les logiques de description (OBSERVER (Mena, 2000)), un modèle probabiliste faisant intervenir l'expression en langue naturelle des concepts (GLUE (Doan, 2002), Labo ISI/USC (Hovy, 1998)), des techniques lexicales (co-occurrences de termes dans des corpus) (ONION (Mitra, 2000)). L'environnement à utiliser doit être choisi en fonction de la tâche envisagée (fusion, différence, appariement, transformation, etc.) et des ontologies locales dont on dispose.

Le problème posé de l'intégration de connaissances hétérogènes s'apparente à la comparaison, l'alignement et la fusion d'ontologies. L'ambition du travail présenté n'est pas de comparer formellement les méthodologies d'alignement mais plutôt d'adapter des outils existants au domaine spécifique de l'anatomie pathologique afin d'étudier l'intérêt et la portée de ces approches dans notre domaine. Notre contribution est donc essentiellement applicative bien que par certains aspects, la réflexion puisse être généralisée. Pour des raisons pratiques, liées à l'historique du projet IDEM, le choix d'un éditeur d'ontologie s'est porté sur Protégé.

### **3 Construction d'un système multi ontologique pour la comparaison des points de vue en anatomie pathologique**

Nous avons suivi l'approche hybride décrite par Wache qui se compose de trois étapes, 1) la construction d'ontologies locales, 2) la construction du vocabulaire partagé et 3) la définition des relations entre ce vocabulaire partagé et les ontologies locales (Buccella 2003) (cf. figure 1). Une fois le système multi ontologique construit, nous sommes capables de comparer les concepts des ontologies locales

entre elles, et plus particulièrement dans notre cas, d'exprimer un concept d'une ontologie dans une autre ontologie (opération de traduction).

### 3.1 Construction des ontologies locales

Nous avons défini un modèle de description des cas anatomopathologiques puis identifié les différents points de vue de pathologistes existant dans ce domaine et nous les avons représentés au sein d'ontologies locales.

- Définition des points de vue

Il existe dans la littérature, un modèle de démarche diagnostique en anatomie pathologique, formalisé par des guides de bonnes pratiques diagnostiques et appelé « anatomie pathologique basée sur le niveau de preuve » (ou « evidence-based pathology »). Selon ce modèle, la démarche diagnostique consiste à reconnaître des anomalies morphologiques au sein d'images et à les interpréter en accord avec des règles diagnostiques publiées résultant de consensus d'experts (Fleming, 2002). Des systèmes de classification diagnostique explicitant ces règles permettent de conclure à un diagnostic à partir de l'analyse d'anomalies morphologiques. Nous nous sommes restreints au domaine de la pathologie tumorale mammaire et en particulier du diagnostic du Carcinome Canalaire In Situ (CCIS) pour lequel les trois systèmes de classification diagnostique principaux sont ceux de Holland, Lagios et Van Nuys (Wells, 2000).

Le système de classification est défini par l'ensemble constitué d'une part des termes associés aux anomalies morphologiques complexes (e.g. « bas grade nucléaire ») ou élémentaires (e.g. « augmentation modérée de la taille nucléaire ») et aux diagnostiques morphologiques (e.g. « CCIS de bas grade »), et d'autre part des règles d'interprétation permettant d'inférer un diagnostic morphologique à partir d'anomalies morphologiques (par exemple, selon Van Nuys, la présence d'un « bas grade nucléaire » ou d'un « grade nucléaire intermédiaire » associé à l'« absence de nécrose » permet de conclure au diagnostic morphologique de « CCIS de bas grade »).

- Construction des ontologies locales

Nous avons utilisé la méthode de construction d'ontologie décrite par Uschold et Grüninger (Uschold, 1996). En pratique, deux experts ont extrait des trois systèmes de classification du CCIS – Holland, Lagios et Van Nuys – les anomalies morphologiques pertinentes par rapport au diagnostic de grade de CCIS.

L'organisation des concepts de l'interprétation des images anatomopathologiques a bénéficié des efforts de standardisation de la discipline ayant abouti à la constitution de ressources terminologiques électroniques telles que la SNOMED CT (Lieberman, 2003). Conformément à la SNOMED CT, nous avons organisé les anomalies morphologiques selon une hiérarchie taxinomique (relations « est un ») par région anatomique observée (par exemple, noyau, cellule, canal, etc.) puis par type de caractéristique morphologique (par exemple, taille, forme, etc). De plus, les anomalies morphologiques complexes telles que « CCIS de haut grade » et « grade nucléaire élevé » ont été mises en relation avec les anomalies morphologiques justifiant ces interprétations selon les systèmes de classification de référence choisis

grâce à la relation « est inféré par ». Les ontologies locales ont été codées en utilisant l'éditeur d'ontologie Protégé (Noy, 2003) (cf. figure 2).

- Description et validation d'un cas à partir d'une ontologie locale

Une ontologie locale permet de décrire un cas selon un système de classification donné en fournissant aux pathologistes les termes standardisés du domaine ainsi que les connaissances nécessaires à la description des cas (règles d'interprétation).

Concrètement, un plug-in Protégé a été développé afin d'afficher le formulaire permettant aux pathologistes de décrire le cas en utilisant les termes désignant les concepts de l'ontologie locale (Van Nuys ou Holland ou Lagios). Ce plug-in produit une description XML d'un cas. Une originalité du travail réalisé a consisté à développer un algorithme permettant de valider les descriptions de cas selon une ontologie locale. La validité d'un cas selon un point de vue se définit comme la conformité de ce cas aux règles d'interprétation exprimées dans l'ontologie locale. On dit alors que le cas est valide ou « prototypique ». L'algorithme de validation exploite les règles d'interprétation explicitées au sein de l'ontologie locale de référence par la relation : « est inféré par ». Cet algorithme utilise la méthode classique de validation syntaxique d'un fichier XML (ici, la description du cas) selon un schéma XML (ici, exprimant les règles d'interprétation) et permet de signaler si, au niveau de la description, les diagnostics morphologiques sont correctement inférés.

### **3.2 Construction du vocabulaire partagé**

La méthode de construction d'un vocabulaire partagé est proche de la méthode de fusion d'ontologies qui consiste à construire une ontologie à partir de plusieurs ontologies sources. La différence, en ce qui concerne la construction d'un vocabulaire partagé, est qu'il s'agit de choisir parmi les concepts appariés des ontologies locales celui qui fournit le terme à placer dans le vocabulaire partagé et surtout de créer un lien entre ce terme (dit terme « préféré ») et chacun des termes désignant les concepts des ontologies locales. Dans le cas de la fusion d'ontologies, les concepts appariés des ontologies locales sont fusionnés en un nouveau concept (au sein d'une nouvelle ontologie) sans que soit conservé le lien entre ce concept et les concepts sources. L'idée est donc de « passer » par la terminologie (le vocabulaire partagé) pour conserver ce lien.

Nous avons par ailleurs fait l'hypothèse, qu'il était possible d'adapter une solution logicielle permettant de faire une fusion d'ontologies afin d'obtenir un vocabulaire partagé pour finalement réaliser un appariement entre les concepts. Nous avons vu que le choix d'un environnement multi ontologique dépendait de la tâche envisagée et des ontologies locales dont on dispose. Dans la mesure où les ontologies locales que nous avons construites ne contiennent pas d'instances et que les relations ont une grande importance, nous avons choisi d'utiliser la suite PROMPT (plug-in Protégé), adaptée à ce type d'ontologies et qui permet, entre autres, de comparer et de fusionner des ontologies. De plus, la suite PROMPT possède une interface de développement sous licence libre permettant de réaliser les adaptations envisagées (Noy, 2003).

Lors de l'analyse de deux ontologies, le mode fusion de la suite PROMPT (iPROMPT) affiche un tableau de suggestions sous forme de propositions d'actions de fusion. Si un même concept semble exister dans les ontologies locales, alors iPROMPT propose une fusion des concepts sources dans la nouvelle ontologie. Une fusion peut, et c'est souvent le cas, entraîner d'autres fusions.

Le mode dit « mapping » que nous avons développé à partir du mode fusion pour construire notre vocabulaire partagé s'appuie sur la première étape de suggestion de fusion puis permet de choisir un terme « préféré », placé dans le vocabulaire partagé, et de créer une relation entre ce terme « préféré » et le ou les termes « apparié(s) » désignant les concepts sources suggérés pour la fusion. La relation peut être la relation « *identique* » lorsque le terme « apparié » est celui qui a été choisi comme terme « préféré ». La relation peut-être la relation « *synonyme* » lorsque le concept désigné par le terme « préféré » est identique au concept désigné par le terme « apparié » mais que les termes « préféré » et « apparié » sont synonymes. Enfin, pour augmenter le nombre de concepts traduisibles, la relation peut-être aussi la relation « *similaire* » lorsque le concept désigné par le terme « préféré » est jugé similaire au concept désigné par le terme « apparié ». La relation entre terme « préféré » et terme « apparié » a été formalisée sous forme de relation entre les concepts désignés par ces termes. La relation « MappedToConcept » est définie par quatre propriétés :

- *toTerm* : contient le nom du terme de destination de la relation (terme « apparié »),
- *mapping-type* : contient le type sémantique de la relation entre terme « préféré » et terme « apparié » (identique, synonyme ou similaire),
- *toOntology* : contient le nom de l'ontologie source,
- *mapping-comment* : contient une phrase de justification de la relation.

L'intérêt de l'approche réside dans le fait qu'un terme « préféré » peut être en relation avec plusieurs termes « appariés » désignant des concepts de différentes ontologies sources grâce aux différentes instanciations de la relation « MappedToConcept ».

Une fois toutes les suggestions de fusion traitées, nous pouvons considérer que la construction du vocabulaire partagé est terminée.

### 3.3 La traduction

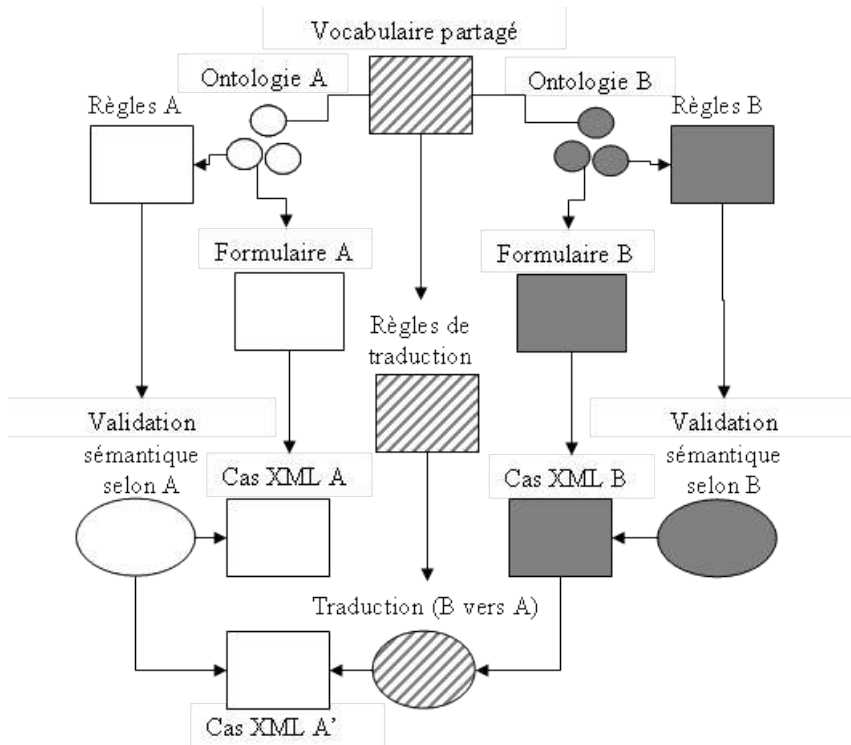
L'algorithme de traduction utilise l'ensemble des instances de la relation « MappedToConcept » pour fournir, à partir d'un cas décrit selon un point de vue, une description de ce même cas décrit selon un point de vue différent.

La méthode de traduction utilise trois paramètres – le concept à traduire, l'ontologie source et l'ontologie destination – afin de retourner le concept de l'ontologie de destination résultant de la traduction

Deux scénarios sont envisageables :

- le concept source est désigné par un terme « préféré » du vocabulaire partagé. Il suffit de trouver, parmi les termes « appariés » au terme « préféré », celui qui désigne un concept de l'ontologie de destination (ce

- concept existe s'il existe une instance de la relation « MappedToConcept » entre ce concept de l'ontologie de destination et le concept désigné par le terme « préféré » dans le vocabulaire partagé).
- le concept source n'est pas désigné par un terme « préféré ». Dans ce cas, il faut d'abord identifier le concept en relation avec ce concept source qui est désigné par un terme « préféré ». On retrouve ensuite le premier scénario décrit ci-dessus.



**Fig. 1** – L'approche hybride permettant l'intégration de multiples ontologies. Construction et évaluation des ontologies locales correspondent aux multiples points de vue. La méthode de validation d'une description de cas selon un point de vue est basée sur un schéma XML exprimant les règles d'interprétation correspondant à l'ontologie locale de référence; construction du vocabulaire partagé et développement d'un algorithme de traduction de la description d'un cas d'un point de vue vers un autre basé sur un appariement entre le vocabulaire partagé et les ontologies locales.



## 4 Résultats

### 4.1 Ontologies locales

Les ontologies locales constituées dans Protégé à partir des classifications de Holland, Van Nuys et Lagios contiennent respectivement 27, 18 et 24 concepts correspondant aux grades de CCIS et aux caractéristiques morphologiques observées au niveau des images. Les relations « est un » et « est inféré par » ont été définies au niveau des trois ontologies locales construites.

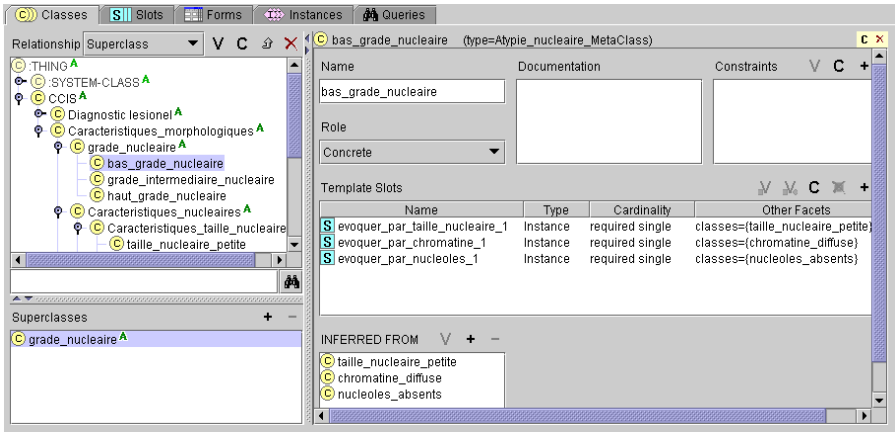


Fig 2. – Ontologie locale

### 4.2 Vocabulaire partagé et module de traduction

Le nouveau mode de fonctionnement, appelé « mapping » intégré dans le système multi ontologique PROMPT s'apparente dans un premier temps au mode fusion, l'utilisateur choisit un des concepts à fusionner et le terme qui le désigne devient le terme « préféré » placé dans le vocabulaire partagé. Puis le logiciel demande à l'utilisateur le type de la relation existant entre le terme « préféré » et chacun des termes désignant les concepts source et lui permet d'explicitier son choix.

Après l'analyse de toutes les suggestions de fusion des concepts des trois ontologies locales prises 2 à 2, le vocabulaire partagé constitué comporte 30 termes « préférés ». De plus, à chaque terme « préféré » de ce vocabulaire partagé sont associées les instanciations disponibles de la relation « MappedToConcept » définissant les règles d'appariement entre le concept désigné par le terme « préféré » et des concepts des différentes ontologies sources.

Le plug-in développé dans Protégé permettant d'extraire, sous forme de schéma XML, les règles d'appariement du vocabulaire partagé permet à l'algorithme de traduction de produire une représentation XML des appariements et d'exprimer un concept d'une ontologie A (classification A) dans une ontologie B (classification B).

## 5 Évaluation de la traduction

- Matériel

Trente-trois cas ont été décrits par un expert à deux reprises en utilisant dans le module de description de la plateforme IDEM le formulaire basé sur l'ontologie locale de Holland d'une part et celui basé sur l'ontologie locale de Van Nuys d'autre part correspondant aux deux systèmes de classification dont ils étaient familiers. La description saisie via le formulaire est ensuite soumise au module de validation, c'est-à-dire qu'on vérifie si elle est conforme aux règles d'interprétation de la classification. Ce module génère des messages d'erreur en cas de non-conformité. Par convention avec les pathologistes, un cas est valide s'il génère au plus un message d'erreur pour la classification Holland et s'il ne génère aucun message d'erreur pour la classification Van Nuys. Ces différences de convention entre les deux classifications sont dues à leur nature différente. La classification de Holland s'appuie sur des descriptions de prototypes et est naturellement plus imprécise et ambiguë que la classification Van Nuys qui s'appuie sur un arbre de décision.

Pour la classification Van Nuys, onze cas sur trente trois (soit 33%) sont conformes aux règles d'interprétation, vingt deux cas ne sont pas conformes dont dix-sept (soit 52%) génèrent un message d'erreur et cinq (soit 15%) deux messages d'erreur ou plus. En ce qui concerne la classification de Holland, aucun des trente trois cas n'est rigoureusement conforme aux règles d'interprétation. C'est à dire qu'aucune des descriptions n'est prototypique d'un des trois grades de carcinome canalaire in situ. Mais, dans la mesure où les cas générant un seul message d'erreur sont considéré comme valides, douze cas sur trente trois (soit 36%) sont valides. Vingt et un cas sur trente trois (soit 64%) génèrent deux messages d'erreur ou plus.

- Traduction des concepts des ontologies locales

Le taux de traduction des ontologies locales a été évalué en calculant le pourcentage de concepts d'une ontologie locale qui étaient traduisibles vers une autre ontologie locale. Nous avons calculé ce taux de traduction pour les trois classifications disponibles sous forme d'ontologies locales (VN : Van Nuys ; HL : Holland ; LA : Lagios). La traduction concept par concept d'une façon générale donne des résultats satisfaisants en terme de pourcentage de concepts traduisibles comme le montre le tableau 1.

**Table 1.** Pourcentage de concepts traduisibles

|           | Nb total de concepts | <i>Concepts traduisibles (%)</i> |           |           |
|-----------|----------------------|----------------------------------|-----------|-----------|
|           |                      | <i>VN</i>                        | <i>HL</i> | <i>LA</i> |
| <i>VN</i> | 18                   |                                  | 100       | 83,3      |
| <i>HL</i> | 27                   | 66,6                             |           | 66,6      |
| <i>LA</i> | 24                   | 62,5                             | 75        |           |

- Traduction des trente trois cas d'une ontologie locale à une autre

La traduction d'une description de cas d'un point de vue vers un autre a été évaluée en utilisant les trente trois cas décrits selon la classification de Van Nuys puis traduits selon la classification de Holland (et réciproquement). Une analyse qualitative de la traduction a été réalisée par un expert.

Par ailleurs, nous nous sommes intéressés à savoir si la traduction conservait la validité des cas. Nous avons comparé la validité des cas décrit selon une classification à la validité des cas obtenus après traduction selon la nouvelle classification. La concordance obtenue est bonne ( $Kappa = 0,94$ ) lors de la traduction de Holland vers Van Nuys et mauvaise ( $Kappa = 0,02$ ) en ce qui concerne la traduction de Van Nuys vers Holland.

## 6 Discussion

Notre problématique consistait à proposer une solution de représentation de points de vue différents de pathologistes pour la démarche diagnostique en anatomie pathologique et à offrir une solution méthodologique et technique pour rendre ces points de vue interopérables. Nous avons mis en œuvre une approche décrite par Wache et expérimentée par Buccella consistant à représenter les points de vue sous forme d'ontologies locales, à construire un vocabulaire partagé et à exprimer les correspondances existant entre les concepts des ontologies locales au niveau du vocabulaire partagé (appariement). Un intérêt important de cette approche est de pouvoir envisager de nouvelles ontologies locales qui venant s'apparier au vocabulaire partagé peuvent alors rapidement interopérer avec les autres ontologies locales. Cet appariement a permis de développer un module de traduction de descriptions structurées de cas exprimés selon une ontologie locale vers une autre ontologie locale. La méthode de traduction repose sur la modification d'une description de cas au format XML selon des règles d'appariement représentées dans le formalisme XML.

Il apparaît dans notre expérimentation que les pathologistes produisent généralement des descriptions de cas de CCIS qui ne correspondent pas rigoureusement aux cas prototypiques décrits dans les systèmes de classification selon lesquels ils interprètent les images. Le module de validation vérifiant la conformité des descriptions permet aux pathologistes de prendre conscience de cet écart entre leur pratique et la situation décrite par le système de classification. De plus, nous avons noté les descriptions sont beaucoup plus souvent conformes en ce qui concerne le système de classification de Van Nuys qui est basé sur un arbre de décision non ambigu alors que le système de classification de Holland, (comme celui de Lagios), repose sur des descriptions prototypiques des grades de CCIS. Or, les cas réels ne correspondent pas souvent à ces descriptions prototypiques.

En ce qui concerne la traduction, les résultats montrent que les concepts sont majoritairement traduisibles d'une ontologie locale à une autre. Un taux élevé de concepts traduisibles (par exemple de Van Nuys vers Holland) n'est pas suffisant pour préjuger de la qualité de la traduction. En effet, si le nombre de concepts de

l'ontologie source est inférieur à celui de l'ontologie de destination (plus riche), même si tous les concepts de l'ontologie source sont traduisibles, il manquera vraisemblablement à la description une fois traduite des concepts de description nécessaires à la bonne application des règles d'interprétation de l'ontologie de destination. A ce titre, nous avons constaté que la traduction des descriptions de cas ne conservait pas toujours la validité de ces cas selon le point de vue considéré. Dans notre expérimentation, c'est particulièrement le cas lors de la traduction de Van Nuys vers Holland. Ce résultat est lié d'une part à la plus grande richesse en concepts de description de la classification de Holland et d'autre part à une expression encore trop rigide des règles d'interprétation au niveau des ontologies locales correspondant aux systèmes de classification basés sur des descriptions prototypiques.

Des améliorations à court terme semblent importantes. En ce qui concerne la traduction, l'algorithme de fusion de PROMPT peut être enrichi afin d'améliorer la qualité des suggestions réalisées lors de la constitution du vocabulaire partagé. En ce qui concerne les ontologies locales, il est nécessaire de définir des relations d'interprétation diagnostique plus spécifiques. En effet, nous n'avons utilisé que la relation de type «est inféré par». Or, les pathologistes utilisent également d'autres types de relations que l'on pourrait implémenter (« est un critère diagnostique nécessaire et suffisant » ou « est un critère diagnostique optionnel »). Cela permettrait de mieux prendre en compte la sémantique des règles d'interprétation fournies par les classifications et d'améliorer la conformité des cas décrits

A plus long terme, nous envisageons d'intégrer les outils développés à la plate forme collaborative d'interprétation d'images anatomopathologique IDEM destinée à des pathologistes qui n'utilisent pas obligatoirement les mêmes classifications pour déterminer leurs diagnostics. La traduction des descriptions permettrait de quantifier la part de la variabilité diagnostique liée à l'utilisation de règles d'interprétation diagnostique par rapport à la variabilité d'observation de caractéristiques morphologiques dans l'image.

## **Références**

- BACHIMONT B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologie en ingénierie des connaissances, *Eyrolles*, pp. 305-23, 2000.
- BUCCELLA A, CECLICH A. (2003). An Ontology Approach to Data Integration, *JCS&T*; 3(2), 2003; pp 62-68.
- DOAN A, MADHAVAN J, DOMINGOS P, HALEVY A. (2002). Learning to map between ontologies on the semantic web. *In Proceedings of the eleventh international conference on World Wide Web*, pp. 662-673. ACM Press, 2002.
- FALISSARD B. (2001). Mesurer la subjectivité en santé: perspective méthodologique et statistique, Paris, *Masson*, 2001.
- FLEMING KA. (1996). Evidence-based pathology, *J Pathol*, 1996 Jun; 179(2): 127-8.
- FLEMING KA. (2002). Evidence-based cellular pathology, *Lancet*, 2002; 359 (9312): 1149-50.
- GRÜBER TA. (1993). Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition* 5, pp. 199-220, 1993.

- HOVY EH. (1998). Combining and Standardizing Large-Scale, Practical Ontologies for Machine Translation and Other Uses, *Proc. 1st Int'l Conf. Language Resources and Evaluation (LREC)*, European Language Resources Assoc., Paris, 1998, pp. 535-542.
- LIEBERMAN MI, RICCIARDI TN, MASARIE FE, SPACKMAN KA. (2003). The use of SNOMED CT simplifies querying of a clinical data warehouse. *AMIA Annu Symp Proc.* 2003;910.
- MENA E, ILLARRAMENDI A, KASHYAP V, SHETH AP. (2000). OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *Distributed and Parallel Databases*, 8(2); pp. 223-271, 2000.
- MITRA P, WIEDERHOLD G, KERSTEN ML. (2000). A Graph-Oriented Model for Articulation of Ontology Interdependencies. *In Intl. Conference on Extending Database Technology (EDBT)*, pp. 86-100, 2000.
- NOY N, MUSEN MA. (2003). The prompt suite: Interactive tools for ontology merging and mapping. *Journal of Human-Computer Studies*, 59(6); pp. 983-1024, 2003
- NOY N. (2003). Tools for Mapping and Merging Ontologies, Handbook on Ontologies, S.Staab and R. Studer editors, pp. 365-384. *Springer-Verlag*, 2003.
- PREECE A, HUI K, GRAY A, MARTI P, BENCH-CAPON T, JONES D AND CUI Z. (1999). The KRAFT Architecture for Knowledge Fusion and Transformation, in M Bramer, A Macintosh & F Coenen (eds), *Research and Development in Intelligent Systems XVI (Proc ES99)*, Springer, New York, 1999, pp.23-38.
- RIBIÈRE M. (1999). Représentation et gestion de multiples points de vues dans le formalisme des graphes conceptuels, *Thèse de l'université de Sophia-Antipolis*; 1999.
- THIEU M, STEISHEN O, ZAPLETAL E, JAULENT MC ET CHRISTEL LE BOZEC. (2004). Mesures de similarité pour l'aide au consensus en anatomie pathologique. *IC 2004* ; pp. 225-236.
- USCHOLD M, GRÜNINGER M. (1996). Ontologies: Principles, Methods and Applications, *Knowledge Engineering Review*, 11(2), pp. 93-155, 1996.
- WACHE H, VOGELE T, VISSER U, STUCKENSCHMIDT H, SCHUSTER G, NEUMANN H AND HUBNET S. (2001). Ontology-based integration of information - a survey of existing approaches. In Stuckenschmidt, *IJCAI-01 Workshop: Ontologies and Information Sharing*, 2001, pp. 108-117.
- WELLS WA, CARNEY PA, ELIASSEN MS, GROVE MR, TOSTESON AN. (2000). Pathologists' Agreement With Experts and Reproducibility of Breast Ductal Carcinoma-in-Situ Classification Schemes, *Am J Surg Pathol*; 5(24); 2000; pp. 651-670.