

Apport d'une modélisation ontologique pour la détection du signal en pharmacovigilance

Cédric Bousquet^{1,2}, Cornéliu Henegar¹, Agnès Lillo-Le Louët² et Marie-Christine Jaulent¹

¹INSERM ERM202,

{Cedric.Bousquet, Corneliu.Henegar, Marie-Christine.Jaulent}@spim.jussieu.fr

²Centre régional de Pharmacovigilance de l'Hôpital Européen Georges Pompidou
pharmacovigilance.hegp@hop.egp.ap-hop-paris.fr

Résumé : La détection du signal en pharmacovigilance consiste à mettre en évidence toute association potentielle entre un médicament et les effets indésirables qu'il est susceptible de provoquer, cette relation étant inconnue ou insuffisamment documentée. Des méthodes de fouille de données basées sur l'apprentissage statistique permettent d'explorer de façon non supervisée de grandes quantités de données dans les bases nationales de pharmacovigilance. Notre objectif est de développer une méthode de détection du signal qui tient compte de l'information sémantique contenue dans les termes retenus pour le codage des effets indésirables des médicaments (dictionnaire MedDRA). Nous avons modélisé les termes MedDRA au moyen de la logique de description SHIQ dans l'éditeur d'ontologie OIEd. La mise en oeuvre du moteur d'inférence RACER sur les données ainsi traitées permet de regrouper les cas sémantiquement proches. L'intégration de ce raisonnement terminologique aux méthodes classiques de détection du signal améliore significativement les résultats.

Mots-clés : Pharmacovigilance, Modélisation ontologique, MedDRA

1 Introduction

1.1 La détection du signal avec MedDRA

La pharmacovigilance est la discipline ayant pour objet la détection, l'évaluation et la prévention des effets indésirables liés aux médicaments. En France, les cas déclarés par les professionnels de santé sont enregistrés dans la base nationale de pharmacovigilance au moyen de terminologies standardisées. Un signal désigne en pharmacovigilance toute relation inattendue entre un médicament et un effet indésirable. Un signal doit être validé au moyen d'une enquête dont les conclusions conduiront à des mesures correctives telles que le retrait du médicament, la restriction des indications, des lettres d'information aux médecins ou l'ajout de ces conclusions dans le résumé des caractéristiques du produit destiné aux professionnels de santé.

L'augmentation de la charge de travail en pharmacovigilance justifie le développement de méthodes automatisées pour la détection du signal. En pratique, on parlera de signal quand le nombre de cas observé est supérieur au nombre de cas attendu. Le seuil, dont le dépassement détermine l'apparition d'un signal peut être établi au moyen d'un modèle mathématique. Des tests statistiques habituels ont été appliqués aux données de pharmacovigilance tels que Proportionate Reporting Ratio (PRR) (Evans *et al*, 2001), Reporting Odds Ratio (ROR) associé ou non à la régression logistique (Van Puijbroek *et al*, 1999), χ^2 (avec la correction de Yates lorsqu'un des effectifs est inférieur à 5). Par exemple la Medical Control Agency en Grande Bretagne détecte un signal dans sa base quand $PRR > 2$, $\chi^2 > 4$ et le nombre de cas est supérieur ou égal à trois (Evans *et al*, 2001). L'OMS – Organisation Mondiale de la Santé - (Bate *et al*, 1998) et la Food and Drug Agency (FDA) aux Etats Unis (Szarfman *et al*, 2002) proposent la mise en oeuvre des probabilités bayésiennes.

On retrouve par exemple de façon rétrospective une association statistiquement significative entre la dextrofenfluramine et le risque d'hypertension artérielle pulmonaire dans la section HEGP (Hôpital Européen Georges Pompidou) de la base nationale de pharmacovigilance. L'indice « information component (IC) » proposé par l'OMS auquel nous avons appliqué deux écart-types est nettement supérieur à 0 (3,8) ce qui indique que sa valeur est comprise dans l'intervalle de confiance à 95 %. Le risque relatif de développer une hypertension artérielle pulmonaire est de 322 par rapport à un patient qui est traité par un autre médicament.

Les effets indésirables de la base nationale de pharmacovigilance sont actuellement codés au moyen du WHO-ART (World Health Organisation - Adverse Reference Terminology) développé et maintenu par l'OMS. MedDRA (Medical Dictionary for Regulatory Activities) est une nouvelle terminologie qui doit remplacer le dictionnaire WHO-ART pour la description des effets indésirables des médicaments et l'échange d'information par voie électronique (Brown 2002). MedDRA ne dispose pas d'une représentation formelle des concepts et n'est pas vraiment multiaxiale (Bousquet 2004). La granularité des termes proposée par MedDRA par rapport à WHO-ART est susceptible de diminuer la sensibilité de l'ensemble des méthodes d'apprentissage statistique en raison d'un regroupement insuffisant de termes décrivant des conditions cliniques similaires. Par exemple MedDRA dispose d'une quinzaine de termes préférés différents pour décrire une paresthésie. L'impact du choix d'une terminologie pour le codage des effets indésirables sur les résultats des algorithmes de détection du signal n'a encore jamais été évalué de façon quantitative.

1.2 Objectifs

Notre objectif est de construire une ontologie afin de représenter de façon formelle la signification des termes désignant les effets indésirables des médicaments. En effet l'ajout d'une sémantique formelle à MedDRA permettrait d'appliquer des modèles de raisonnement terminologique afin de permettre le regroupement automatique de termes MedDRA décrivant une même condition clinique. L'augmentation du nombre

de cas identifiés lors du regroupement de cas de pharmacovigilance similaires devrait améliorer la sensibilité des méthodes de détection automatisée du signal appliquées sur les cas regroupés.

Nous décrivons dans la deuxième section les principes de construction d'une ontologie, les principaux systèmes terminologiques utilisés en médecine et les limites liées à MedDRA. La troisième section présente les cas de pharmacovigilance extraits de la base nationale, la quatrième décrit la construction d'une ontologie pour définir les termes MedDRA et l'application du raisonnement terminologique à nos données. Les résultats obtenus sont donnés dans la cinquième section. Enfin nous discutons l'intérêt d'une représentation formelle des termes MedDRA ainsi que la charge de travail nécessaire pour décrire l'ensemble de la terminologie.

2 Etat de l'art

Une ontologie est la spécification explicite et formelle d'une conceptualisation partagée, en vue de la réalisation d'une tâche (Bachimont, 2000). Le processus général de modélisation est complexe, mais il s'agit d'abord d'identifier les concepts utilisés dans le domaine pour la tâche, leurs propriétés et les relations qu'ils entretiennent. Après quoi une contrepartie formelle est construite, en faisant correspondre des termes primitifs et leurs attributs aux concepts et à leurs propriétés, et des liens formels entre termes primitifs aux relations entre concepts.

Propulsées par les travaux réalisés autour du Web sémantique, les logiques de descriptions (LD) se retrouvent parmi les langages les plus utilisés pour la représentation de la connaissance et l'opérationnalisation des ontologies (Charlet, 2003). Dans cette famille de formalismes, les connaissances d'un domaine sont représentées à l'aide de « descriptions » qui peuvent être des concepts (classes d'individus), des rôles (relations binaires entre individus) et des individus (instances) (McGuinness, 2001). Au sein de la grande famille des LD une des plus utilisées actuellement est celle appelée SHIQ (Horrocks et al., 2000).

Le raisonnement terminologique en LD (raisonnement sur les descriptions) est basé sur des opérations de classification et d'instanciation sur les hiérarchies de concepts et de rôles. La classification permet de déterminer automatiquement la position d'un concept ou d'un rôle dans leurs hiérarchies respectives, alors que l'instanciation permet de retrouver tous les concepts dont un individu est susceptible d'être une instance.

2.1 Principaux systèmes terminologiques en médecine

Les spécialistes en terminologie appartenant à la communauté d'informatique médicale ont classé les systèmes terminologiques en trois générations (Rossi-Mori *et al.*, 1998).

- Les systèmes de première génération tels que la classification internationale des maladies (CIM) (WHO, 1993) sont présentés sous la forme de manuels

qui peuvent être consultés textuellement. La recherche d'information est contrainte par les rubriques déjà existantes dans la terminologie. Un outil informatique améliore la mise en œuvre de ce type de terminologie mais n'est pas obligatoire lors du codage.

- Les systèmes de deuxième génération tels que SNOMED (Systematized Nomenclature of Medicine) (Coté *et al*, 1993) décrivent les concepts sous forme compositionnelle selon plusieurs axes : topographie, morphologie, fonction, etc. Par exemple, le concept ulcère gastrique hémorragique est décrits au moyen de trois concepts juxtaposés ulcère, hémorragie (axe morphologique) et estomac (axe topographique) mais sans indiquer les relations entre ces trois concepts.
- Les systèmes de troisième génération par exemple GALEN (Generalized Architecture for Languages, Encyclopedias and Nomenclatures) (Rector *et al*, 1995) proposent une représentation formelle des termes dans une ontologie. Les concepts sont décrits au moyen d'un langage formel basé sur les logiques de description. Certains de ces systèmes sont associés à un système de raisonnement terminologique, par exemple dans le serveur GALEN.

UMLS (Unified Medical Language System) (Lindberg *et al*, 1993) présente un statut particulier par rapport à cette classification. Il s'agit d'un réseau sémantique mettant en relation une soixantaine de terminologies utilisées en médecine (CIM, SNOMED, WHO-ART, MedDRA, etc mais pas GALEN qui est un système indépendant). La signification des concepts n'est pas formalisée au moyen d'un langage de description formel mais par les relations que ces concepts entretiennent les uns avec les autres dans le réseau.

2.2 Problèmes liés à MedDRA

MedDRA est un système de première génération comportant cinq niveaux : 26 classes organes ou *System Organ Class* (SOC) – Ex : *Troubles du foie et des voies biliaires*, 332 termes de groupe de haut niveau ou *high level group term* (HLGT) – Ex : *Maladies hépatobiliaires*, 1683 termes de haut niveau ou *high level term* (HLT) – Ex : *Hépatite*, 16 102 termes préférés ou *preferred terms* (PT) – Ex : *Hépatite cytolitique*¹, 56 981 termes de bas niveau ou *low level terms* (LLT). Une classe organe regroupe l'ensemble des concepts liés à un organe, par exemple le système digestif ou le système rénal. La classe organe examens ou investigations réunit les termes décrivant les résultats des tests de laboratoire ou des explorations fonctionnelles.

Chaque terme préféré est associé à une classe organe unique et peut appartenir de façon optionnelle à une ou plusieurs classes organes secondaires. Par exemple la *néphropathie diabétique* appartient à la classe organe des *troubles rénaux* mais il existe un lien secondaire vers la classe organe des *troubles métaboliques*. Aucun lien

¹ La cytolyse hépatique est une destruction des cellules

n'a été défini entre la classe des examens et les pathologies associées dans les autres classes organe. Ainsi, une *élévation des transaminases* et une *hyperbilirubinémie* qui signent la présence d'un trouble hépatique ne sont pas liées au terme préféré hépatite dans la classe organe des troubles du foie et des voies biliaires.

Bien qu'un terme préféré puisse appartenir à plusieurs classes organe, il ne peut pas avoir plus d'un parent à l'intérieur de la même classe organe. Par exemple une *hépatite cholestatique*² est correctement classée dans la catégorie des *cholestases et ictères*. Cependant aucun lien pointant vers la catégorie des *hépatites* n'est présent dans cette classe organe. Cette limite est liée à l'organisation hiérarchique de MedDRA selon des critères taxinomiques. Ce type d'organisation est nécessaire lors de la description de l'arbre des concepts primitifs. Ce n'est pas le cas de MedDRA qui comporte essentiellement des concepts définis. Dans ce cas l'organisation devrait être un treillis de concepts.

3 Matériel

3.1 Cas de pharmacovigilance

Nous avons utilisé 3028 cas de pharmacovigilance correspondant à l'extrait de la base nationale enregistré par le centre régional de l'Hôpital européen Georges Pompidou (HEGP). A partir de ces cas, nous avons identifié 12955 couples {médicament – effet indésirable} parmi lesquels 9646 associations étaient uniques.

3.2 Editeur d'ontologie et moteur d'inférence

Les concepts ont été décrits dans l'éditeur d'ontologie OilEd. Le langage de représentation des connaissances était la logique de description SHIQ (DAML + OIL). Les méthodes statistiques et bayésiennes de détection du signal ont été implémentées dans l'outil Pharmaminer (Henegar et al, 2004) qui intègre le moteur d'inférence RACER pour la mise en œuvre du raisonnement terminologique.

3.3 Sélection des termes MedDRA à modéliser

Après avoir identifié un nombre total de 846 termes préférés WHO-ART employés pour le codage des effets indésirables dans la section HEGP de la base nationale, nous avons sélectionné les 694 termes préférés MedDRA homologues, à l'aide d'une table de correspondances terminologiques WHO-ART – MedDRA. Cette table a été établie en s'appuyant sur l'information existante dans le réseau sémantique et dans les tables de termes synonymes désignés pour chaque concept UMLS.

² La cholestase est une diminution de l'écoulement de la bile dont les éléments refluent dans le sang

4 Méthode

4.1 Modélisation ontologique

4.1.1 Construction d'une hiérarchie de concepts primitifs et de relations

La première étape de la modélisation ontologique a été la construction de la hiérarchie de concepts primitifs. La réutilisation de ressources terminologiques existantes n'a pas été possible. Le réseau sémantique d'UMLS ne propose pas suffisamment de relations entre MedDRA et SNOMED pour envisager une modélisation à partir de l'arbre des primitives contenues dans SNOMED. Le GALEN Core couvre de façon inégale et insuffisante les différents sous-domaines médicaux disponibles dans MedDRA. Dans ces conditions la solution choisie a été de construire une hiérarchie de concepts primitifs inspirée partiellement du modèle offert par la structure de GALEN Core et de la développer ensuite en fonction des besoins de modélisation des concepts complexes MedDRA, au fur et à mesure de l'extension de l'ontologie.

Le tronc de cette hiérarchie, conçue en respectant le principe des taxinomies indépendantes (chaque primitive conceptuelle dérive d'un père unique), est divisé en 2 branches principales. La première branche est celle des concepts primitifs indépendants (« Self-standing Concepts ») comme par exemple inflammation, infection, hémorragie, foie. Cette branche, la plus riche en concepts, est structurée en 4 subdivisions : les niveaux structurels physiques, les fonctions physiologiques, les agents pathogènes et les types d'entités pathologiques. L'autre branche principale contient des primitives conceptuelles permettant de raffiner les descriptions conceptuelles complexes (« Refining Concepts »), exprimant des nuances et des degrés d'intensité par rapport aux propriétés des concepts complexes comme par exemple aigu, chronique, hypo, hyper, normal, anormal, etc.

En parallèle avec la hiérarchie de concepts primitifs nous avons développé une hiérarchie de relations. Cette hiérarchie a été construite sur le même modèle que celle des concepts primitifs, en proposant d'une part des rôles permettant de relier les concepts atomiques primitifs afin de composer des descriptions conceptuelles complexes (« Relating Properties ») et d'autre part des rôles utiles pour le raffinement des descriptions antérieurement réalisées (« Refining Properties »).

4.1.2 Modélisation des concepts définis MedDRA

La modélisation des concepts complexes désignés par les termes MedDRA a été réalisée en parallèle avec l'extension des hiérarchies de concepts primitifs et de relations. Le positionnement des concepts complexes à l'intérieur de la structure ontologique, a été inféré automatiquement en rapport avec les descriptions réalisées, en utilisant le moteur d'inférence RACER. Afin de faciliter le contrôle du résultat de la modélisation ontologique, mais aussi pour permettre une implémentation facile des

techniques de raisonnement terminologique au sein d'un outil logiciel, nous avons individualisé les deux hiérarchies de concepts (primitifs et définis) en interdisant les liens taxinomiques directs « *is a* » entre les deux types de concepts. Par exemple au lieu de définir le concept de « hépatite » comme « *est une inflammation localisée au niveau du foie* » nous l'avons défini comme « *est un trouble structurel de niveaux structurels hépatiques de type inflammation* ».

Hepatitis = (*isStructuralDisorderOf*SOME HepaticStructuralLevels) AND (*hasTypeOfDisorder* SOME InflammationTypes)

Nous avons défini plusieurs règles de modélisation. Le modèle le plus général permettant la description d'une entité pathologique doit préciser dans chaque cas le type de trouble impliqué (structurel ou fonctionnel). S'il s'agit d'un trouble fonctionnel on doit indiquer le type de fonction affectée et la structure physique qui en est responsable. Dans le cadre d'un trouble uniquement structurel il suffit d'indiquer la structure affectée. Les troubles mixtes peuvent être composés par conjonction ou disjonction de troubles fonctionnels et/ou structurels.

Hepatic_enzymes_and_function_abnormalities = (*isStructuralDisorderOf* SOME HepatocytesStructuralLevels) OR (*isFunctionalDisorderOf* some (Functions AND (*isFunctionOf*SOME HepaticStructuralLevels)))

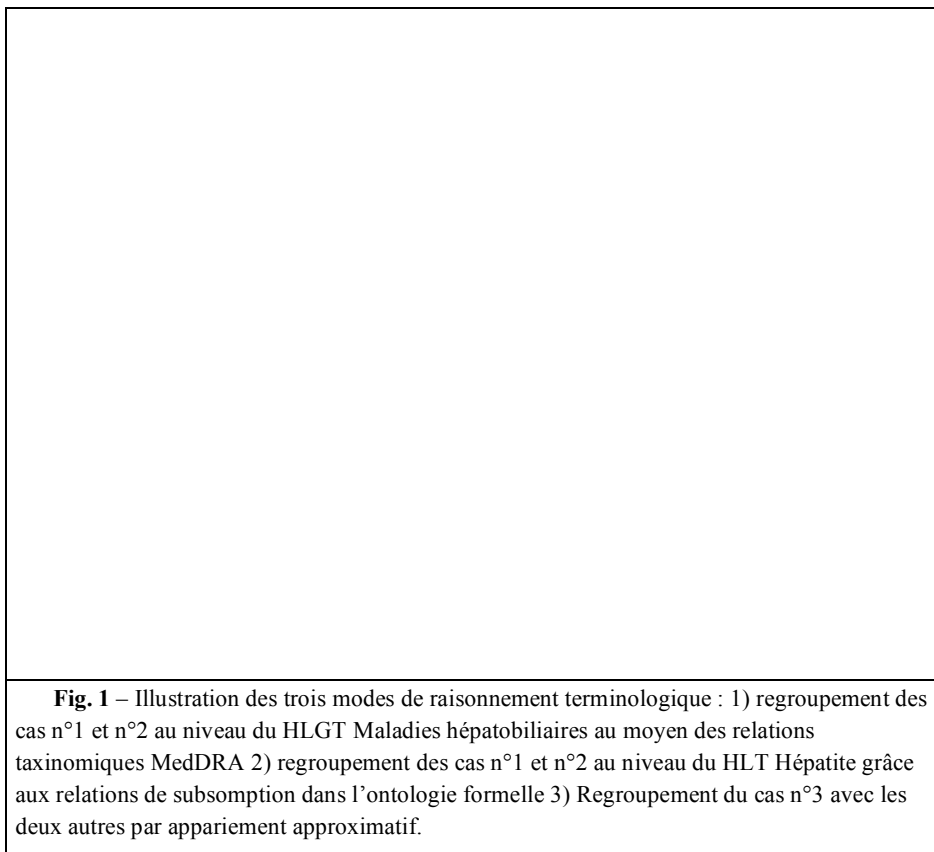
Ensuite les descriptions peuvent être raffinées en précisant le sens d'un trouble (hypo ou hyperfonction par exemple) ou bien en le situant dans un contexte temporel (aigu, chronique, fulminant) et ainsi de suite.

Hepatitis_fulminant = (*isStructuralDisorderOf*SOME HepaticStructuralLevels) AND (*hasTypeOfDisorder* SOME InflammationTypes) AND (*hasEvolution* SOME FulminantEvolutionType)

4.2 Raisonnement terminologique

4.2.1 Subsumption basée sur la hiérarchie taxinomique MedDRA

Le dictionnaire MedDRA comporte des relations hiérarchiques qui permettent de regrouper des termes appartenant à une même rubrique. Nous proposons par exemple dans la figure 1 trois cas de pharmacovigilance avec le même médicament mais des effets indésirables différents. Le cas n°1 est une hépatite cytolytique, le cas n°2 une hépatite cholestatique et le cas n°3 comporte deux effets indésirables qui sont des examens de laboratoire perturbés (augmentation des alanine amino transférases ou ALAT, bilirubine augmentée). Par subsumption dans la hiérarchie MedDRA, les cas n°1 et n°2 sont regroupés au niveau du HLGT *Maladies hépatobiliaires*.



4.2.2 Subsomption sur l'ontologie formelle

La subsomption au moyen de l'ontologie formelle permet de retrouver de façon algorithmique l'ensemble des relations qui existent entre les termes mais qui ne sont pas définies dans MedDRA. Par exemple, les cas n°1 et n°2 sont regroupés au niveau de la catégorie Hépatite qui présente un caractère plus spécifique que le terme générique *Maladies hépatobiliaires*.

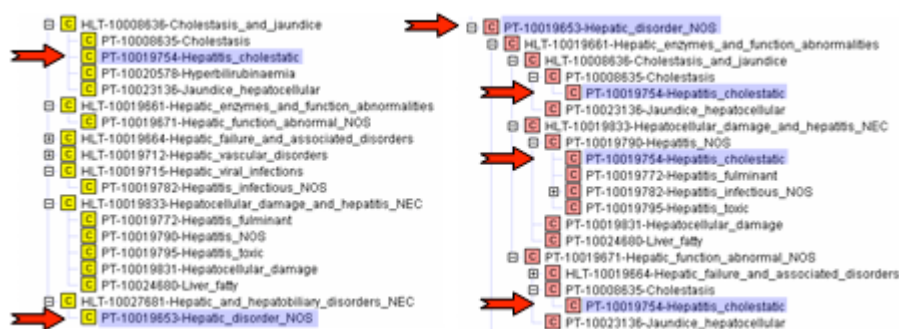
4.2.3 Raisonnement par appariement approximatif

Nous avons traduit le terme anglais *approximate matching* par appariement approximatif. Il s'agit d'un modèle de subsomption étendue qui permet d'inclure des formes incomplètes d'une maladie qui peut être suspectée sur la base des examens de laboratoire. Ce mode de raisonnement est basé sur les relations « *test diagnostique de* ». Dans l'exemple de la figure 1, le cas n°3 qui indique l'existence d'une hyperbilirubinémie et d'une augmentation des ALAT est associé par appariement approximatif aux deux cas d'hépatite n°1 et n°2.

5 Résultats

5.1 Modélisation ontologique

L'ontologie a été réalisée en 300 heures par un médecin – informaticien. Elle comporte 390 concepts primitifs, 530 concepts définis MedDRA et 42 relations. Certaines relations entre concepts qui présentent un intérêt d'un point de vue médical ne sont pas présentes dans MedDRA. La modélisation ontologique permet de les rétablir de façon automatique par simple calcul à partir des descriptions des concepts en logique de description. Par exemple, l'hépatite cholestatique qui était définie dans MedDRA comme appartenant au HLT *Cholestase et ictères* bénéficie dans l'ontologie d'un héritage double sous les concepts *cholestase* et *hépatite* (figure 2). Des concepts qui étaient mal placés dans MedDRA sont mis à une position appropriée dans l'ontologie. Ainsi, le terme *Hepatitis NOS* (Not Otherwise Specified) désigne une hépatite pour laquelle on ne dispose pas de suffisamment d'information pour pouvoir la coder de façon plus spécifique. Cet élément qui était positionné comme un terme préféré dans MedDRA sert maintenant à désigner les hépatites à un niveau plus général dans l'ontologie (figure 2).



couple à l'appariement approximatif, raisonnement sur l'ontologie, raisonnement basé sur la hiérarchie taxinomique MedDRA) et l'application immédiate du test

Information Component de l’OMS. Les courbes ROC (Receiver Operating Curve) étudient le rapport entre le nombre de vrais positifs (sensibilité) et le nombre de faux positifs (1 – spécificité). L’aire sous la courbe est de 0,590 pour Information Component sans regroupement terminologique préalable (table 1). Les meilleurs résultats sont obtenus au moyen d’une approche mixte qui associe le raisonnement terminologique et l’appariement approximatif avec une aire sous la courbe de 0,908. Les résultats obtenus avec les autres modes de raisonnement terminologique sont une aire sous la courbe de 0.890 avec l’ontologie et 0.815 avec les liens disponibles dans MedDRA.

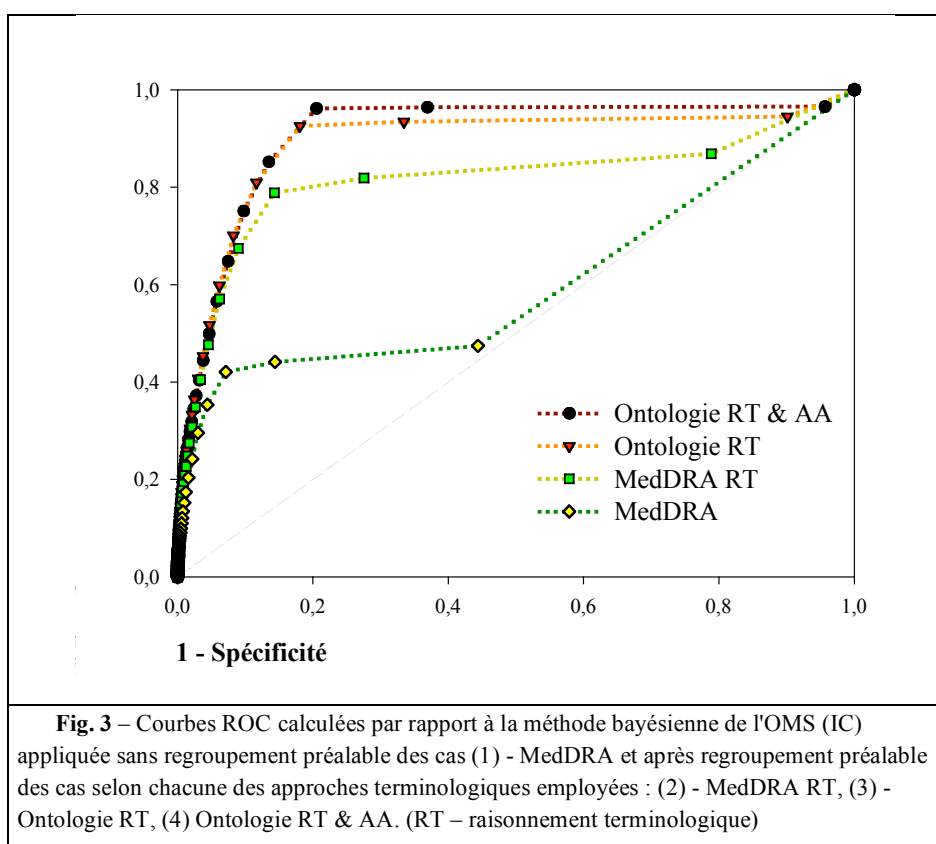


Fig. 3 – Courbes ROC calculées par rapport à la méthode bayésienne de l’OMS (IC) appliquée sans regroupement préalable des cas (1) - MedDRA et après regroupement préalable des cas selon chacune des approches terminologiques employées : (2) - MedDRA RT, (3) - Ontologie RT, (4) Ontologie RT & AA. (RT – raisonnement terminologique)

Table 1. L’aire sous les courbes ROC calculées par rapport à la méthode bayésienne de l’OMS (IC) pour chacune des approches terminologiques employées

Variables	Aire sous la courbe	Intervalle de confiance à 95%	
		Lim. inf.	Lim. sup.
Ontologie RT & AA (OTRAA)	0,908	0,903	0,913

Ontologie RT (OTR)	0,890	0,885	0,896
MedDRA RT (MTR)	0,815	0,807	0,823
MedDRA (M)	0,599	0,590	0,609

6 Discussion et conclusion

Nous avons montré de façon quantitative l'intérêt d'une approche basée sur l'ingénierie des connaissances pour l'exploitation automatisée d'une base de données codée au moyen d'un système terminologique de première génération. En particulier, le raisonnement terminologique améliore les performances des algorithmes statistiques de détection du signal. Les meilleurs résultats sont obtenus par une méthode qui combine le raisonnement terminologique sur l'ontologie et l'application de l'appariement approximatif. On note que la mise en œuvre des liens déjà définis dans le dictionnaire MedDRA permet de réaliser un raisonnement terminologique plus élémentaire. Les performances sont moins bonnes mais la méthode est applicable sans avoir besoin d'une phase de modélisation ontologique préparatoire.

Il est probable que MedDRA continuera à évoluer sous la forme d'une hiérarchie stricte dans les prochaines versions. Bien que le passage à une ontologie apporte de nombreux avantages (cohérence, maintenance, définition formelle des termes), la structure rigide de MedDRA présente des avantages en termes de reproductibilité du codage et de présentation des données. L'héritage unique à l'intérieur d'une classe organe est nécessaire lors de l'exploitation statistique des données afin d'éviter qu'un même concept puisse être compté plusieurs fois dans des catégories différentes.

Nous souhaitons néanmoins continuer la modélisation ontologique de MedDRA. La description de l'ensemble des termes MedDRA (plus de 15 000 termes préférés) nécessite un très gros effort de modélisation si nous souhaitons conserver le niveau de détail choisis pour la représentation des concepts dans le présent travail. Cependant, la difficulté de la modélisation doit être mise en balance avec les avantages que nous apporte l'ontologie. Un plus grand nombre de cas pourraient être identifiés lors de la recherche d'information dans la base nationale de pharmacovigilance. Une ontologie permettrait l'ajout d'outils d'aide au codage basés sur la description conceptuelle des cas de pharmacovigilance.

Le couplage d'un système terminologique de première génération qui a l'avantage d'être plus intuitif avec un système de troisième génération est une piste de développement prometteuse. Par exemple la CCAM (Classification Commune des Actes Médicaux) est une nouvelle terminologie destinée aux codages des actes chirurgicaux qui a été développée au moyen des outils disponibles sur le serveur Galen. Les termes apparaissent sous la forme d'une hiérarchie qui permet un codage efficace par le médecin. Chaque terme bénéficie d'une description formelle en logique de description ce qui permet d'inférer automatiquement des regroupements d'actes similaires selon des critères différents.

Références

- BACHIMONT B. (2000) Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances. *Charlet J, Zacklad M, Kassel G, Bourigault D. Ingénierie des connaissances*. Paris: Eyrolles: 305-23.
- BATE A., LINDQUIST M., EDWARDS I.R., OLSSON S., ORRE R., LANSNER A. & DE FREITAS R.M. (1998). A bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*;54:315-21.
- BOUSQUET C., LAGIER G., LILLO-LE LOUËT A., LE BELLER C., VENOT A. & JAULENT MC. (2004). Appraisal of MedDRA conceptual structure for describing and grouping adverse drug reactions. Soumis le 3 juillet 2003 à la revue *Drug safety*. Accepté sous forme révisée le 9 décembre 2003.
- BROWN EG. (2002). Effects of coding dictionary on signal generation: a consideration of use of MedDRA compared with WHO-ART. *Drug Saf*; 25(6):p 445-52.
- CHARLET J. (2003) L'ingénierie des connaissances développements, résultats et perspectives pour la gestion des connaissances médicales *Mémoire d'habilitation à diriger des recherches. Université Pierre et Marie Curie*.
- CÔTÉ R.A., ROTHWELL D.J., PALOTAY J.L., BECKET R.S. & BROCHU L. (1993). *SNOMED International*. College of American Pathologists.
- EVANS S.J., WALLER P.C. & DAVIS S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*; 10(6):p 483-6.
- HENEGAR C., BOUSQUET C., LILLO-LE LOUËT A., DEGOULET P., JAULENT M.C. (2004). A knowledge based approach for automated signal generation. Article (cinq pages) accepté au congrès *Medinfo* à San Francisco, septembre 2004.
- HORROCKS I, SATTLER U, TOBIES S. (2000) Reasoning with Individuals for the Description Logic SHIQ. *Proceedings of the 17th International Conference on Automated Deduction (CADE-17)*. David MacAllester.
- LINDBERG DA, HUMPHREYS BL, MCCRAY AT. (1993) The Unified Medical Language System. *Methods Inf Med* 1993; 32(4):281-91.
- MCGUINNESS DL. Description Logics Emerge from Ivory Towers. *Description Logics (proceedings)*. 2001.
- RECTOR A.L., SOLOMON W.D., NOWLAN W.A., RUSH T.W., ZANSTRA P.E. & CLAASSEN W.M.A. (1995). A terminology server for medical languages and medical information systems. *Methods Inform Med*;34:p 147-57.
- ROSSI MORI A., CONSORTI F. & GALEAZZI E. (1998). Standards to support development of terminological systems for healthcare telematics. *Methods Inform Med*;37:p 551-63.
- SZARFMAN A., MACHADO S.G. & O'NEILL R.T. (2002). Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports Database. *Drug Saf*;25(6):381-92.
- VAN PUJENBROEK E.P., EGBERTS A.C., HEERDINK E.R. & LEUFKENS H.G. (2000). Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. *Eur J Clin Pharmacol*;56(9-10):733-8.
- WHO (1993). World Health Organisation. *Manual of the international statistical classification of diseases, injuries and causes of death. Tenth revision of the international classification of diseases*. Geneva (1993).