



HAL
open science

Au delà du DSM : les ontologies comme aide aux classifications descriptives psychiatriques ?

Marion Richard, Xavier Aimé, Marie-Odile Krebs, Jean Charlet

► To cite this version:

Marion Richard, Xavier Aimé, Marie-Odile Krebs, Jean Charlet. Au delà du DSM : les ontologies comme aide aux classifications descriptives psychiatriques ?. 2e édition du Symposium sur l'Ingénierie de l'Information Médicale, Jul 2013, Lille, France. hal-00840700

HAL Id: hal-00840700

<https://hal.science/hal-00840700>

Submitted on 9 Jul 2013

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.

Au delà du DSM : les ontologies comme aide aux classifications descriptives psychiatriques ?

Marion Richard¹, Xavier Aimé¹, Marie-Odile Krebs², Jean Charlet¹³

¹ Equipe Ingénierie des Connaissances et Santé (INSERM UMR 872 équipe 20)
marion.richard75@gmail.com

² Laboratoire de Pathophysiologie des Troubles Psychiatriques, Centre Hospitalier Sainte-Anne
mo.krebds@ch-sainte-anne.fr

³ Assistance Publique - Hôpitaux de Paris DRCD

Résumé : La caractérisation des pathologies en psychiatrie pose généralement plus de difficultés que dans les autres domaines de la médecine. Les consensus au sujet des critères définissant les pathologies mentales restent minimaux. Dans ce contexte, notre but est de construire une ressource terminologique et ontologique permettant de caractériser les patients résistant aux traitements en particulier dans le contexte de la schizophrénie.

Pour cela, nous utiliserons une méthodologie maintenant éprouvée de construction de ressource terminologique et ontologique à partir de ressources textuelles. Cette méthodologie sera complétée par les éléments permettant de préciser les déterminants des résistances aux traitements (référentiels structurés du domaine, médicaments, . . .). Les corpus textuels seront élaborés à partir des comptes rendus d'hospitalisation issus du centre hospitalier Sainte-Anne, les ressources structurées seront la CIM-10, le DSM et, pour les médicaments, l'ATC.

Mots-clés : Ingénierie des connaissances, ontologie, psychiatrie.

1 Introduction

La caractérisation des pathologies en psychiatrie pose généralement plus de difficultés que dans les autres domaines de la médecine. Les consensus au sujet des critères définissant les pathologies mentales restent minimaux. Ils sont souvent soumis à des variations en fonction des courants de pensées du moment, mêlant connaissances historiques, observations cliniques et référentiels théoriques. Des nomenclatures et classifications ont été développées dont les deux principales sont la *Classification statistique*

Internationale des Maladies et des problèmes de santé connexes (CIM - World Health Organization (1992)) dont le Chapitre 05 – « Troubles mentaux et du comportement » – comporte 930 codes et le *Diagnostic and Statistical Manual of Mental Disorders* (DSM - American Psychiatric Association (2000)) qui se réfère à la CIM. La première est élaborée par l'Organisation Mondiale de la Santé (OMS). Elle est utilisée principalement pour le codage PMSI¹. La seconde est une classification de l'Association Psychiatrique Américaine (APA). Si ces deux classifications sont largement répandues, elles laissent cependant le médecin souvent insatisfait. En effet, elles ne rendent compte que d'une partie de la réalité clinique, laissant de nombreux patients « hors critères ». De plus, elles ne reprennent pas certaines évolutions du savoir psychiatrique ou, au contraire, oublient certains fondamentaux pour les redécouvrir ultérieurement. Les auteurs du DSM promettent une réactivité plus grande pour la cinquième version qui vient de sortir en juin 2013 mais les changements sont inégaux en fonction des sections et globalement décevants.

Les comptes rendus hospitaliers (CRH) psychiatriques présentent, aujourd'hui, une assez forte hétérogénéité de contenu en fonction, ou non, de l'utilisation de ces classifications et des pratiques locales. Au delà, ces difficultés de délimitation des syndromes compliquent la définition des populations incluses dans les essais thérapeutiques et les recherches étiologiques ou de biomarqueurs. Les catégories actuelles correspondent encore à des situations hétérogènes. En termes médico-économiques, la résistance aux traitements est une caractéristique au moins aussi importante que la catégorie du diagnostic principal. Il est enfin à noter que la notion de comorbidité, donnée clinique essentielle, est souvent absente des codages de dossiers qui se limitent la plupart du temps à un item issu de la CIM-10².

L'objectif premier de nos travaux, situés à mi-chemin entre l'Informa-

1. Le Programme de médicalisation du système d'information (PMSI), vise à introduire des concepts de comptabilité analytique dans la gestion administrative des hôpitaux : les diagnostics et actes effectués dans un établissement de santé sont codés et comptabilisés, rapportés à un patient et aux différents coûts de la structure. Cela permet ainsi de bâtir des indices de coûts relatifs par groupe homogène de malades. Le PMSI utilise un système de codage international, la CIM-10, pour les diagnostics, et un système français, développé grâce à une approche ontologique, la CCAM, pour les actes. Le codage des diagnostics se fait en posant un diagnostic principal et, si nécessaire – au maximum 5 –, des diagnostics associés. Le PMSI a évolué vers une comptabilité qui vise à analyser le coût de chaque acte : c'est la tarification à l'activité ou T2A mais elle ne concerne pas la psychiatrie.

2. Cette limite n'est pas inhérente au système PMSI qui permet d'associer plusieurs codes à un malade (*cf.* note 1) mais est liée aux systèmes internes de l'hôpital.

tique et la Médecine, est de contribuer au développement d'un consensus sur les catégories descriptives des troubles psychiatriques à travers les outils et méthodes de l'Ingénierie des connaissances, en premier lieu le développement d'une ontologie de la psychiatrie. Les ontologies sont des modèles informatiques qui permettent de représenter le sens des concepts d'un domaine dans un formalisme logique. Les concepts du domaine et leurs relations sémantiques y sont organisés de manière à pouvoir être interprétés et manipulés par des algorithmes de raisonnement automatique. Les ontologies connaissent actuellement un important développement avec l'essor du Web sémantique ou du Web des données. Dans le domaine de la médecine, de nombreuses ontologies ont été réalisées : par exemple, *Gene Ontology* en génétique (Ashburner *et al.*, 2000), *Foundational Model of Anatomy* en anatomie (Rosse *et al.*, 2003) ou, dans notre équipe, *OntoMénélas*³ ou *ONTOORPHA*⁴.

L'objectif *in fine* de l'ontologie que nous voulons développer est de pouvoir répondre à au moins deux problématiques rencontrées par les praticiens du centre hospitalier Sainte-Anne : (1) apporter un outil consensuel permettant de mieux rendre compte des diagnostics posés et des actes pratiqués, et ainsi (2) mieux indexer les dossiers de patients en vue de recherche clinique par exemple.

La suite de cet article est structurée comme suit. La section 2 présente le projet et ses enjeux. La section 3 détaille la méthodologie adoptée et la section 4 conclue la présentation de ce travail.

2 Description du projet

2.1 Le centre hospitalier Sainte-Anne

Le centre hospitalier Sainte Anne est un établissement psychiatrique situé dans le XIV^e arrondissement de Paris. Il s'agit d'un hôpital de secteur accueillant les patients de cinq arrondissements parisiens (5, 6, 14, 15 et 16). Il dispose d'un service hospitalo-universitaire (SHU), d'une clinique des maladies mentales et de l'encéphale (CMME), d'un service médico-psychologique régional (SMPR), d'un service spécialisé en addictologie et d'un service d'appui santé mentale et exclusion sociale.

3. <http://pertomed.spim.jussieu.fr/~onto/ontologies/public/OntoMenelas/> (Zweigenbaum *et al.*, 1995).

4. <http://bioportal.bioontology.org/ontologies/1586> (Dhombres *et al.*, 2011).

Depuis plus de dix ans, le SHU du centre hospitalier Sainte-Anne réalise des comptes rendus semi-standardisés pour l'ensemble des patients hospitalisés au titre du secteur (psychiatrie générale couvrant l'ensemble des pathologies) et de sa mission hospitalo-universitaire. Cette dernière représente une psychiatrie de recours tournée plus particulièrement vers les pathologies résistantes associées à des troubles somatiques et vers les difficultés diagnostiques chez le sujet âgé ou chez le sujet jeune.

2.2 Enjeux

Il s'agit, dans un premier temps, de construire une ontologie du domaine de la psychiatrie à partir *a)* des différentes nomenclatures utilisées depuis dix ans pour annoter les dossiers patients, mais également *b)* d'un traitement de type Traitement Automatique de la Langue (TAL) sur un ensemble de dossiers patients numérisés et anonymisés au préalable. La granularité de l'ontologie devra être réglée selon les branches par rapport aux concepts qui s'avèreront utiles pour établir des profils de patients résistants.

Une évaluation du modèle ontologique développée sera réalisée par son opérationnalisation au sein d'applications dédiées au partage de l'information en psychiatrie. Ces applications pourront fournir les services suivants :

- indexation (semi-)automatique de comptes rendus et de dossiers patients ;
- (semi-)automatisation du codage médico-économique (PMSI) ;
- assistance à la recherche d'informations sur les cas de résistance pour l'ensemble des pathologies, avec un focus sur la dépression et la schizophrénie.

Une partie du projet sera également consacrée aux aspects éthiques et légaux, spécialement importants en psychiatrie.

2.3 Approche choisie

Afin de permettre une meilleure appropriation de l'ontologie développée, Aimé & Charlet (2013) établissent une liste de recommandations à suivre en préambule de toute construction d'ontologies. Ces recommandations, à destination de l'ontologue, visent à prendre en compte le fonctionnement de l'écosystème dans lequel l'ontologie va se développer puis être utilisée. L'objectif est de rendre l'ontologie plus ergonomique, dans le sens d'une meilleure adéquation système d'information–utilisateur.

Choix des experts

Derrière cette question se cache le problème des sources choisies et des personnes associées au projet. Premièrement, il n'est pas possible, pour des raisons évidentes de disponibilités du personnel soignant, d'associer tous les utilisateurs à la démarche de construction. Néanmoins, dans le choix des sources (cf. section 3.1), nous privilégions une relative exhaustivité quant à leurs auteurs de manière à pouvoir retrouver les points de vue d'une majorité des praticiens. Secondement, en ce qui concerne la validation itérative de la conceptualisation, elle devrait se faire de manière semi-directive et participative. Par conséquent et dans un premier temps, le choix des experts n'est pas un problème. Nous allons donc tendre vers une solution de type semi-normative.

Rôle des ontologues

Les ontologues qui vont travailler sur ce sujet (1) ont des connaissances dans le domaine de la santé mentale et (2) ont déjà travaillé (hors ontologie) au sein de cette structure psychiatrique. Cela devrait permettre de minimiser leur influence extérieure, mais également une meilleure intégration au sein de cet écosystème pour pouvoir recueillir les informations nécessaires sans les biaiser.

Inclusion de normes

Deux types de normes sont susceptibles d'être prises en compte pour le développement de cette ontologie. Il y a en premier lieu les normes de la profession, les terminologies, que les praticiens se sont appropriées : la CIM-10, le DSM, ... Ces ressources seront à utiliser soit dans la structuration soit à fin d'alignements. Il y a en second lieu les normes liées à la notion même d'ontologie. Elle sera développée en OWL 2.0. La question est de savoir s'il est opportun d'utiliser une top-ontologie pour structurer les concepts de haut-niveau, en sachant que ce choix influence l'organisation même de l'ontologie. Nous prenons le parti, dans un premier temps, de ne pas choisir de top-ontologie, d'autant que nous avons déjà fait l'expérience, en cours de développement, de mettre une ontologie de domaine sous une top-ontologie et que c'est un travail de quelques heures. Les choses se compliquent au niveau des relations dans la mesure où, avec une top-ontologie et une core-ontologie, peuvent être associées des relations. Notre choix n'est pas de minimiser cette problématique mais plutôt de reporter les décisions de ce type à plus tard quand la partie taxinomique

– *i.e.* liée aux concepts – sera stabilisée.

3 Méthodologie mise en œuvre

3.1 Dossiers sélectionnés

Le corpus à étudier est composé de CRH (pseudo-standardisés et sous format Word) sur une période couvrant dix années. Le diagnostic est donné en fin de CRH depuis quelques années, ainsi que la cotation selon la CIM-10. Parallèlement, pour chaque patient, l'activité est recueillie par le département d'Informatique Médicale (DIM) selon le codage EDGAR (Entretien, Démarche, Groupe, Accompagnement, Réunion).

3.2 Anonymisation

La loi 78-17 du 6 janvier 1978, modifiée le 6 août 2004 encadre la circulation des données dites sensibles, dont font parties les données de santé. Ainsi, pour sortir les CRH de l'enceinte de l'hôpital Sainte Anne, il nous est indispensable de procéder à leur anonymisation. Cette opération devra être effectuée sur site et sous contrôle de l'équipe médicale impliquée dans le projet. Afin de permettre une réutilisation par les praticiens des dossiers traités par nos soins, une clé sécurisée permettant le lien entre un patient et un dossier devra également être générée. Le chaînage des données sera sécurisé via l'utilisation de méthodes de hachages irréversibles ainsi que d'un cryptage des tables contenant les informations relatives au chaînage (Quantin *et al.*, 2005; Goldberg *et al.*, 2008; Kalam *et al.*, 2004).

Les recherches dans le domaine médicale étant croissantes, les besoins en anonymisation de CRH poussent au développement de méthodes automatiques d'anonymisation. Pour l'anglais, Uzuner *et al.* (2007) a réalisé une étude de 200 publications sur des outils d'anonymisation. Une liste exhaustive permet d'identifier les résultats des outils développés et de discuter des méthodes utilisées. De plus, les États-Unis ont renforcé la loi⁵ en matière de protection des données en répertoriant 18 identifiants à supprimer des CRH.

5. La loi américaine HIPAA (Health Insurance Portability and Accountability Act) regroupe une liste officiel d'identifiants (PHI) à supprimer : noms, prénoms, lieux, dates, âge (si + de 89), téléphone, télécopie, adresse mail, numéro de sécurité social, enregistrement médical, complémentaire santé, compte bancaire, carte d'identité, et de permis de conduire, références sur le véhicule, URLs, adresse IP, numéro de série ou identifiant d'appareil implémenté, identifiant biométriques, etc.

Pour le français, le projet ALADIN (Gicquel *et al.*, 2012) a développé un outil pour anonymiser les CRH nécessaires à leur étude, mais les résultats n'ont pas permis d'en envisager un usage automatique. En revanche, le logiciel MEDINA (Grouin *et al.*, 2009), développé dans le cadre de la thèse de l'auteur, permet d'anonymiser des dossiers avec un taux de précision d'environ 91% et un taux de rappel d'environ 85%. Ce logiciel utilise des expressions régulières pour anonymiser les entités numériques et des dictionnaires et listes d'entités nommées pour anonymiser les noms propres. Une première anonymisation est réalisée sur le corpus et une deuxième sur le voisinage des termes anonymisés. Ce double traitement permet de repérer les noms qui ne seraient pas dans les dictionnaires. Ainsi le premier tour va permettre d'annoter les prénoms via le dictionnaire et permettra au deuxième tour d'annoter les noms qui les suivent ou les précèdent⁶. Cet outil permet également de conserver la chronologie des événements médicaux ainsi que les intervalles de temps en établissant un décalage pour garantir l'anonymisation des dates. Les possibilités offertes par MEDINA et les résultats l'accompagnant nous ont conforté dans le choix de travailler avec cet outil afin d'anonymiser les données nominatives contenues dans les CRH : noms, prénoms, adresses y compris codes postaux et villes, dates, noms d'hôpitaux, numéros de sécurité social et de téléphone.

3.3 Extraction des termes

L'extraction de termes dans des textes spécialisés s'organise principalement autour de deux approches : (1) une approche fondée sur des thésaurus et les outils de balisage et d'extraction associés, et (2) une approche fondée sur des extracteurs de termes candidats.

Pour la première approche, le méta-thésaurus UMLS (Bodenreider, 2004) fournit de nombreuses ressources pour l'extraction de termes médicaux selon des domaines spécialisés. Actuellement, 145 thésaurus⁷ sont disponibles dans 20 langues dont l'anglais, le français, l'allemand, le japonais et le russe. En France, différents projets se développent tel que l'UMLF (Zweigenbaum *et al.*, 2005), puis INTERSTIS (Cartoni & Zweigenbaum, 2010), ou encore le CISMEF (Darmoni & Joubert, 2000). Ils visent à définir, pour le français, des ressources aussi complètes que celles contenues dans l'UMLS. Ces méthodes s'appuient notamment sur des outils de re-

6. Une autre possibilité pas mise en œuvre par manque de temps, serait de construire un dictionnaire de noms propres interne au SHU.

7. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html>

pérage permettant de retrouver dans les textes les termes contenus dans les thésaurus.

La seconde méthode s'appuie sur l'extraction de termes candidats en corpus spécialisés. Les outils développés utilisent principalement des techniques du TAL telles que les analyses syntaxiques (permettant de reconnaître les phrases correspondant à la syntaxe d'une langue), les annotations morpho-syntaxiques (attribuant à chaque mot d'un texte son étiquette grammaticale), et les méthodes de statistiques linguistiques afin d'obtenir une liste de termes candidats à valider manuellement. Parmi les outils utilisant ces méthodes, on peut citer YATEA (Aubin & Hamon, 2006; Hamon, 2012), SYNTAX-UPERY (Bourigault & Lame, 2002), TTC TERMSUITE (Rocheteau & Daille, 2011) – anciennement ACABIT – ou THERMOSTAT (Drouin, 2003). Le logiciel THERMOSTAT a la particularité de s'appuyer sur la mise en opposition d'un corpus spécialisé et non spécialisé. TTC TERMSUITE utilise des corpus bilingues comparables afin d'aligner les termes spécialisés. SYNTAX-UPERY se compose de deux modules : SYNTAX qui crée un réseau de dépendances entre les mots et les syntagmes. Et UPERY qui va permettre de rapprocher les termes du réseau et leurs contextes syntaxiques via des mesures de proximité distributionnelle. Enfin le logiciel YATEA, développé dans le cadre du projet ALVIS, permet d'identifier des groupes nominaux qui peuvent correspondre à des termes spécialisés d'un corpus. Il fournit une analyse syntaxique dans un fichier xml sous forme d'une décomposition en tête et modifieur. L'extraction des termes est réalisée avec des patrons d'analyse simple. Une désambiguïsation endogène est réalisée au préalable, puis des mesures de pondération statistique permettent de discriminer les termes candidats. Il est à noter que YATEA prend en entrée des données étiquetées morphologiquement via l'annotateur TREETAGGER⁸ pour le français. Alecu *et al.* (2012) met en avant la capacité de ces logiciels à fournir des résultats complémentaires et préconise donc une utilisation « multi-extractions » pour garantir de meilleurs résultats.

Nos travaux ayant pour objectif la construction d'une ressource termino-ontologique⁹ (RTO), les méthodes d'extraction fondées sur des thésaurus ne répondaient pas à nos besoins. Nous avons donc opté pour une méthode par extraction de termes. Notre choix s'est porté sur YATEA. Cependant l'annotateur TREETAGGER donnait de très mauvais résultats sur le corpus

8. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

9. Nous entendons par là, une ressource combinant concepts et termes et dont les structures associées peuvent être développées séparément (Aussenac-Gilles, 2005).

d'étude. Nous avons donc opté pour l'annotateur MELT (Denis & Sagot, 2012) qui offre un meilleur taux d'annotation sur notre corpus spécialisé. Pour ensuite utiliser le corpus annoté par MELT un convertisseur de formats et d'étiquettes a été développé. Enfin notre choix devait respecter deux contraintes : l'outil devait s'utiliser conjointement avec TERMINAE (Després *et al.*, 2007; Aussenac-Gilles *et al.*, 2008) pour la construction d'ontologie et devait pouvoir s'installer sur une machine que nous maîtrisons, à l'opposé d'un traitement sur serveur, afin de garantir l'anonymat des données.

3.4 La construction de la RTO

Comme dit au début, cette RTO devra inclure les concepts pertinents pour la prise en charge puis l'analyse clinique de patients résistants aux traitements des pathologies psychiatriques (*cf.* section 2.2).

Au regard du faible développement des RTO nécessaires, le cœur de notre travail sera d'analyser les corpus textuels de CRH pour expliciter les concepts du domaine. Mais nous savons déjà que ce travail devra être complété dans au moins 3 directions :

- vérification du fait que les termes de la CIM-10 dans son versant psychiatrique devront être analysés précisément pour, *a priori*, les représenter ;
- prise en compte de la DSM-IV de façon différentielle par rapport à la CIM-10 ;
- prise en compte des médicaments. Les traitements et les médicaments faisant évidemment partis des traitements devront être modélisés. Pour les médicaments, nous partirons de l'ATC¹⁰.

Ce travail impliquant de nombreux acteurs et des sources différentes – textuelles *vs* structurées – il sera proche, en terme de ressources et de méthodologies, de celui mis en œuvre dans le cadre du projet LeRUDI (Charlet *et al.*, 2012).

4 Perspectives et conclusion

La méthodologie de travail pour construire la RTO a été *a priori* stabilisée – même si elle pourra évoluer en fonction des circonstances – et les

10. Le Système de Classification Anatomique, Thérapeutique et Chimique (ATC) est utilisé pour classer les médicaments. C'est le Collaborating Centre for Drug Statistics Methodology de l'OMS qui le contrôle <http://www.whocc.no/atc_ddd_index/>.

choix des outils nécessaires à l'analyse de corpus et à la construction de la RTO faits.

A partir de là, le travail a été jusqu'ici de vérifier la disponibilité des corpus, préciser les contraintes liées au domaine médical (déclarations CNIL, etc.) connues mais toujours un peu différentes d'un cas à l'autre et, enfin, faire le choix des outils d'anonymisation et s'assurer pratiquement et légalement de leur disponibilité.

La méthodologie développée ici n'est pas originale, en particulier par rapport à ce qui se fait dans notre équipe. Ce qui est nouveau, c'est le domaine que l'on veut modéliser, la psychiatrie, qui a plusieurs écoles de pensées. Cela nous amènera à affiner nos méthodologies de construction d'ontologies et surtout de comparaison par rapport à un domaine proche – *i.e.* une autre école de pensée. Par ailleurs, en même temps que les contraintes de la CNIL se renforcent, notre expertise pour organiser le processus pour satisfaire ces contraintes, dans une collaboration entre informaticiens et cliniciens, se renforcent.

Enfin, il existe peu d'initiatives de part le monde sur le sujet. Citons parmi elles les travaux de Smith & Ceusters sur une ontologie des maladies mentales fondée entre autres sur la DSM-IV (Ceusters *et al.*, 2010). Ils utilisent Basic Formal Ontology (BFO) comme ontologie fondationnelle et Ontology of General Medical Science (OGMS) comme ontologie noyau.

Références

- AIMÉ X. & CHARLET J. (2013). IC : Ingénierie des Connaissances ou Ingénierie du Conformisme ? In *IC 2013 : Actes des 25es Journées Francophones d'Ingénierie des Connaissances IC 2013 (à paraître)*.
- ALECU B. P., THOMAS I. & RENAHY J. (2012). La "multi-extraction" comme stratégie d'acquisition optimisée de ressources terminologiques et non terminologiques. In *Actes de la 19e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2012)*, p. 511–518, Grenoble, France.
- AMERICAN PSYCHIATRIC ASSOCIATION (2000). *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*. Washington, DC : American Psychiatric Publishing, 4th edition.
- ASHBURNER M., BALL C., BLAKE J., BOTSTEIN D., BUTLER H., CHERRY J., DAVIS A., DOLINSKI K., DWIGHT S., EPPIG J. *et al.* (2000). Gene ontology : tool for the unification of biology. *Nature genetics*, **25**(1), 25–29.
- AUBIN S. & HAMON T. (2006). Improving term extraction with terminological resources. In *Advances in Natural Language Processing*, p. 380–387. Springer.

- AUSSENAC-GILLES N. (2005). *Méthodes ascendantes pour l'ingénierie des connaissances*. Habilitation à diriger des recherches, Université Paul-Sabatier, Toulouse.
- AUSSENAC-GILLES N., DESPRES S. & SZULMAN S. (2008). The terminae method and platform for ontology engineering from texts. In *Proceedings of the 2008 Conference on Ontology Learning and Population : Bridging the Gap Between Text and Knowledge*, p. 199–223, Amsterdam, The Netherlands, The Netherlands : IOS Press.
- BODENREIDER O. (2004). The unified medical language system (umls) : integrating biomedical terminology. *Nucleic acids research*, **32**(suppl 1), D267–D270.
- BOURIGAULT D. & LAME G. (2002). Analyse distributionnelle et structuration de terminologie. application à la construction d'une ontologie documentaire du droit. *Traitement automatique des langues*, **43**(1).
- CARTONI B. & ZWEIGENBAUM P. (2010). Semi-automated extension of a specialized medical lexicon for french. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta.
- CEUSTERS W., SMITH B. *et al.* (2010). Foundations for a realist ontology of mental disease. *Journal of biomedical semantics*, **1**(1), 10.
- CHARLET J., DECLERCK G., DHOMBRES F., GAYET P., MIROUX P. & VANDENBUSSCHE P.-Y. (2012). Construire une ontologie médicale pour la recherche d'information : problématiques terminologiques et de modélisation. In S. SZULMAN, Ed., *Actes du 23th French Knowledge Engineering Workshop*, p. 33–48, Paris, France.
- DARMONI S. & JOUBERT M. (2000). Cismef. *Methods of information in medicine*, **39**(1), 30–35.
- DENIS P. & SAGOT B. (2012). Coupling an annotated corpus and a lexicon for state-of-the-art pos tagging. *Language resources and evaluation*, **46**(4), 721–736.
- DESPRÉS S., FÜRST F., SZULMAN S. *et al.* (2007). Construction d'une ontologie du domaine hse. *18es Journées Francophones d'Ingénierie des Connaissances*, p. 133–144.
- DHOMBRES F., VANDENBUSSCHE P., RATH A., OLRYS A., HANAUER M., URBERO B. & CHARLET J. (2011). OntoOrpha : an ontology to support edition and audit of rare diseases knowledge in orphanet. In *Proceedings of the International Conference on Biomedical Ontology*, p. 241–3, Buffalo, NY, USA : Barry Smith, University at Buffalo, Buffalo, NY, USA. [poster].
- DROUIN P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, **9**(1), 99–115.
- GICQUEL Q., PROUX D., MARCHAL P., HAGÉGE C., BERROUANE Y., DARMONI S. J., PEREIRA S., SEGOND F. & METZGER M. (2012). Évaluation d'un outil d'aide à l'anonymisation des documents médicaux basé sur le traite-

- ment automatique du langage naturel. In *Systèmes d'information pour l'amélioration de la qualité en santé*, p. 165–176. Springer.
- GOLDBERG M., QUANTIN C., GUÉGUEN A. & ZINS M. (2008). Bases de données médicoadministratives et épidémiologie : intérêts et limites. *Courrier des Statistiques (124)*.
- GROUIN C., ROSIER A., DAMERON O. & ZWEIGENBAUM P. (2009). Une procédure d'anonymisation à deux niveaux pour créer un corpus de comptes rendus hospitaliers. In *Risques, technologies de l'information pour les pratiques médicales*, p. 23–34. Springer.
- HAMON T. (2012). Acquisition terminologique pour identifier les mots clés d'articles scientifiques. *Actes du huitième DÉfi Fouille de Textes*, p.28.
- KALAM A. A. E., DESWARTE Y., TROUÉSSIN G. & CORDONNIER E. (2004). Une démarche méthodologique pour l'anonymisation de données personnelles sensibles. In *Actes du symposium SSTIC04*.
- QUANTIN C., GOUYON B., ALLAERT F. A. & COHEN O. (2005). Méthodologie pour le chaînage de données sensibles tout en respectant l'anonymat : application au suivi des informations médicales. *Courrier des statistiques*, **2005**, 113–114.
- ROCHETEAU J. & DAILLE B. (2011). Ttc termsuite. *Proceedings of System Demonstrations*, p.9.
- ROSSE C., JR J. L. M. *et al.* (2003). A reference ontology for biomedical informatics : the foundational model of anatomy. *Journal of biomedical informatics*, **36**(6), 478–500.
- UZUNER O., LUO Y. & SZOLOVITS P. (2007). Evaluating the state-of-the-art in automatic de-identification. *Journal of the American Medical Informatics Association*, **14**(5), 550–563.
- WORLD HEALTH ORGANIZATION (1992). *The ICD-10 classification of mental and behavioural disorders : clinical descriptions and diagnostic guidelines*. World Health Organization.
- ZWEIGENBAUM P., BACHIMONT B., BOUAUD J., CHARLET J. & BOISVIEUX J.-F. (1995). Issues in the structuring and acquisition of an ontology for medical language understanding. *Methods of Information in Medicine*, **34**(1/2).
- ZWEIGENBAUM P., BAUD R., BURGUN A., NAMER F., JARROUSSE E., GRABAR N., RUCH P., DUFF F. L., FORGET J. F., DOUYERE M. *et al.* (2005). Umlf : a unified medical lexicon for french. *International Journal of Medical Informatics*, **74**(2), 119–124.