HAL open science

Choix méthodologiques pour la construction d'une ontologie de domaine en médecine prénatale

Ferdinand Dhombres, Jean-Marie Jouannic, Marie-Christine Jaulent, Jean Charlet

To cite this version:

Ferdinand Dhombres, Jean-Marie Jouannic, Marie-Christine Jaulent, Jean Charlet. Choix méthodologiques pour la construction d'une ontologie de domaine en médecine prénatale. 21èmes Journées Francophones d'Ingénierie des Connaissances, Jun 2010, Nîmes, France. pp.171-182. hal-00487736

HAL Id: hal-00487736

https://hal.science/hal-00487736v1

Submitted on 30 May 2010

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.

Choix méthodologiques pour la construction d'une ontologie de domaine en médecine prénatale

Ferdinand Dhombres[1,2], Jean-Marie Jouannic[2], Marie-Christine Jaulent[1] et Jean Charlet[1,3]

[1] INSERM UMR_S 872, Éq. 20, Paris, France

[2] Service de Gynécologie-Obstétrique et Centre de Diagnostic Prénatal de l'Est Parisien, Hôpital Armand Trousseau, AP-HP, Paris, France

[3] AP-HP, Paris, France

ferdinand.dhombres@upmc.fr

Résumé : Nous décrivons une méthode de construction d'ontologie d'interface en médecine prénatale avec une double approche d'expert du domaine et d'ingénieur de connaissance ; notre stratégie est à la fois *bottom-up* à partir de l'analyse des documents du domaine (SYNTEXUPERY, ARCHONTE), et *top-down* par réutilisation de ressources terminologiques et ontologiques existantes (ONTOMÉNÉLAS, FMA, ORPHAnet, CCAM, CIM-10).

Mots-clés : Ingénierie des connaissances, ontologie de domaine, méthode de construction d'ontologie, médecine prénatale, anatomie fœtale.

1 Introduction

Il existe de nombreuses méthodes de construction d'ontologie. Nous présentons ici nos choix pour chacune des étapes de la construction d'une ontologie médicale dans le domaine du diagnostic prénatal.

Notre méthodologie se fonde sur la méthode ARCHONTE (ARCHitecture for ONTological Elaborating) élaborée par Bachimont *et al.* (2002) : c'est une méthode ascendante de construction d'ontologie à partir des textes du domaine en trois étapes :

1. choix des termes pertinents du domaine, normalisation sémantique précisant les relations de similarités et de différences que chacun des concepts entretient avec ses frères et son père (principes de la *sémantique différentielle*) ;
2. formalisation des connaissances ou *engagement ontologique* : construction d'une *ontologie référentielle* et ajout des propriétés et annotations, définition des domaines et codomaines des relations ;
3. opérationnalisation de l'ontologie dans un langage de représentation des connaissances. On parle alors d'*ontologie computationnelle*.

Dans notre cas, l'éditeur d'ontologies PROTÉGÉ [1] est utilisé pour formaliser l'ontologie. Le *plugin* spécifique du même nom [2], ARCHONTE pour PROTÉGÉ permet l'implémentation des principes de l'approche différentielle, en définissant les similarités entre les frères, les similarités entre père et fils, les différences entre les frères et les différences entre père et fils. PROTÉGÉ comme les autres éditeurs, permet de construire une ontologie respectant les contraintes logiques d'OWL et de l'exporter directement dans ce langage. L'ontologie construite est donc référentielle dans l'éditeur et est exportée sous une forme computationnelle. Dans la suite de cet article, nous confondrons les deux dernières étapes [3].

Après avoir décrit le travail de construction de l'ontologie du diagnostic prénatal, nommée ONTODPN, nous synthétiserons notre démarche et proposerons une méthodologie générale de construction d'ontologie de domaine pour la médecine.

2 Contexte

2.1 Des ressources terminologiques et ontologiques

Une *ontologie* est définie en informatique, pour le domaine de l'ingénierie des connaissances (IC), comme un artefact particulier – formel – permettant de représenter des connaissances. Il est maintenant classiquement admis de distinguer dans une ontologie :

1. la *top*-ontologie, niveau le plus élevé structurant les connaissances de haut niveau avec des catégories dont l'organisation dépend de réflexions philosophiques ;
2. l'ontologie noyau ou *core*-ontologie, fournissant les concepts structurants du domaine et décrivant les relations entre ces concepts ; en médecine, on y trouve des concepts de *diagnostic*, *signe*, *structure anatomique* et des relations comme celles liées à la localisation d'une pathologie sur une structure anatomique ;
3. l'ontologie du domaine, c'est-à-dire les concepts du domaine tels qu'ils sont manipulés par les professionnels.

Nous supposons le lecteur familier de ces notions. Dans ce contexte, les niveaux de *top*-ontologie et *core*-ontologie d'ONTOMÉNÉLAS semblent une ressource permettant la construction de nouvelles ontologies en médecine en ne se focalisant que sur l'ontologie du domaine (Charlet *et al.*, 2009).

2.2 Le domaine du diagnostic prénatal

2.2.1 Caractéristiques des connaissances du domaine

Le pôle d'activité de périnatalité regroupe une très grande diversité d'acteurs : d'une part les soignants (gynécologues, pédiatres, radiologues, généticiens, etc.) qui mani-

1. PROTÉGÉ, Version 4.0, build 111 (`http ://protege.stanford.edu/`)

2. Ce *plugin* a été développé dans notre unité de recherche par L. Mazuel, il correspond à l'intégration d'une partie des fonctionnalités du logiciel DOE dans PROTÉGÉ.

3. Il existe une dernière étape qui correspond au chargement de l'ontologie dans un applicatif et qui peut amener des changements de langage d'expression des connaissances qui peuvent donc être de nature formelle. Nous ne discuterons pas cette problématique ici.

pulent chacun un grand volume de données sous forme de textes et d'images et d'autre part les patients à l'origine des données (mères, fœtus, nouveaux-nés). L'activité de diagnostic prénatal a lieu au sein de ce type de pôle et nécessite une bonne collaboration s'appuyant sur le partage de l'information produite par ces différents acteurs.

Le staff de diagnostic prénatal est une réunion de concertation pluridisciplinaire hebdomadaire qui permet une conjugaison d'expertises pour la prise en charge médicale d'un fœtus *in utero* ; ce staff qui rassemble tous les acteurs impliqués dans le domaine que nous souhaitons modéliser sera donc notre contexte de travail. Les caractéristiques des connaissances mises en jeu sont leur nombre, leur complexité, leur dispersion et leur défaut d'interopérabilité sémantique (en grande partie dû aux langages propres à chacune des spécialités médicales).

2.2.2 Objectifs de la modélisation de ce domaine

Les enjeux et limites d'une modélisation ontologique pour le domaine du diagnostic prénatal ont déjà été discutés avec les acteurs de la périnatalité (Dhombres *et al.*, 2009) ; si une représentation des connaissances selon une modélisation ontologique semble pertinente pour l'amélioration de l'interopérabilité sémantique, cette démarche ne sera validée que par évaluation de son apport réél par le biais de mises en situation applicatives concrètes telles que : a) l'aide au codage médico-économique, b) l'aide à l'indexation de documents, c) l'aide à la décision par recherche assistée dans des bases de données médicales existantes (PUBMED, OMIM, ORPHAnet, *etc...*), d) l'aide à la décision par recherche de cas analogues dans des bases de cas, e) l'aide à la décision par assistance au diagnostic différentiel.

Nous ne considérons pas ce modèle comme un constituant potentiel de référence médicale opposable, ou d'un outil d'évaluation des pratiques professionnelles.

3 Démarche de modélisation

3.1 Étape préliminaire : identification et recueil de supports de la connaissance du domaine

La démarche pour identifier les sources de connaissances du domaine a consisté en une modélisation selon la méthode du processus unifié, en respectant un formalisme UML, du staff de diagnostic prénatal du Réseau de Périnatalité de l'Est Parisien (R.PEP) qui se déroule à l'hôpital Armand Trousseau [4]. Notre participation régulière à ce staff ainsi que dans d'autres centres [5] a permis d'établir des diagrammes de séquences de prises de décision pour différents *scenarii* en prénatal.

Le compte rendu d'échographie apparaît le plus souvent comme point de départ de la prise en charge prénatale, conditionnant les hypothèses initiales puis toute la démarche diagnostique. Son contenu minimum, standardisé depuis 2005 par Sureau & Henrion,

4. AP-HP, 26, avenue du Docteur Arnold-Netter 75012 Paris, France. (`www.reseaupep.fr`)

5. Au sein des Centres Pluridisciplinaires de Diagnostic Prénatal de Poissy-Saint-Germain-en-Laye, de Cochin-Port-Royal et de Pontoise.

comporte pour une large partie la description de signes morphologiques fœtaux. Ainsi, il s'agit d'un support majeur des connaissances pour notre domaine.

La collaboration avec plusieurs services hospitaliers [6] ainsi qu'avec plusieurs membres du Collège Français d'Échographie Fœtale (CFEF) [7] a abouti à un recueil conséquent de ces documents (*cf.* tableau 1). Les documents de référence pour le domaine ont été collectés grâce à une collaboration et un accord d'exploitation de données avec le service INSERM SC11 ORPHAnet [8], une collaboration avec le CFEF, des recherches bibliographiques et l'accord de traitement de manuscrits originaux de manuels de formation.

Tableau 1 – Documents collectés pour la constitution des corpus.

Type de document	*n*
Comptes rendus d'échographie prénatale	194 143
Comptes rendus de radiopédiatrie	917
Comptes rendus de fœtopathologie (caryotypes, sérologie, ...)	5631
Fiches de synthèse ORPHA*net*	1 095
Textes de FMC *et séminaires du* CFEF	65
*Articles encyclopédiques (*EMC*)*	57
Manuels d'échographie et de médecine fœtale	2

3.2 Constitution de corpus textuels

3.2.1 Méthode

La constitution d'un corpus unique à partir de l'ensemble des textes du domaine ne semblait pas pertinente du fait de l'importante hétérogénéité des documents. Ainsi, trois catégories de documents ont été retenues pour constituer trois types de corpus différents [9] :

1er corpus : comptes rendus d'échographie. Le corpus complet, trop volumineux (de l'ordre de 10 millions de mots) pour notre outil de traitement automatique des langues (TAL), a été réduit à l'aide d'une table de hachage supprimant toutes les lignes présentes plus de 50 fois, le corpus étant très redondant, certaines phrases types répétées plusieurs milliers de fois. Ce premier corpus a été fina-

6. Centres de diagnostic prénatal des centres Hospitalo-Universitaires de Clamart (Dr Sénat, hôpital Antoine-Béclère, AP-HP) et de Grenoble (Dr Althuser). Service d'imagerie pédiatrique du Pr Ducou Lepointe (hôpital Armand Trousseau, AP-HP). Service de génétique et embryologie médicales du Pr Siffroi (hôpital Armand Trousseau, AP- HP).

7. Dr Bessis, Dr Althuser, Dr Martin, Dr Bussière et leurs collaborateurs.

8. Portail des maladies rares et des médicaments orphelins : `http ://www.orpha.net`

9. L'outil de TAL utilisé nécessite un formatage *ad hoc*. Nous avons donc du prétraiter tous les documents (.csv, .doc, .pdf, .html, *etc...*). Pour une partie des comptes rendus d'échographie, nous avons généré par un algorithme *perl* le fichier au format requis. Pour les autres documents, nous avons créé, soit manuellement, soit par des scripts JAVA (librairie JDOM v1.1), des fichiers XML respectant la DTD Docbook 4.2. Ces fichiers ont ensuite permis par le biais d'une feuille XSL de générer un fichier au format requis.

lement scindé en deux sous-corpus, echoVILLE et echoHOSPI, respectivement pour les examens réalisés en ville et en centre hospitalier.

2e corpus : comptes rendus de radiopédiatrie. Seul le texte des comptes rendus de radiologie pédiatrique prénatale (imagerie par résonance magnétique et tomodensitométrie fœtales) a été conservé, les comptes rendus de fœtopathologie n'ont pu être traité à cause d'un format propriétaire non exploitable.

3e corpus : documents de référence du domaine. Les textes de référence pour le domaine étant sous des formats très hétérogènes, une grande partie du traitement a été manuelle et spécifique pour chaque type de documents : ouvrages d'enseignement (Dumez *et al.*, 2004; Fredouille & Develay-Morice, 2006), articles encyclopédiques (EMC) et textes de formation médicale continue (FMC) du CFEF.

3.2.2 Analyse du corpus

L'analyse de corpus par extraction des candidats termes, des relations, des verbalisations qui correspondent aux normes et pratiques du domaine de l'imagerie prénatale est assurée par l'outil SYNTEXUPERY développé par Didier Bourigault.

Le premier module, SYNTEX (Bourigault *et al.*, 2005), est un analyseur syntaxique qui, à partir de corpus de textes, extrait un réseau terminologique composé de mots et de syntagmes (nominaux, adjectivaux ou verbaux). Chaque syntagme est relié d'une part à sa tête et d'autre part à son (ses) expansion(s) et est caractérisé par une fréquence d'apparition dans le corpus. Ces syntagmes constituent des *candidats termes* (CT) ; ce sont potentiellement des *termes* du domaine. SYNTEX les caractérise également par des indices de productivité (nombre de contextes dans lesquels chacun apparaît) et de répartition (nombre d'articles différents dans lesquels chacun apparaît). Le second module, UPERY (Bourigault, 2002), effectue une analyse distributionnelle à partir du réseau terminologique construit par SYNTEX et calcule les proximités distributionnelles entre les termes du réseau : il rapproche ainsi les couples de CT qui partagent les mêmes configurations syntaxiques, c'est-à-dire qui sont exprimés en tête d'un même ensemble de termes, ce sont les *voisins en tête*. Le résultat de l'analyse des corpus est présenté dans le tableau 2.

Tableau 2 – Résultats d'analyse des 3 corpus par SYNTEXUPERY.

corpus	1 (hosp.)	1 (ville)	2 (radio)	3 (réf.)	total
nombre de documents	35 619	47 434	927	-	-
nombre de mots	1 892 999	1 991 237	191 896	921 072	4 997 204
syntagmes nominaux	58 367	44 427	6 626	102 385	211 805
syntagmes verbaux	21 643	15 710	2 837	81 374	-

3.3 Normalisation sémantique

Les *termes* ont été dégagés manuellement par l'expert parmi les *candidats termes* dans cette étape méthodologique assistée par les résultats d'analyse d'UPERY, exploités

à l'aide de son interface graphique, TERMONTO. L'objectif central était de réduire l'ambiguïté des *termes*. Nous avons procédé en tant qu'expert du domaine pour identifier les concepts de celui-ci en effectuant une analyse par fréquence des syntagmes nominaux, puis un tri des syntagmes en fonction de leur nature selon quatre axes :

- syntagmes nominaux correspondant à des concepts de structures anatomiques,
- syntagmes nominaux correspondant au déroulement d'un examen,
- syntagmes nominaux correspondant à des concepts de pathologies et de syndromes.
- syntagmes nominaux correspondant à d'autres concepts évidents du domaine.

L'étude a porté sur les syntagmes ayant une fréquence supérieure ou égale à 7 (la rentabilité des termes de moindre fréquence apparaissant très faible) pour les corpus echoVILLE et echoHOSPI, soit respectivement 4220 et 5380 termes. Les concepts anatomiques ont été retenus sans qualificatif d'aspect ou d'état, par utilisation des termes ascendants : ainsi, à partir du syntagme « fémur droit incurvé », le terme ascendant retenu était « fémur ». Les synonymes ont été identifiés par les voisins en tête ; par exemple les termes « diagnostic prénatal », « diagnostic anténatal » et « DPN » ont été aisément identifiés. Les termes plus rares mais correspondant à des concepts pertinents étaient identifiables par l'étude des descendants en tête, comme pour dans le cas de « syndrome » qui comporte ainsi 109 expansions : syndrome de Turner, de Jeune, d'Alagile, *etc.*...

Par la suite, nous sommes passés du *terme* au *termino-concept* [10] puis avons identifié des relations sémantiques existant entre les deux (*est un*, *est une partie de*, ...).

3.4 Engagement ontologique

Introduction à l'engagement ontologique

L'étape précédente ayant permis d'organiser le bas d'une ontologie différentielle, dans cette étape nous nous sommes focalisés sur la réutilisation de ressources existantes, en particulier la *top*-ontologie d'ONTOMÉNÉLAS, passant alors d'une stratégie ascendante à une stratégie descendante.

ONTOMÉNÉLAS*, choix d'une ontologie noyau pour la médecine*

Le domaine du diagnostic prénatal comporte un grand nombre de concepts liés à l'anatomie (signes morphologiques et structures anatomiques), mais aussi d'autres types de concepts tels que les techniques d'imagerie, les types d'examens, les regroupements de signes en syndromes et associations malformatives, et plus généralement les concepts de diagnostic, de pronostic, de gravité, *etc*... Nous avons donc choisi de réutiliser les parties d'ONTOMÉNÉLAS correspondant aux niveaux de *top*-ontologie et de *core*-ontologie pour la médecine, ce qui a permis d'articuler de manière appropriée les concepts anatomiques dérivés du FMA aux concepts sémiologiques définis pour le diagnostic prénatal.

10. Un *termino-concept* est un concept qui n'est pas encore formalisé ; ceci correspond typiquement à une entrée de *theraurus*.

FMA, Foundational Model of Anatomy

L'ontologie ONTODPN a pour premier objectif de représenter la sémiologie de l'imagerie prénatale. Cette sémiologie étant pour une large part morphologique (description de structures anatomiques du fœtus), nous avons choisi de réutiliser l'ontologie de référence pour l'anatomie humaine de la *Washington University School of Medicine* : le *Foundational Model of Anatomy*, FMA (Rosse *et al.*, 1995).

L'utilisation du FMA version 3.0 en l'état, en introduisant très haut dans la hiérarchie des concepts, au niveau du concept MENELAS de *substratum*, la distinction entre les entités anatomiques (correspondant au concept *Anatomical entity*) d'une part et les autres concepts d'autre part, n'a pas été retenue. Le FMA ayant vocation selon ses auteurs eux-mêmes (Rosse & Mejino, 2003) à être réutilisé en partie et adapté à un usage spécifique, nous n'avons pas repris l'intégralité du FMA pour l'inclure dans ONTODPN, mais nous l'avons utilisé comme base pour notre construction d'ontologie.

Nous avons isolé les descriptions de signes d'imagerie fœtale à partir de nos corpus textuels. Pour chacune de ces descriptions, les éléments de l'anatomie correspondants ont été identifiés et recherchés au sein du FMA afin d'être nommés, organisés et en partie annotés de manière analogue au FMA. Un signe morphologique dans ONTODPN a été ainsi relié à une entité anatomique existante dans le FMA, ou à un nouveau concept anatomique le cas échéant. Différents choix de modélisation ont été faits selon la situation :

Concept absent du FMA. Certains concepts anatomiques identifiés par notre méthode sont assez spécifiques au diagnostic prénatal, comme celui de liquide amniotique (*PortionOfAnmioticFluid*). Dans cette situation, un nouveau concept est créé (figure 1.C), en accord avec la logique du FMA : *PortionOfAnmioticFluid* est donc positionné comme fils de *PortionOfBodySubstance*, en dessous du concept spécifique à ONTODPN : *GestationalSubstance*, lui-même créé dans le respect de la logique du FMA.

Concept présent dans le FMA, mais absent pour le fœtus. Certains éléments d'anatomie existent à la fois chez la mère et chez le fœtus et doivent pouvoir être modélisés sans ambiguïté d'appartenance. Sous le concept de *GestationalStructure* du FMA, il existe des concepts propres au fœtus et à l'embryon. Ces concepts sont cependant très insuffisants pour notre ontologie [11] ; nous avons alors choisi de procéder en deux étapes :

- La première est de déplacer le concept de *GestationalStructure* original du FMA au même niveau que celui d'*AnatomicalStructure* (figure 1.B) qui est renommé en *NonGestationalStructure* (figure 1.A) ; ces deux concepts sont frères, ayant comme père commun le concept d'*AnatomicalStructure* qui permet de faire l'interface avec le concept d'*Inanimate* du bas de la partie *core*-ontologie de ONTOMÉNÉLAS. Nous appelons ce type de concept, *concept d'interface* (*cf.* éléments marqué d'une $*$, figure 1).
- La seconde étape consiste en la reprise des concepts du FMA pour la mère en dessous du concept *NonGestationalStructure*, en utilisant les même labels que dans le FMA. Pour les concepts d'anatomie fœtale, les labels sont annotés avec

11. Seuls le cœur et l'utérus du fœtus y sont représentés comme organes spécifiques.

les caractères "_F" en suffixe et leur hiérarchie est reprise de manière identique à celle du FMA pour l'adulte (ce sont les concepts *type FMA*, *cf.* figure 1).

Nous avons ainsi distingué les structures anatomiques pour les organismes en développement (sous *GestationalStructure*) des structures anatomiques post-natales (sous *NonGestationalStructure*), en assurant le haut niveau de granularité nécessaire à notre modèle pour l'anatomie prénatale.

Concept présent dans le FMA, mais dont un des pères correspond à un concept d'ONTOMÉNÉLAS. Par exemple, les espaces anatomiques définis dans le FMA correspondent à des fils du concept de *PseudoObjet* d'ONTOMÉNÉLAS. Nous avons choisi la hauteur de raccordement entre les deux ontologies pour pouvoir utiliser les autres fils des concepts MENELAS utiles au domaine comme les concepts de *SystemicObject* et de *SociologicObject* (figure 1.D). Les concepts d'espaces anatomiques étant déjà spécialisés pour l'embryon et le fœtus sous *GestationalSpace*, nous les avons repris tels quels.

Concept présent dans le FMA, mais ayant une position inconsistante avec ONTOMÉNÉLAS. Les concepts de systèmes d'organes (*OrganSystem* et *OrganSystemSubdivision*) dans le FMA ne sont pas intégrables sous le concept d'*AnatomicalStructure* qui ne comporte pas la notion d'organisation autour d'une fonction particulière, notion présente dans la définition de système en physiologie. Afin de conserver cette sémantique, ces concepts sont définis comme fils du concept d'ONTOMÉNÉLAS *PhysicalSystemicObject* (figure 1.E). De plus, la distinction entre les systèmes gestationnels et non gestationnels est assurée par deux concepts d'interface *GestationalPhysicalSystemicObject* et *NonGestationalPhysicalSystemicObject*, de manière analogue au concept d'*AnatomicalStructure*.

Utilisation d'autres ressources

En plus d'ONTOMÉNÉLAS, très structurante, nous avons utilisé d'autres ressources, à même de compléter les notions du domaine, en particulier la CCAM (Classification Commune des Actes Médicaux), la CIM-10 (Classification Internationale des Maladies) et la partie de la base ORPHAnet pour les malformations congénitales. L'enregistrement des identifiants des concepts dans ces terminologies via des annotations permettent d'envisager, entre autres et par la suite, des liens vers les bases de données génétiques comme Gene, OMIM, Swissprot (par ORPHAnet) et l'aide au codage médico-économique (par la CCAM et la CIM-10).

Tests de cohérence et annotations linguistiques

L'opérationalisation de l'ontologie en OWL-DL est assurée par PROTÉGÉ, le raisonneur FaCT++ permet de contrôler la cohérence du modèle. Par ailleurs, un formalisme SKOS [12] permet de définir les synonymes et les labels pour les concepts de l'ontologie de manière standardisée, en supportant le multilinguisme.

12. `http ://www.w3.org/2004/02/skos/`

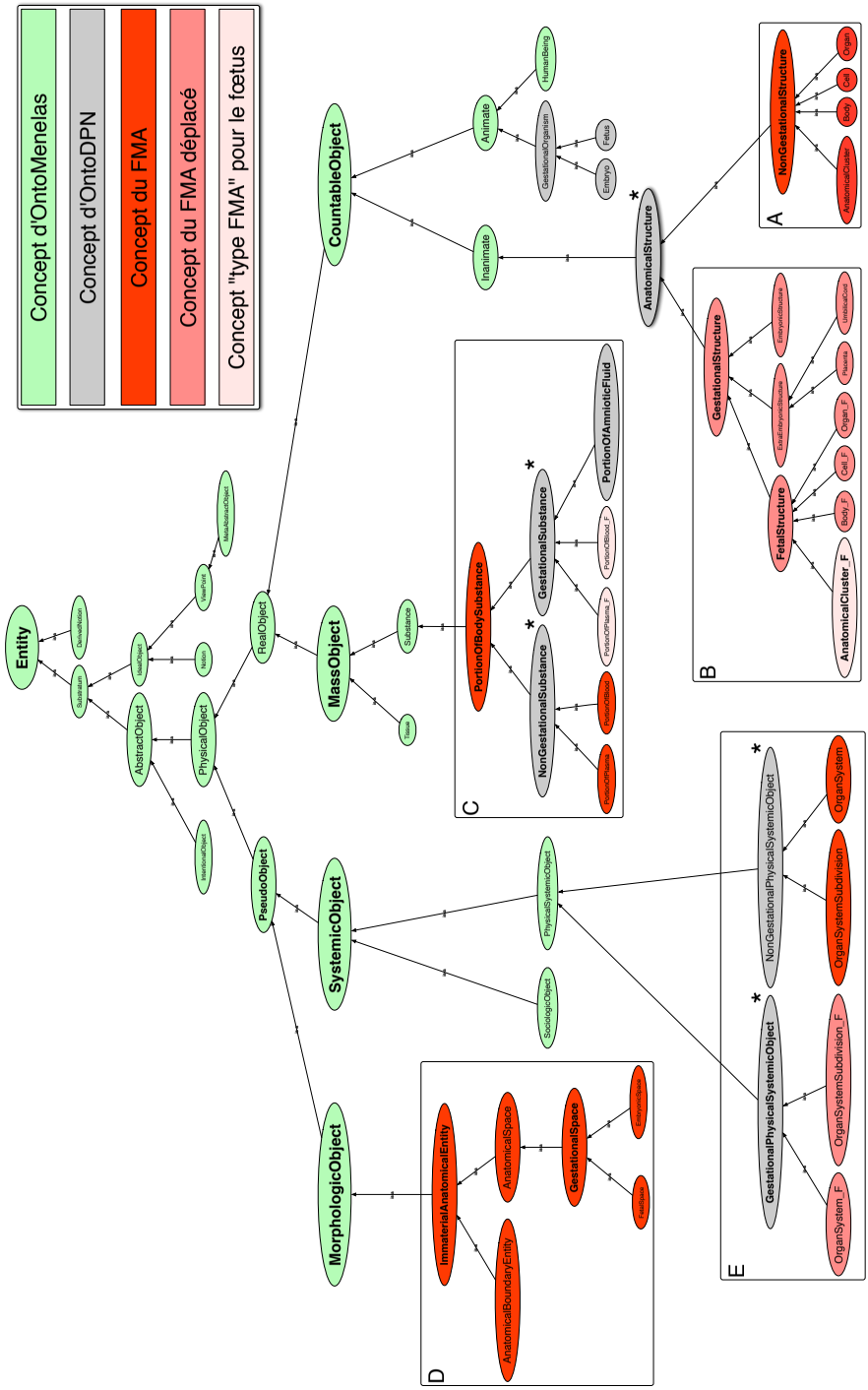

Figure 1 – Raccordement MENELAS–FMA dans ONTODPN. A : concepts de structures anatomiques pour l'anatomie adulte. B : concepts de structures anatomiques pour l'anatomie du fœtus et de l'embryon. C : Concepts des substances du corps humain. D : Concepts d'espaces anatomiques. E : Concepts de systèmes d'organes.

4 Discussion

Évaluation de l'ontologie

L'évaluation de l'ontologie passe par plusieurs étapes que nous n'avons pas encore abordées dans ce travail précis mais pour lesquelles un certain nombre de travaux, en particulier sur l'évaluation de l'ontologie en situation (Köhler *et al.*, 2009) et sur la réutilisabilité de la ressource (Baneyx & Charlet, 2006), permettent d'envisager des processus intéressants. En amont, les corpus constitués pour ce travail ont été extraits de bases de données par des algorithmes réutilisables pour des comptes rendus d'échographie provenant de deux plateformes les plus utilisées en échographie prénatale (VIEWPOINT® et *Top'*ECHO®). La forme des corpus, dans un format XML (Docbook), assure une portabilité de la ressource. Enfin, le volume (194 143 documents) et la diversité des données du corpus d'échographie en assurent une certaine représentativité.

Reproductibilité de la méthode.

Nous résumons notre méthologie dans le tableau 3 ; la reproductibilité de la méthode a pu être observée au sein de notre unité de recherche où d'autres ontologies médicales sont construites : en particulier ONTOLURGENCES (Giroud, 2009) pour les urgences et ONTOPNEUMO pour la pneumologie (Baneyx *et al.*, 2005). Par rapport aux différentes étapes du tableau, l'étape 2.c correspondant à l'organisation différentielle de l'ontologie n'est pas placée exactement où elle s'est déroulée. En effet, nous n'avions pas d'outil permettant d'éditer l'ontologie différentielle séparément de l'ontologie référentielle comme l'aurait permis DOE. Le plugin ARCHONTE, s'il permet de renseigner les principes différentiels, le fait au niveau de l'édition de l'ontologie référentielle. Dans ce travail, l'organisation différentielle a ainsi été intégrée dans l'étape d'engagement ontologique. Par ailleurs, le processus n'est pas linéraire. Si cette non-linéarité est maintenant admise, il faut noter qu'il y a de très nombreux allers-retours en particulier au sein des étapes 2 et 3 (*cf.* figure 2).

Lourdeur de la méthode.

La méthode utilisée ici, correspondant à ARCHONTE enrichie de la réutilisation de ressources existantes, est d'autant plus longue et lourde qu'il n'existe pas à ce jour de solution disponible qui intègre les différents outils nécessaires à cette démarche de modélisation ontologique ; ainsi il a été nécessaire, pour chacune des étapes, d'utiliser différents logiciels (SYNTEX, UPERY et TERMONTO, PROTÉGÉ avec plusieurs *plugins* comme ARCHONTE, ...). La perspective de pouvoir réaliser tout le processus dans une seule et même plateforme de travail comme celle de DaFOE (Szulman *et al.*, 2009) permet d'envisager une bien meilleure efficacité de notre démarche.

5 Conclusion

Nous avons décrit les principes d'une méthode de construction d'ontologie d'interface qui semble robuste. La part du travail de modélisation où l'expert du domaine

0	DÉFINITION DE L'USAGE
	a. Modélisation du domaine (UML) b. Identification du défaut d'interopérabilité sémantique
1	IDENTIFICATION DES RESSOURCES
	a. Constitution de corpus textuels de référence b. Constitution de corpus textuels de cas c. Collecte des ressources liées au domaine d. Choix d'une *top*-ontologie et d'une *core*-ontologie
2	NORMALISATION SÉMANTIQUE
	a. Analyse de corpus b. Identification des termes du domaine c. Organisation différentielle des termino-concepts
3	ENGAGEMENT ONTOLOGIQUE
	a. *Bottom-up* : passage du termino-concept au concept formel b. *Top-down* : réutilisation des RTO et d'une top-core c. Différentiation/élaboration des concepts primitifs et définis
4	ÉVALUATION DE LA RESSOURCE EN PRATIQUE (en attente)

Tableau 3 – Étapes de modélisation d'une ontologie de domaine complète.

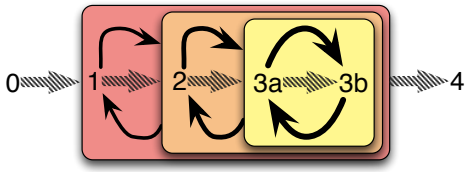

Figure 2 – De l'intérêt d'une plateforme de modélisation intégrée : une méthode par itérations entre les différentes étapes de modélisation décrites dans le tableau 3.

est nécessaire reste à définir, car pour ce travail nous avons été à la fois ingénieur de connaissance et expert du domaine ; il nous semble toutefois crucial qu'une interdisciplinarité active préside à l'entreprise de tels travaux. La prochaine étape du travail sera la complétion de l'ontologie, en particulier au niveau des relations. Des mises en situation d'utilisation du modèle sont à venir : pour l'indexation de bases de données médicales (ORPHAnet et CFEF), pour l'automatisation de procédures de codage médico-économique (CCAM et CIM-10). Des perspectives de recherche en aide à la décision sont également envisagées.

Références

BACHIMONT B., ISAAC A. & TRONCY R. (2002). Semantic commitment for designing ontologies : A proposal. In *EKAW '02 : Proceedings of the 13th Internatio-*

nal Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, p. 114–121, London, UK : Springer-Verlag.

BANEYX A. & CHARLET J. (2006). Évaluation, évolution et maintenance d'une ontologie en médecine : état des lieux et propositions méthodologiques. *Revue Information - Interaction - Intelligence*.

BANEYX A., CHARLET J. & JAULENT M. (2005). Building medical ontologies based on terminology extraction from texts : an experimentation in pneumology. *Studies in Health Technology and Informatics*, **116**, 659–64.

BOURIGAULT D. (2002). Upery : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. In *Actes de la 9ème conférence annuelle sur le Traitement Automatique des Langues*, p. 75–84, Nancy, France.

BOURIGAULT D., FABRE C., FRÉROT C., JACQUES M. & OZDOWSKA S. (2005). Syntex, analyseur syntaxique de corpus. In *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles*, Dourdan, France.

CHARLET J., BACHIMONT B., MAZUEL L., DHOMBRES F., JAULENT M. & BOUAUD J. (2009). OntoMenelas : motivation et retour d'expérience sur l'élaboration d'une ontologie noyau de la médecine. In *Journées Francophones des Ontologies*, Poitiers.

DHOMBRES F., JAULENT M., MAZUEL L., CHARLET J., BESSIS R., JOUANNIC J. & BÉNIFLA J. (2009). Apport des nouvelles technologies de l'information pour le partage des données et des connaissances en prénatal : une ontologie des signes de l'imagerie prénatale. In *5ème Journée de Périnatalité de l'Est Parisien*, Paris, France.

DUMEZ Y., BENACHI A. & COLLECTIF (2004). *Médecine foetale et diagnostic prénatal*. Doin Editions.

FREDOUILLE C. & DEVELAY-MORICE J. (2006). *Coeur foetal pratique : 3 temps - 10 points-clés*. Sauramps Médical, 2e édition revue et augmentée edition.

GIROUD M. (2009). L'accès au dossier médical personnel par le médecin régulateur du samu. In *3eme Congrès Urgences*, p. 807–816, Paris, France : Société Française de Médecine d'Urgence. Ce texte s'appuie sur les travaux du Groupe animé par l'ASIP santé (ex GIP-DMP).

KÖHLER S., SCHULZ M. H., KRAWITZ P., BAUER S., DÖLKEN S., OTT C. E., MUNDLOS C., HORN D., MUNDLOS S. & ROBINSON P. N. (2009). Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *American Journal of Human Genetics*, **85**(4), 457–464.

ROSSE C. & MEJINO J. L. V. (2003). A reference ontology for biomedical informatics : the foundational model of anatomy. *Journal of Biomedical Informatics*, **36**(6), 478–500.

ROSSE C., SAID M. B., ENO K. R. & BRINKLEY J. F. (1995). Enhancements of anatomical information in UMLS knowledge sources. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, p. 873–877.

SUREAU C. & HENRION R. (2005). *Rapport du Comité National Technique de l'échographie de dépistage prénatal*. Rapport interne, Ministère des solidarités, de la santé et de la famille.

SZULMAN S., CHARLET J., AUSSENAC-GILLES N., NAZARENKO A., SARDET E. & TEGUIAK V. (2009). DAFOE : an ontology building platform from texts or thesauri. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. Poster.