

---

# MEANS : une approche sémantique pour la recherche de réponses aux questions médicales

Asma Ben Abacha\* — Pierre Zweigenbaum\*\*

\* CRP Henri Tudor, Luxembourg

\*\* LIMSI-CNRS, France

asma.benabacha@tudor.lu, pz@limsi.fr

---

*RÉSUMÉ.* Nous présentons une approche pour la recherche de réponses à des questions médicales posées en langage naturel, appelée MEANS. Cette approche se fonde sur des techniques de TAL pour l'extraction des entités médicales et des relations sémantiques exprimées dans les questions et les corpus médicaux. MEANS utilise les langages du Web sémantique pour représenter et interroger les informations recherchées par les utilisateurs. Cette caractéristique permet de partager les informations extraites à partir des corpus et de considérer l'acquisition cumulative de connaissances à moyen et long terme. MEANS construit une requête SPARQL initiale et plusieurs requêtes relaxées comme interprétation sémantique formelle de la question. L'évaluation de MEANS sur un jeu de données réelles a abouti à des résultats encourageants en termes de précision et de MRR et a permis de constater les améliorations significatives apportées par sa méthode de relaxation de requêtes.

*ABSTRACT.* We present a medical question answering approach, called MEANS. This approach relies on natural language processing techniques to extract medical entities and relations from the user questions and medical corpora. It also uses semantic Web languages to represent and query the information searched by the users. This feature allows to share the information extracted from textual corpora using standard languages and to consider incremental knowledge acquisition in mid-and-long terms. MEANS constructs an initial SPARQL query and several relaxed queries as semantic interpretations of a user question. The evaluation of MEANS on a real dataset shows promising results for both precision and MRR and showed the significant benefits of the query relaxation technique.

*MOTS-CLÉS :* système de Questions-réponses, extraction d'information, recherche sémantique.

*KEYWORDS :* Question-Answering System, Information Extraction, Semantic Search.

---

## 1. Introduction

Au cours des dernières décennies, la quantité d'information a augmenté de façon exponentielle dans tous les domaines et notamment le domaine médical. En effet, le volume des connaissances médicales double tous les cinq ans (Engelbrecht, 1997), voire tous les deux ans (Hotvedt, 1996). Avec la numérisation à large échelle, retrouver automatiquement une information de haute précision est devenu un défi. Dans ce contexte, plusieurs moteurs de recherche spécialisés dans ce domaine ont vu le jour : citons par exemple PubMed<sup>1</sup>, CISMef<sup>2</sup> ou Health On the Net<sup>3</sup>. À une requête donnée, ces moteurs retournent un ensemble de documents et délèguent à l'utilisateur la tâche de trouver l'information cherchée si elle existe dans les documents renvoyés.

Les systèmes de questions-réponses présentent une alternative axée sur la précision pour faciliter et accélérer la recherche d'information. Un système de questions-réponses vise à répondre directement à des questions posées en langue naturelle avec une réponse précise extraite à partir d'une collection de documents, du Web ou d'une base de données. La nature et la complexité de cette tâche varient selon le domaine traité et les types de questions posées par les utilisateurs.

Ainsi, plusieurs critères et conditions du domaine ouvert ne sont pas valables pour le domaine médical, notamment en ce qui concerne le rôle des pronoms interrogatifs. Par exemple, le pronom '*When*' indique souvent une date en domaine ouvert alors que dans le domaine médical, il indique souvent une condition (e.g. « *When should you suspect community-acquired MRSA?* »). D'autres caractéristiques particulières sont aussi à prendre en compte en domaine médical, citons par exemple :

- les nouveaux types de questions (non nécessairement factuelles) ;
- les nouveaux types de réponses attendues ;
- les entités de domaine autres que les habituelles « entités nommées » du domaine ouvert (e.g. maladies ou traitements plutôt que noms de personnes connues) ;
- les relations sémantiques spécialisées (e.g. traiter, prévenir, diagnostiquer) ;
- la disponibilité de ressources sémantiques et terminologiques spécialisées comme le thésaurus MeSH (*Medical Subject Headings*) et le Metathesaurus UMLS (*Unified Medical Language System*) et son réseau sémantique ;
- les types de documents ciblés (articles scientifiques et leur résumé, recommandations pour la pratique clinique, ouvrages didactiques, etc.) : collections en général « fermées » mais potentiellement très vastes et « fiables ».

Le domaine médical a donc besoin de nouvelles approches adaptées pour l'extraction d'information à partir des corpus textuels, pour l'analyse de questions et l'extraction de réponses.

1. <http://www.pubmed.com>

2. <http://www.chu-rouen.fr/cismef/>

3. <http://www.hon.ch>

Cet article porte sur l'étude de méthodes permettant de rechercher sémantiquement des réponses à des questions médicales. Les principales questions posées sont :

- 1) Quelles sont les caractéristiques des questions en domaine médical ? Quelles classifications sont les plus pertinentes pour les questions médicales ?
- 2) Quelles représentations sont les plus robustes et efficaces pour les informations extraites à partir de questions et de corpus médicaux ?
- 3) Comment traduire une question écrite en langage naturel en une requête structurée qui retranscrit sa sémantique ?
- 4) Comment pallier les erreurs et les limites des méthodes d'extraction d'information utilisées ?

Nous discutons les questions 1 et 2 dans les sections 3 et 4. Nous utilisons une méthode en six étapes pour la construction de requêtes SPARQL<sup>4</sup> (*SPARQL Protocol and RDF Query Language*) à partir de questions posées en langage naturel. Cette méthode utilise des patrons génériques pour déterminer le type de la question et donc le type de requête qui s'ensuit. La technique proposée met aussi en œuvre des procédés de simplification et de construction de la forme affirmative de la question avant sa transformation finale en requête SPARQL.

Nous proposons une nouvelle méthode pour pallier les erreurs et les limites des techniques d'extraction d'information (section 5). Cette méthode se fonde sur la relaxation de requête SPARQL et permet d'atteindre les réponses erronément annotées ou partiellement annotées. Cette relaxation est effectuée de façon incrémentale en éliminant les maillons les moins importants sémantiquement en premier lieu. Cette approche constitue l'aspect novateur de cet article qui conclut un ensemble de travaux effectués sur les tâches en amont de la recherche de réponses, à savoir l'extraction d'information à partir de textes médicaux (Ben Abacha et Zweigenbaum, 2011a ; Ben Abacha et Zweigenbaum, 2012b ; Ben Abacha et Zweigenbaum, 2012c) et l'analyse de questions (Ben Abacha et Zweigenbaum, 2012b). Plus précisément, dans (Ben Abacha et Zweigenbaum, 2012b) nous avons présenté la méthode proposée pour l'analyse des questions médicales et la construction automatique de requêtes SPARQL. Nous avons évalué cette méthode sur un corpus de questions médicales sans s'adresser à la recherche de réponses. Dans cet article, nous nous concentrons sur la recherche de réponses et nous présentons notre méthode pour la recherche « sémantique » de réponses aux questions médicales. Le présent article décrit aussi l'architecture générale du système de questions-réponses MEANS ainsi que son évaluation sur un nouveau corpus de questions-réponses.

Dans ce travail, nous nous intéressons essentiellement à des corpus de la littérature biomédicale (e.g. Medline<sup>5</sup>, PubMed<sup>6</sup>). L'approche a aussi l'avantage de pouvoir intégrer l'interrogation de bases de connaissances RDF locales ou distantes grâce à l'ou-

4. <http://www.w3.org/TR/rdf-sparql-query/>

5. <http://www.nlm.nih.gov/pubs/factsheets/medline.html>

6. <http://www.nlm.nih.gov/pubs/factsheets/pubmed.html>

verture apportée par les technologies et les langages du Web sémantique. Cet aspect n'est cependant pas testé dans le présent article qui porte uniquement sur l'extraction de réponses à partir de l'annotation de corpus textuels.

Cet article est structuré de la façon suivante. La section 2 donne une présentation générale des systèmes de questions-réponses et un résumé de quelques classifications proposées et travaux existants. Dans la section 3 nous présentons notre approche pour répondre aux questions médicales. La section 4 détaille notre méthode pour l'analyse des questions et la section 5 détaille notre méthode pour la recherche de réponses. Enfin, l'évaluation de notre système de questions-réponses **MEANS** est présentée en section 6.

## 2. Systèmes de recherche de réponses à des questions

La recherche de réponses précises à des questions formulées en langue naturelle est un des défis majeurs posés dans le champ de la recherche d'information. Les premiers systèmes de questions-réponses tels que **BASEBALL** (Green *et al.*, 1961) et **LUNAR** (Woods, 1973) ont vu le jour dans les années 60. Le système **BASEBALL** traite les questions qui portent sur les dates, les lieux ou les résultats des matchs de baseball du championnat américain. **LUNAR** est un des premiers systèmes de questions-réponses scientifiques. Il a été conçu pour assister l'analyse géologique des pierres retournées par la mission Apollo. Lors de son expérimentation, il a pu répondre avec succès à 90 % des questions posées par les utilisateurs. Les deux systèmes **BASEBALL** et **LUNAR** exploitaient des bases de connaissances écrites manuellement par des experts de leurs domaines d'application.

Un système de questions-réponses (SQR) vise à fournir une réponse à une question posée en langage naturel. La réponse visée peut aussi bien être un morceau de texte extrait d'une collection de documents ou provenant du Web, qu'une donnée récupérée d'une base de données ou base de connaissances. Dans d'autres cas plus rares, les réponses retournées sont des informations multimédias. C'est le cas par exemple du système **START** (Katz, 1999), introduit en 1993 comme un « serveur d'information », qui permet de répondre à des questions écrites en anglais avec des informations multimédias. Il utilise des techniques issues du TAL pour analyser les phrases et associe des annotations linguistiques à des segments d'information multimédias. Ces annotations sont ensuite exploitées pour retourner les bons segments d'information aux questions posées par les utilisateurs.

Dans un cadre général, un SQR peut être décomposé en trois tâches principales : (i) analyse de la question utilisateur, (ii) analyse du corpus à partir duquel les réponses vont être extraites, et (iii) recherche et extraction des réponses en sélectionnant le meilleur appariement question-réponse. La deuxième tâche ne sera pas requise pour les systèmes qui extraient leurs réponses à partir de bases de données ou de bases de connaissances structurées. Les méthodes employées pour l'analyse de questions et/ou de corpus peuvent être surfaciques (*e.g.* fondées sur l'indexation des mots-clés de la

question et du corpus), sémantiques (*i.e.* formalisation explicite de tout ou partie du sens de la question et du corpus<sup>7</sup>) ou hybrides (combinant les deux approches).

### 2.1. Classification des systèmes de questions-réponses

Pour cause de limitation d'espace nous ne présentons pas ici un état de l'art étendu sur les méthodes proposées pour trouver des réponses aux questions. Plusieurs travaux portent sur la présentation et la classification des SQR existants (Athenikos et Han, 2010 ; Lopez *et al.*, 2011). Dans cette section, nous présentons et discutons trois classifications des systèmes existants.

Plusieurs classifications ont été proposées pour les SQR. Par exemple, Moldovan *et al.* (2003) se sont appuyés sur la complexité des questions traitées et la difficulté de la tâche d'extraction de réponses pour proposer cinq classes de systèmes avec une complexité croissante :

- 1) systèmes capables de répondre à des questions factuelles ;
- 2) systèmes employant des processus de raisonnement simples ;
- 3) systèmes qui constituent les réponses à partir de plusieurs sources à la fois ;
- 4) systèmes qui proposent un dialogue interactif avec l'utilisateur ;
- 5) systèmes capables d'effectuer un raisonnement analogique.

La première catégorie de systèmes proposée couvre un grand nombre de systèmes traitant les questions WH (*What, Who, Where, etc.*). Dans la classification proposée cette catégorie regroupe les approches qui extraient des réponses constituées de morceaux de textes provenant d'un ou de plusieurs documents. La deuxième catégorie de systèmes est une sous-classe de la première catégorie où l'extraction de réponses nécessite d'effectuer des inférences logiques. Dans la troisième classe de systèmes proposée, la réponse est éparpillée dans plusieurs documents et une fusion est nécessaire pour la constituer. Le quatrième type de systèmes exploite les interactions précédentes avec l'utilisateur pour extraire les réponses. Enfin la dernière catégorie, jugée plus complexe, consiste à extraire des réponses qui ne sont pas explicites dans les documents. C'est le cas, par exemple, des questions de type prédiction (*e.g.* « *Is the Fed going to raise interests at their next meeting?* »). Afin de répondre à de telles questions, les SQR doivent effectuer un raisonnement par analogie complété potentiellement par un raisonnement temporel, spatial ou conditionnel.

Cependant cette approche n'utilise pas les entrées/sorties comme critère de classification des SQR car elle s'est fixé un cadre avec toujours le même type d'entrées (questions en langage naturel) et le même type de sorties (extraits textuels). Elle divise aussi les méthodes exploitant des inférences ou des raisonnements en raisonnement simple et raisonnement avancé (par analogie), ce qui constitue un critère de classification de précision élevée qui ne peut pas être évalué pour tous les systèmes. Aussi les

7. Le mot *sémantique* sera utilisé dans ce sens dans le reste de l'article.

types de techniques utilisées pour l'analyse de la question et/ou des corpus et l'appariement questions/réponses ne sont pas exploités comme un critère de classification. Enfin, cette classification n'explique pas non plus les sources desquelles les réponses sont collectées ou extraites.

D'autres travaux explicitent les sources des réponses et les entrées/sorties des SQR comme des critères de classification. Par exemple, Lopez *et al.* (2011) proposent un état de l'art sur une sous-catégorie des SQR qui exploitent des métadonnées sémantiques présentes dans des bases de connaissances structurées et des ontologies. Ils proposent pour cela deux classifications des systèmes à base d'ontologies. Une première classification en trois catégories : (i) interfaces en langage naturel pour les bases de données, (ii) recherche de réponses à des questions à partir de documents textuels et (iii) recherche de réponses à des questions avec des données, des textes et des langages propriétaires. Une deuxième classification suivant les sources des réponses : (i) bases de données structurées, (ii) textes non structurés et (iii) bases de connaissances sémantiques précompilées. Athenikos et Han (2010) ont proposé une classification des techniques utilisées pour la recherche de réponses et fondées sur des connaissances sémantiques en trois catégories: (i) SQR sémantiques, (ii) SQR basés sur les inférences et (iii) SQR fondés sur des représentations logiques. Cette classification reste cependant ambiguë car les trois catégories ont beaucoup d'intersections ce qui ne permet pas de trier efficacement les différents systèmes. Athenikos et Han (2010) proposent aussi une classification des SQR médicaux qu'ils partagent en deux catégories : (i) approches s'appuyant sur des connaissances sémantiques et (ii) approches s'appuyant sur des connaissances non sémantiques.

À propos de ces trois travaux, nous observons que l'objectif des auteurs n'était pas de définir des critères qui auraient permis de trier et étiqueter efficacement les SQR mais plutôt de donner quelques grandes directions de recherche et synthétiser les défis actuels et ceux à venir pour les SQR. Ainsi, bien que ces travaux fournissent des descriptions et des résumés pertinents pour chaque SQR individuellement, les trois classifications de SQR proposées présentent certaines ambiguïtés. Par exemple, le type de représentation de la question (*e.g.* SQL, XQuery, SPARQL, formules logiques) et le type de représentation des informations extraites des corpus (*e.g.* XML, relationnel, RDF, textuel) ne sont pas explicités comme critères de classification. Les techniques d'extraction d'information appliquées sur la question en langage naturel ne sont pas précisées dans la catégorisation (*e.g.* patrons utilisant les catégories morpho-syntaxiques des mots, les arbres de dépendance, techniques d'apprentissage artificiel). Pour les SQR sémantiques, les éventuelles techniques de désambiguïsation du sens des mots ne sont pas non plus prises en compte comme critère de classification.

## **2.2. Systèmes de questions-réponses en domaine biomédical**

Dans cette section, nous décrivons quelques-uns des travaux les plus significatifs sur les SQR « sémantiques » dans le domaine biomédical.

Plusieurs approches ont été proposées pour la réalisation de SQR dans le domaine biomédical. Par exemple, le système ExtrAns est un système conçu pour s'adapter aux domaines restreints, il a été expérimenté sur des articles de recherche en génomique (Rinaldi *et al.*, 2004). Les documents de ce corpus médical sont analysés hors ligne avec des techniques de TAL puis transformés en une représentation sémantique, appelée « *Minimal Logical Form* », sauvegardée dans une base de connaissances. Pendant la phase d'interrogation, la question de l'utilisateur est analysée avec la même méthode utilisée pour l'analyse du corpus. Le système recherche ensuite un appariement entre la représentation sémantique de la question utilisateur et la base de connaissances. Si un appariement est découvert, les phrases qui ont permis cet appariement sont retournées comme réponses possibles à la question.

Terol *et al.* (2007) utilisent la taxinomie de questions proposée par Ely *et al.* (2000) et ciblent uniquement les dix catégories de questions les plus fréquentes dans leur SQR médical. Ce système procède en quatre étapes : (i) analyse de questions, (ii) recherche de documents en ligne et/ou en local, (iii) sélection de passages pertinents et (iv) extraction de réponses. L'analyse de la question et des documents est effectuée en dérivant des formes logiques à partir des relations de dépendance entre les mots de la phrase (ou la question) en utilisant la codification de la version étendue de WordNet (*i.e.* les prédicats de la forme logique ne relient pas les mots mais les ensembles de synonymes auxquels ils appartiennent). Les réponses sont extraites parmi les phrases des passages sélectionnés en comparant la forme logique de la question à celle de la phrase.

Embarek (2008) a travaillé sur la mise en place d'un SQR pour le domaine médical (le système Esculape) en adaptant un SQR généraliste existant (le système Œdipe). L'analyse de questions utilise un ensemble de règles construites manuellement pour déterminer la relation principale de la question, en partant de l'hypothèse qu'une part significative des questions médicales peuvent être modélisées sous la forme d'une relation unissant une ou plusieurs entités explicitement instanciées dans la question et une entité absente correspondant à la réponse cherchée (ce qui est conforme au premier modèle de Jacquemart et Zweigenbaum (2003)). Le module suivant cherche la réponse dans une ontologie médicale, construite à partir de ressources sémantiques disponibles, qui représente les entités et les relations les plus communément manipulées en médecine. Si la réponse n'est pas trouvée, une autre méthode est appliquée qui se base sur des patrons linguistiques (appelés patrons multiniveaux) appris de façon supervisée pour repérer et extraire la réponse à partir d'une collection de documents.

Le système EPoCare (Niu et Hirst, 2004) effectue une recherche par mots-clés dans une base de documents XML pour répondre aux questions utilisateurs posées en langage naturel. Plus précisément la question de l'utilisateur est traduite par le système en une requête formée de mots-clés combinant les termes recherchés et les catégories médicales dans le format PICO. Ce format distribue les caractéristiques d'une question dans quatre champs : les problèmes médicaux et les patients (*P*), les interventions (*I*), les relations de comparaison (*C*) et les résultats d'interventions (outcome : *O*). Le processus d'extraction de réponses utilise un ensemble de patrons de réponses

préconstruits qui consiste en des chemins XML pour chacune des quatre catégories définies dans le format PICO. Le système identifie les chemins XML qui contiennent tous les mots-clés de la question tout en filtrant les chemins qui ne correspondent pas aux catégories PICO associées à ces mots-clés. Dans un travail qui précède EPoCare, Niu *et al.* (2003) ont proposé une approche pour répondre à des questions médicales reposant sur l'identification des catégories PICO. Le processus employé consiste à déterminer les frontières textuelles de chaque catégorie puis à identifier les relations entre les différentes catégories détectées dans la phrase ou la question. Les relations traitées étaient restreintes aux relations thérapeutiques (traitement).

Demner-Fushman et Lin (2005) ont aussi proposé une approche qui utilise le format PICO pour décrire leurs composants d'extraction de connaissances sémantiques dans le cadre d'un SQR médical. Ils identifient les éléments correspondant aux catégories PICO dans les résumés d'articles de MEDLINE. Demner-Fushman et Lin (2006) ont proposé une approche hybride pour les SQR en domaine médical qui repose sur des techniques issues de la recherche d'information et du résumé automatique de textes. Ils ont ciblé une classe fréquente de questions de la forme « *What is the best drug treatment for X?* ». Étant donné un ensemble initial de résumés de MEDLINE retrouvés avec PubMed, le système identifie d'abord les médicaments en cours d'étude en utilisant le composant d'extraction des interventions proposé dans (Demner-Fushman et Lin, 2005). Il regroupe ensuite dans des clusters sémantiques les résumés MEDLINE retrouvés en exploitant (i) les interventions principales identifiées dans le texte du résumé et (ii) un algorithme de partitionnement agglomératif hiérarchique afin de calculer les similarités entre les interventions. Pour chaque résumé de MEDLINE, le système produit un résumé plus court composé de l'intervention principale, du titre du résumé et de la meilleure phrase citant un résultat d'intervention, calculée par le composant d'extraction des résultats d'interventions, présenté dans (Demner-Fushman et Lin, 2005).

Cependant, comme observé par Sackett *et al.* (2000), la représentation PICO ne permet pas d'exprimer la sémantique complète de la question en langage naturel. Huang *et al.* (2006) ont examiné davantage le format PICO en étudiant sa compatibilité avec les questions cliniques posées en langage naturel en classifiant manuellement un ensemble de cent questions. Les principales observations qu'ils ont effectuées sur l'adéquation du format avec les questions évaluées sont que la représentation PICO est mieux adaptée pour traiter les questions de type *therapy* (traitement) et moins adéquate pour les questions de type *diagnosis* (diagnostic), *etiology* (cause) et *prognosis* (pronostic). Par ailleurs, le format PICO ne permet pas de reconstruire la question originale à partir des informations extraites (*P*, *I*, *C* et *O*). Ceci est dû principalement à l'incapacité d'exprimer des relations précises entre les entités médicales (*e.g.* est-ce que [Problem: hypomagnesemia, Intervention: ?] correspond à « *What is the most effective treatment for hypomagnesemia?* » ou à « *What are the causes of hypomagnesemia?* » ?). D'autres limitations incluent l'ambiguïté de ce format (*e.g.* il représente le problème médical et aussi la population par le même élément *P*) et son incapacité à identifier les relations anatomiques.

Cairns *et al.* (2011) ont proposé la plate-forme multisource MiPACQ pour répondre automatiquement aux questions cliniques. Dans le cadre de leurs expérimentations, l'application de techniques d'apprentissage artificiel a permis une amélioration du MRR<sup>8</sup> (0,266) et de la précision de la première réponse retournée (0,173). Leur système utilise des annotations sémantiques de la question et des réponses potentielles mais de manière statistique uniquement, *i.e.* seule l'occurrence de la même entité UMLS dans la question et la réponse est considérée, les relations sémantiques entre ces entités ne sont ni extraites ni recherchées. Cela constitue une différence majeure par rapport à la méthode que nous proposons et qui se fonde aussi bien sur les entités que sur les relations qui les lient. Leur système a aussi pour point de départ les cent premiers paragraphes retournés par un système de recherche d'information classique à base de mots. Cette méthode de base peut ainsi écarter dès le départ des paragraphes pertinents qui contiennent des synonymes des mots de la question.

### 2.3. Synthèse et positionnement

Nous proposons un SQR « sémantique » pour le domaine médical. Nous pensons que les approches sémantiques sont les plus appropriées pour ce domaine d'application car elles permettent de suivre son évolution en termes de connaissances, notamment parce qu'elles permettent d'interroger des bases de connaissances provenant de sources différentes ou construites de multiples façons. En effet, ces bases peuvent aussi bien être construites à partir de bases de données publiques ou propriétaires (*e.g.* BIO2RDF (Belleau *et al.*, 2008)) que par l'annotation de corpus médicaux. Aussi, dans le cadre du projet Linked Open Data<sup>9</sup> plusieurs bases de connaissances biomédicales ont été publiées en ligne (*e.g.* BioGateway<sup>10</sup>, DrugBank<sup>11</sup>). L'avantage ici étant que l'intégration se fait avec des langages standard qui donnent une formalisation unique aux liens entre les différentes bases et facilite ainsi leur partage et leur accès (*e.g.* les axiomes standard OWL<sup>12</sup> `sameAs` pour la réconciliation d'individus ou `equivalentClass` pour l'alignement de concepts ontologiques).

Cependant, bien que l'interrogation directe de ces bases permette de répondre aux questions des utilisateurs avec des faits (triplets), les éléments retournés en réponses (*e.g.* URI de concepts ontologiques ou d'instances de ces concepts) ne sont pas forcément pertinents ou parlants pour les utilisateurs. Ainsi, le fait de pouvoir justifier ou retrouver une réponse grâce à un corpus textuel devient important. Dans notre vision, ces deux types de réponses ne sont pas disjoints mais complémentaires. Un SQR sémantique efficace devrait pouvoir rechercher des réponses à partir des annotations sémantiques de corpus cibles et compléter les informations manquantes dans ces annotations en interrogeant les bases de connaissances disponibles. L'exploita-

8. *Mean Reciprocal Rank*

9. <http://linkeddata.org>

10. <http://www.semantic-systems-biology.org/biogateway>

11. <http://www4.wiwiw.fu-berlin.de/drugbank/>

12. <http://www.w3.org/2004/OWL/>

tion conjointe des bases de données et des corpus annotés est aussi employée par le système de questions-réponses START (Katz *et al.*, 2002), cependant, l'ordre y est inversé car les réponses sont recherchées d'abord dans une base de données (donc sans justifications) puis dans les documents textuels.

Sur un deuxième plan, une des observations principales qui peut être faite sur les systèmes existants est que les processus d'analyse de questions se fixent souvent un cadre dans lequel une question ne peut avoir qu'un seul focus, un seul type de réponse attendu et une seule relation principale, ce qui ne permet pas de saisir complètement les questions posées par les utilisateurs. L'approche que nous proposons prend en compte ces aspects au moment de la formalisation ou de l'interprétation des questions de l'utilisateur en construisant une ou plusieurs requêtes structurées pour une question posée en langage naturel.

### 3. Approche proposée

Dans cette section nous commençons par comparer trois types de représentation ou d'interrogation des informations extraites et des questions dans le cadre des systèmes de questions-réponses :

- 1) les langages propriétaires ou *ad hoc* ;
- 2) la représentation relationnelle et l'interrogation avec SQL ;
- 3) la représentation « sémantique » avec les langages du Web sémantique.

Nous présentons ensuite les langages du Web sémantique que nous avons choisis pour notre approche ainsi que l'architecture de notre SQR, appelé **MEANS**.

#### 3.1. Choix de représentation

Plusieurs SQR en domaine ouvert ou en domaine médical ont opté pour un format de représentation *ad hoc* des informations extraites du corpus et de la question une fois analysée. C'est le cas, par exemple, du format PICO qui a été utilisé par Sackett *et al.* (2000) pour représenter les informations extraites des corpus ou des questions dans le domaine médical. Bien que ce format ait été à l'origine conçu pour aider à formuler des recherches bibliographiques pour des humains, il a été exploité par plusieurs approches pour représenter des informations médicales extraites automatiquement.

La représentation avec un langage propriétaire ou *ad hoc* limite la couverture du domaine d'une part et l'interopérabilité des SQR d'autre part. Par exemple, il ne sera pas possible pour un système fondé sur le format PICO d'exploiter directement les résultats d'extraction d'information exprimés dans d'autres langages, ce qui implique que pour exploiter de nouveaux corpus ou de nouvelles méthodes d'extraction il sera nécessaire soit de modifier les outils d'extraction, soit de construire des adaptateurs spécifiques qui permettront de faire le lien entre les différents langages. Sachant que les connaissances dans le domaine médical évoluent très rapidement, il sera de moins

en moins concevable pour un SQR d'adopter une telle stratégie car elle ne permet pas, à l'échelle mondiale, de construire, d'extraire ou de modifier les connaissances de manière cumulative.

D'autres SQR représentent les informations extraites par des tables relationnelles et transforment les questions des utilisateurs en requêtes SQL (Popescu *et al.*, 2003). L'avantage de cette approche est qu'elle permet d'interroger à la fois les informations extraites à partir d'un corpus de documents mais aussi les informations préexistant dans les bases de données relationnelles métiers utilisées par une institution ou une entreprise donnée. Cependant, pour ce faire, le même schéma de base de données doit être utilisé. Ceci est le principal inconvénient de la représentation relationnelle car l'intégration de données provenant de sources différentes requiert le développement d'adaptateurs ad hoc pour intégrer les différents schémas de données.

Dans la dernière décennie, plusieurs métalangages tels que RDF(S)<sup>13</sup> et OWL<sup>14</sup> ont été normalisés par le W3C afin de formaliser la représentation du sens sur le Web. Ces langages fournissent un niveau élevé d'expressivité et sont de plus en plus utilisés dans des applications sémantiques et soutenus par des systèmes de stockage efficaces ainsi que des API (*e.g.* Sesame<sup>15</sup>, Jena<sup>16</sup>) qui facilitent la lecture et l'interrogation des données. Mis à part l'avantage de l'accessibilité rendue uniforme grâce au langage d'interrogation SPARQL<sup>17</sup>, les schémas de données dans le cadre du Web sémantique sont des ontologies qui sont plus facilement partageables que les schémas de bases de données relationnelles. En effet, l'intégration de données dans le cadre du Web sémantique peut être automatisée grâce aux axiomes par défaut OWL qui ont standardisé la façon avec laquelle les alignements entre différentes ontologies sont exprimés et interprétés.

Notre choix s'est porté sur les langages du Web sémantique RDF(S)/SPARQL car notre vision est que la conception de SQR efficaces dans le futur va requérir la disponibilité de bases de connaissances universelles accessibles qui peuvent être interrogées et mises à jour avec une interopérabilité aussi bien technique que sémantique.

Dans le cadre de notre approche, nous utilisons le langage RDF(S) pour définir l'ontologie de domaine de référence et le langage SPARQL qui permet d'interroger des données au format RDF. Le langage RDF<sup>18</sup> représente les données comme un ensemble de triplets (sujet, prédicat, objet). Le sujet et le prédicat d'un triplet sont des ressources RDF (rdfs:Resource) définies par des IRI (*Internationalized Resource Identifiers*)<sup>19</sup>. L'objet d'un triplet peut aussi bien être une ressource RDF qu'un littéral. Une ressource RDF conforme à une ontologie est soit une propriété (rdf:Property),

13. <http://www.w3.org/TR/rdf-syntax/>

14. <http://www.w3.org/TR/owl-ref/>

15. <http://www.openrdf.org>

16. <http://jena.sourceforge.net/>

17. <http://www.w3.org/TR/rdf-sparql-query/>

18. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>

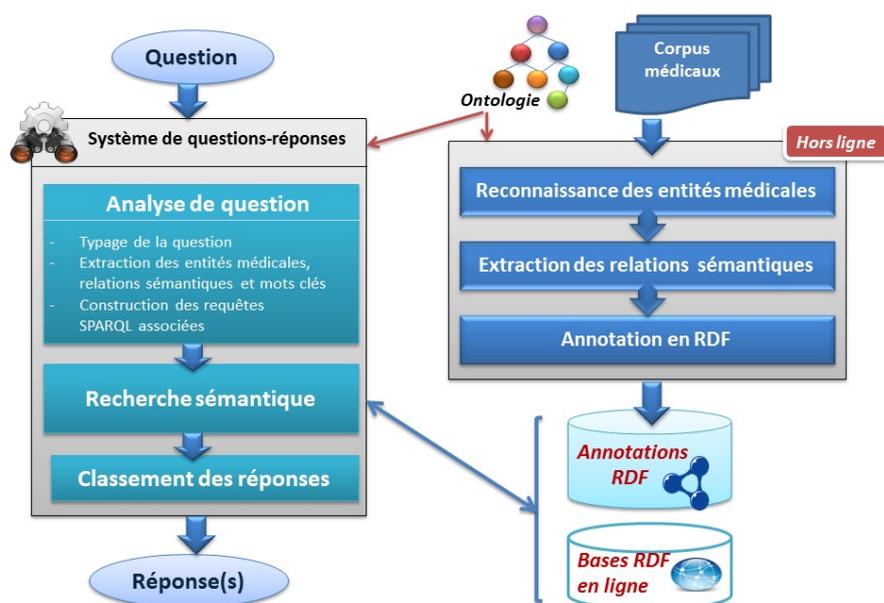
19. <http://tools.ietf.org/html/rfc3987>

une classe (`rdfs:Class`) ou une instance de classe  $i$  vérifiant ( $i$  `rdf:type`  $C$ ) avec  $C$  la classe dont  $i$  est l'instance. Les prédicats sont toujours des propriétés RDF. Un ensemble de données RDF forme un graphe dirigé dont les arcs sont les prédicats et les nœuds sont des ressources RDF ou des littéraux.

### 3.2. Architecture du système de questions-réponses MEANS

Un SQR nécessite deux entrées : le corpus utilisé pour extraire les réponses pertinentes et la question elle-même. Chacune de ces deux entrées doit être analysée correctement pour pouvoir trouver le meilleur appariement question-réponse. Dans une perspective de réutilisation, il est aussi important de représenter à la fois les questions et les réponses candidates avec une représentation formelle homogène pouvant être traitée par les systèmes d'information.

Nous proposons une approche sémantique pour répondre aux questions médicales (cf. figure 1) qui combine l'utilisation de connaissances du domaine médical, de techniques de TAL et de technologies du Web sémantique.



**Figure 1.** Approche proposée pour la réalisation d'un système de questions-réponses pour le domaine médical

Cette approche effectue une interprétation sémantique des documents et de la question et ramène le problème de recherche de réponses à des questions en langage naturel à l'interrogation de métadonnées structurées suivant une ontologie de domaine.

L'idée est d'associer des graphes sémantiques à la question d'une part, et aux phrases du corpus d'autre part, puis de rechercher les appariements pour trouver l'extrait de document qui répond à la question posée. Ceci permet de traiter les questions indépendamment de leur correspondance avec un type de question prédéterminé (*e.g.* booléenne, factuelle, cause/conséquence).

Notre approche comporte les trois composantes suivantes :

1) **analyse et annotation hors ligne des documents utilisés pour trouver les réponses.** Il s'agit d'annoter les corpus qui seront utilisés pour trouver les réponses. Cette annotation se fait hors ligne et fournit en sortie des triplets RDF qui seront interrogés pour trouver les réponses (cf. section 5.1). L'annotation comporte deux étapes : (i) la reconnaissance des entités médicales présentes dans les textes (*e.g.* maladie, médicament) et (ii) l'extraction des relations sémantiques qui les relient (*e.g.* traiter, prévenir, causer) ;

2) **analyse des questions en anglais.** Cette étape consiste à analyser la question et à extraire ses caractéristiques (*i.e.* type de la question, type de la réponse attendue, entités médicales, etc.), puis à construire à partir de ces informations extraites une ou plusieurs requêtes SPARQL avec des degrés de précision différents (cf. section 4) ;

3) **recherche sémantique et classement des réponses.** Cette étape se fonde sur un moteur de recherche RDF pour effectuer l'appariement requêtes et documents et trouver les phrases ou les entités médicales pertinentes. Le classement des réponses est effectué en fonction du degré de précision de la requête SPARQL utilisée et du nombre de justifications retrouvées pour la même réponse (cf. section 5.2).

RDF et SPARQL permettent une grande expressivité en représentant et en interrogeant les données comme des instances de concepts et de relations définies dans une ontologie de référence. Cette approche a été implémentée dans le SQR MEANS. L'utilisation de cette approche et de ces technologies par le SQR MEANS nous permet d'envisager d'exploiter les bases de connaissances RDF en ligne (*e.g.* BIO2RDF (Belleau *et al.*, 2008), BioGateway<sup>20</sup>, DrugBank<sup>21</sup>) comme sources complémentaires de réponses aux annotations de corpus médicaux.

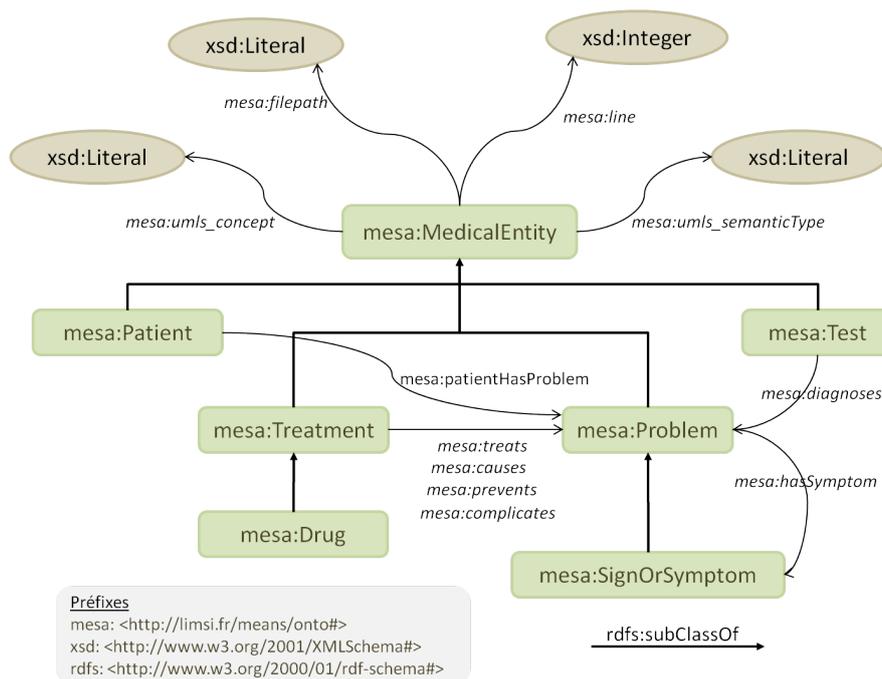
### 3.3. Ontologie de référence

Pour formaliser la question de l'utilisateur et les informations extraites des corpus médicaux nous exploitons l'ontologie de référence MESA (cf. figure 2) que nous avons construite à cet effet. L'ontologie MESA définit des concepts et des relations décrivant les fragments de texte qui seront retournés comme des réponses finales de notre SQR aussi bien que des concepts et des relations du domaine médical.

20. <http://www.semantic-systems-biology.org/biogateway>

21. <http://www4.wiwiw.fu-berlin.de/drugbank/>

L'ontologie et en particulier les concepts et les relations médicales ont été choisis après notre étude de différentes collections de questions médicales. MESA n'est pas vouée à être une ontologie de domaine « complète » mais plutôt une ontologie d'application qui représente ce qui est nécessaire pour répondre aux questions médicales fréquentes dans les classifications de la littérature. Elle représente aussi la composante textuelle du corpus qui a permis de générer les annotations, un lien utile pour pouvoir retourner des passages textuels aux utilisateurs.



**Figure 2.** MESA : ontologie de référence pour le système MEANS regroupant les entités et les relations les plus pertinentes pour le domaine médical

L'ontologie MESA représente actuellement six catégories ou classes d'entités médicales organisées hiérarchiquement par la relation `rdfs:subClassOf` et sept relations de domaine pouvant être exprimées entre les entités médicales appartenant aux catégories représentées. La granularité sémantique des classes et des relations de l'ontologie MESA limite le degré d'ambiguïté rencontré lors de l'identification des entités médicales et des relations, et permet d'augmenter la qualité du prétraitement sémantique des documents et des questions. Au niveau contenu, chaque entité médicale est associée à son concept UMLS (propriété `mesa:umls_concept`), son type sémantique UMLS (propriété `mesa:umls_semanticType`), le chemin de fichier qui contient l'entité médicale (propriété `mesa:filepath`) et le numéro de la ligne contenant l'entité dans le

fichier (propriété *mesa:line*). Les fichiers sources du corpus sont structurés de façon à contenir une phrase par ligne.

#### 4. Analyse et transformation des questions médicales en requêtes SPARQL

Dans le cadre des SQR, l'analyse et la transformation de la question en une représentation structurée n'est pas une tâche triviale. Cette tâche a été mise en évidence, entre autres, dans le cadre des interfaces en langage naturel pour bases de données.

Répondre automatiquement à des questions médicales requiert une analyse des questions différente du traitement des questions en domaine ouvert à cause des spécificités de ce domaine de spécialité. Dans cette section, nous étudions en premier lieu les caractéristiques des questions médicales, que nous essayons ensuite de classer. En second lieu, nous présentons notre méthode d'analyse de questions médicales.

##### 4.1. Caractéristiques des questions médicales

Plusieurs caractéristiques des questions médicales ont été plus ou moins mises en évidence et/ou utilisées dans les différents travaux menés sur l'analyse de questions. Certaines de ces caractéristiques sont communes avec les questions en domaine ouvert, d'autres sont plus spécifiques. Nous présentons ici quelques caractéristiques des questions médicales fondées sur notre étude des questions médicales réelles incluant des questions cliniques posées par des médecins<sup>22</sup> et des questions posées par des patients<sup>23</sup>.

1) Le type de question. Il faut différencier entre les questions Y/N (ou booléennes) et les questions WH qui peuvent être des questions de définition (*e.g. What is Depression?*), des questions listes (*e.g. What are the symptoms of blood cancer?*) ou des questions complexes (*e.g. avec Why ou How*) qui nécessitent des réponses détaillées et non pas uniquement des entités médicales.

2) Le type de la réponse attendue (TRA). Pour les questions WH, le type de la réponse attendue est introduit par le pronom ou l'adjectif interrogatif. Dans le domaine médical, ce type peut être un traitement (*e.g. What is the best treatment for Psoriasis?*), un examen médical (aussi appelé test ; *e.g. Colon Cancer: Which screening test should I have?*), etc.

3) Le focus. Le focus de la question est l'entité médicale la plus proche de la réponse attendue. Par exemple, « pyogenic granuloma » est le focus de la question *What's the best treatment for pyogenic granuloma?*.

4) La relation principale. Pour les questions factuelles, nous définissons la relation principale comme étant celle reliant la réponse attendue et le focus. Dans les ques-

22. <http://cliniques.nlm.nih.gov/>

23. [www.askthedoctor.com/topics-a-z/topics-a-z-all.html?view=category&id=8](http://www.askthedoctor.com/topics-a-z/topics-a-z-all.html?view=category&id=8)

tions booléennes, cette relation correspond à la relation la plus importante (objet de la question) entre deux entités médicales (deux focus).

5) Les entités médicales. Reconnaître les entités médicales (*e.g. headache*) et leurs catégories (*e.g. Medical Problem*) est une étape importante qui permet de déterminer le focus et les autres entités médicales indispensables pour trouver la réponse exacte. La reconnaissance des entités médicales traite plusieurs problèmes comme la grande variation terminologique du domaine médical (un terme médical peut avoir plusieurs termes synonymes, des abréviations, etc.) et aussi l'évolution continue de cette terminologie (nouveaux termes médicaux, nouvelles maladies, etc.).

6) Les relations sémantiques. Elles apportent plus de précision sur la sémantique de la question (*e.g* entre une réponse attendue de type Traitement et un focus de type problème (PB), cherche-t-on un traitement qui traite PB, le prévient, le cause ou le complique). Extraire les relations sémantiques de la question permet d'identifier la relation principale mais aussi les relations contextuelles (*e.g.* des relations autour du patient : son âge, son historique familial, etc.), point clé pour une analyse de questions efficace.

Une observation clé dans notre travail est que la définition d'un seul focus ou d'un seul type de réponse attendue limite le traitement de certains types de questions et donc la couverture de l'analyse des questions. Dans notre approche nous proposons une solution à cette restriction en construisant autant de requêtes SPARQL que de TRA et de focus dans la même question. Différents cas sont à considérer, par exemple [le nombre de focus (F) et le nombre de TRA sont indiqués] :

- *How to treat both Psoriatic Arthritis and Psoriasis?* [TRA=1,F=2]
- *How to treat Minor, moderate and Deep burns?* [TRA=1,F=3]
- *How to diagnose and manage common parasitic pneumonias?* [TRA=2,F=1]

#### 4.2. Classification des questions médicales

Nous proposons dans le tableau 1 une classification des questions médicales en six catégories. Des cas particuliers de ces catégories peuvent aussi être rencontrés. Par exemple, certaines questions factuelles recherchent une liste et non une réponse unique (*e.g.* « *What are the symptoms of Alport's Syndrome?* », « *What are the causes of hypomagnesemia?* », « *What are the clinical features and prognosis of post-concussion syndrome?* »). Les questions des différentes catégories peuvent aussi être chaînées (*e.g.* « *What is serum sickness and what are the causes?* », « *What is cerebral palsy? How do you diagnose and treat it? What is the etiology? What is the pathogenesis?* ») ou inclure une description des patients concernés (*e.g.* « *74-year-old female with memory loss and a negative workup. What is the dose of Aricept?* »).

Cat.	Exemple
<i>Yes/No : questions booléennes</i>	« <i>Can Group B streptococcus cause urinary tract infections in adults?</i> »
<i>Explication/Raison : questions « why »</i>	« <i>Why do phenobarbital and Dilantin counteract each other?</i> »
<i>Condition/cas : la majorité des questions « when »</i>	« <i>When would you use gemfibrozil rather than an HMG (3-hydroxy-3-methylglutaryl) coenzyme A inhibitor?</i> »
<i>Manière : questions « how »</i>	« <i>How are homocysteine and folic acid related to hyperlipidemia?/How can you do a screening motor exam of the hand?</i> »
<i>Définition</i>	« <i>What is seronegative spondyloarthropathy?</i> »
<i>Factuelle</i>	<p>Le type de réponse attendue est une entité médicale, une entité nommée ou plus généralement une information spécifique.</p> <p><b>Type 1 :</b> Les réponses attendues correspondent à des entités médicales (cas le plus fréquent). Questions exprimées généralement avec les pronoms « what », « which » et « how » : « <i>What is the complete workup for a meconium plug in a newborn?</i> », « <i>How should you treat osteoporosis in a man, caused by chronic steroid use?</i> », « <i>Which medication is causing neutropenia in this baby?</i> »</p> <p><b>Type 2 :</b> Autres types de réponses attendues. Questions exprimées généralement avec « when » (recherche de temps dans ce cas), where, who, et quelques-unes avec « how » (e.g. « <i>When will inhaled insulin be available?/Where do I go to work up a jaw mass?/Where would a stroke be located that involved a right facial palsy and dysarthria?</i> », « <i>How much Delsym cough syrup do I give?</i> », « <i>How often should someone have tonometry as a screen for glaucoma?</i> », « <i>How old should a patient be before I quit doing prostate specific antigens (PSA's)?</i> »</p>

**Tableau 1.** Classification des question médicales en six catégories

Dans le cadre de notre approche nous nous intéressons aux questions booléennes et aux questions factuelles de type 1<sup>24</sup>. Nous présentons dans la section suivante notre méthode pour l'analyse de ces questions, qui permet aussi le traitement de celles ayant plusieurs focus et/ou plusieurs types de réponses attendues. Cette approche se fonde sur l'extraction d'information à partir de la question, plus précisément la reconnaissance des entités médicales et l'identification des relations sémantiques. Précisons que

24. Nous avons choisi de travailler en premier lieu sur les questions booléennes et factuelles. Dans les travaux futurs, nous allons étendre le système MEANS pour traiter d'autres types de questions : Causes/Conséquences (e.g. questions avec Pourquoi/WHY), Procédures (e.g. questions avec Comment/HOW) ou encore les questions comparatives, très fréquentes dans le domaine médical (e.g. *Which is better to treat pain: ibuprofen or meloxicam?*).

notre approche de transformation des questions en requêtes SPARQL n'est pas dépendante des méthodes d'extraction d'information développées et utilisées et que d'autres outils peuvent être utilisés pour l'extraction d'information à partir de la question.

### 4.3. Méthode utilisée pour l'analyse de questions médicales

Nous avons proposé une méthode originale pour la transformation de questions médicales en requêtes SPARQL (Ben Abacha et Zweigenbaum, 2012b ; Ben Abacha et Zweigenbaum, 2012a). Cette méthode consiste à (i) extraire les informations les plus importantes à partir de la question (*e.g.* type de la question, type de la réponse attendue, entités médicales, relations sémantiques) et (ii) transformer les informations extraites en requête(s) SPARQL, le langage de requête standard pour les données RDF.

Le langage SPARQL permet d'interroger des graphes RDF. Il permet d'exprimer des requêtes recherchant des patrons de graphes RDF requis ou optionnels ainsi que leur conjonction ou leur disjonction. SPARQL inclut plusieurs fonctionnalités ou filtres pour tester les valeurs de littéraux ou d'IRI et permet de spécifier le graphe RDF source à interroger. Les résultats des requêtes SPARQL peuvent aussi bien être des ensembles de réponses (une réponse étant un ensemble d'appariements entre variables et valeurs) que des graphes RDF.

Notre méthode comporte six étapes principales :

- 1) identifier le type de la question (*e.g.* *Yes/No*, définition, factuelle, liste ou complexe) en appliquant un ensemble de patrons construits manuellement ;
- 2) déterminer le ou les types de réponses attendues pour les questions WH ;
- 3) construire la forme affirmative et simplifiée de la question (nouvelle forme)<sup>25</sup> afin d'éliminer les bruits potentiels dans les étapes suivantes ;
- 4) reconnaître les entités médicales dans la nouvelle forme de la question (section 5.1.1) ;
- 5) extraire les relations sémantiques à partir de la nouvelle forme de la question (section 5.1.2) ;
- 6) construire la ou les requêtes SPARQL correspondantes<sup>26</sup>.

Le tableau 2 présente la sortie de chaque étape sur deux exemples de questions. Ces étapes aboutissent à la construction des requêtes SPARQL *R1* et *R2* (figures 3 et 4).

Notre méthode d'analyse de questions a trois caractéristiques principales :

25. Nous construisons une forme simplifiée et affirmative de la question où la séquence de mots indiquant le type de réponse attendue est remplacée par le mot-clé ANSWER.

26. Plusieurs requêtes SPARQL sont associées à la question dans le cas de multiples TRA (*e.g.* « *How to diagnose and treat lagophthalmos patients?* », deux TRA et deux requêtes SPARQL).

Analyse de questions (Q) -> Extraction d'information	Exemples (Questions WH vs Y/N)	
	Question WH	Question Y/N
		<i>What treatment works best for constipation in children?</i>
Identification du type de la réponse attendue	Type de réponse attendue (TRA) = <i>Treatment</i>	—
Simplification et transformation de Q en forme affirmative ( <i>new_Q</i> )	<i>new_Q = What treatment ANSWER works best for constipation in children.</i>	<i>new_Q = Does spinal manipulation relieve back pain.</i>
Reconnaissance des entités médicales (en utilisant <i>new_Q</i> )	<i>ANSWER works best for &lt;PB&gt; constipation &lt;/PB&gt; in &lt;PA&gt; children &lt;/PA&gt;.</i>	<i>&lt;TX&gt; spinal manipulation &lt;/TX&gt; relieve &lt;PB&gt; back pain &lt;/PB&gt;.</i>
Extraction des relations sémantiques (en utilisant <i>new_Q</i> )	<i>treats(ANSWER,PB), avec TRA = Treatment</i>	<i>treats(TX,PB)</i>
Construction des requêtes SPARQL	Requête (R1)	Requête (R2)

**Tableau 2.** Analyse de questions médicales : deux exemples avec des annotations simplifiées (PB: problème, PA: patient, TX: traitement)

```

SELECT ?value3 ?umlsConcept3 ?file ?line WHERE {
  ?concept1 mesa:file ?file .
  ?concept1 mesa:line ?line .
  ?concept1 mesa:value "constipation" .
  ?concept1 mesa:umls_concept "Constipation" .
  ?concept1 mesa:umls_semanticType "Sign or Symptom" .
  ?concept1 mesa:category "sign_or_symptom" .
  ?concept2 mesa:file ?file .
  ?concept2 mesa:line ?line .
  ?concept2 mesa:value "children" .
  ?concept2 mesa:umls_concept "Child" .
  ?concept2 mesa:umls_semanticType "Age Group" .
  ?concept2 mesa:category "patient" .
  ?concept3 mesa:file ?file .
  ?concept3 mesa:line ?line .
  ?concept3 mesa:value ?value3 .
  ?concept3 mesa:umls_concept ?umlsConcept3 .
  ?concept3 mesa:category "treatment" .
  ?concept3 mesa:treats ?concept1 .
} GROUP BY ?umlsConcept3 ?value3

```

**Figure 3.** Requête R1 (WH) : What treatment works best for constipation in children?

```

ASK {
  ?concept1    mesa:value "spinal manipulation" .
  ?concept1    mesa:umls_concept "Manipulation of spine" .
  ?concept1    mesa:umls_semanticType "Therapeutic or Preventive Procedure" .
  ?concept1    mesa:category "treatment" .
  ?concept2    mesa:value "back pain" .
  ?concept2    mesa:umls_concept "Back Pain" .
  ?concept2    mesa:umls_semanticType "Sign or Symptom" .
  ?concept2    mesa:category "sign_or_symptom" .
  ?concept1    mesa:treats ?concept2 . }

```

**Figure 4.** *Requête R2 (Y/N) : Does spinal manipulation relieve back pain?*

1) elle permet de traiter différents types de questions, parmi lesquels les questions avec deux types de réponses attendues ou plus et/ou deux focus ou plus ;

2) elle permet une analyse profonde des questions, en utilisant plusieurs méthodes d'extraction d'information fondées sur les connaissances du domaine (*e.g.* UMLS, MeSH) et des techniques de TAL (*e.g.* utilisation de patrons ou règles, apprentissage automatique). Ces méthodes visent l'extraction des entités médicales, des relations sémantiques et aussi des informations supplémentaires ou contextuelles (*e.g.* informations sur les patients : âge, sexe, etc.) ;

3) elle est fondée sur les technologies du Web sémantique qui offrent plus d'expressivité par rapport à d'autres langages de formalisation des connaissances (*e.g.* logique du premier ordre, tables relationnelles) et fournissent des langages formels standard afin d'augmenter la portabilité des annotations relatives aux questions et aux corpus utilisés pour l'extraction des réponses.

## 5. Recherche de réponses aux questions médicales

Nous détaillons ici les étapes de recherche et d'extraction de réponses.

### 5.1. Annotation RDF hors ligne des corpus médicaux

Les performances du SQR final dépendront fortement de la qualité de l'annotation des collections de textes utilisées pour trouver les réponses. Pour cette raison, nous avons travaillé sur la reconnaissance des entités médicales mais aussi sur l'extraction des relations sémantiques reliant les entités reconnues. Pour ces deux tâches nous avons testé et mis au point des méthodes à base de règles ou patrons et des méthodes statistiques et hybrides. L'analyse et l'annotation finale des corpus comportent trois étapes principales, que nous détaillons brièvement ci-dessous :

- la reconnaissance des entités médicales ;

- l’identification des relations sémantiques reliant ces entités médicales ;
- l’annotation en RDF des entités et des relations extraites ainsi que d’autres informations (e.g. les termes MeSH des articles scientifiques).

#### 5.1.1. Reconnaissance des entités médicales

Cette étape consiste à identifier les termes médicaux et à déterminer leurs catégories (e.g. Traitement, Médicament, Problème médical, Examen ou Test médical, signe ou symptôme). Nous déterminons également les concepts de l’UMLS associés à ces termes médicaux. Utiliser les concepts UMLS permettra de retrouver des réponses au moment de l’interrogation dans les cas de synonymie. Par exemple, si l’utilisateur recherche les derniers traitements développés pour « *Type 2 Diabetes* », utiliser les concepts UMLS permettra de retrouver des réponses exprimant la même maladie avec d’autres termes (e.g. « *Diabetes mellitus type 2* » ou « *non-insulin-dependent diabetes mellitus* »). Nous travaillons sur sept catégories médicales qui ont été choisies après les analyses des différentes taxinomies de questions. Ces catégories sont : Problème médical, Traitement, Test médical, Signe ou Symptôme, Médicament, Nourriture et Patient. Le tableau 3 présente quelques catégories médicales et les types sémantiques de l’UMLS correspondants. Notre approche pour la reconnaissance des entités médicales combine deux méthodes différentes : une méthode à base de règles MetaMap Plus et une méthode statistique BIO-CRF-H. Ces méthodes ont été présentées et évaluées dans (Ben Abacha et Zweigenbaum, 2012c ; Ben Abacha et Zweigenbaum, 2011b).

Catégories	Types sémantiques de l’UMLS correspondants
<b>Problème</b>	<i>Virus, Bacterium, Anatomical Abnormality, Congenital Abnormality, Acquired Abnormality, Sign or Symptom, Pathologic Function, Disease or Syndrome, Mental or Behavioral Dysfunction, Neoplastic Process, Cell or Molecular Dysfunction, Injury or Poisoning</i>
<b>Traitement</b>	<i>Medical Device, Drug Delivery Device, Clinical Drug, Steroid, Pharmacologic Substance, Antibiotic, Biomedical or Dental Material, Therapeutic or Preventive Procedure</i>
<b>Test</b>	<i>Laboratory Procedure, Diagnostic Procedure</i>
<b>Médicament</b>	<i>Clinical Drug, Pharmacologic Substance, Antibiotic</i>

**Tableau 3.** Exemples de catégories médicales

#### 5.1.2. Extraction de relations sémantiques

Après avoir identifié les termes médicaux, nous nous intéressons aux relations sémantiques qui les relient. Cette étape est très importante pour une analyse efficace. Nous visons sept types de relations choisis après une analyse des taxinomies de questions médicales et aussi de questions médicales réelles : (i) *treats* : un traitement améliore ou traite un problème médical, (ii) *complicates* : un traitement empire un problème médical, (iii) *prevents* : un traitement prévient un problème médical, (iv) *causes* : un traitement cause un problème médical, (v) *diagnoses* : un test médical détecte, diagnostique ou évalue un problème, (vi) *DhD* : un médicament a un dosage

et (vii) *P\_hSS* : un problème se manifeste par un signe ou un symptôme. Pour extraire ces relations sémantiques, nous utilisons une combinaison de deux méthodes : une méthode à base de patrons et une méthode statistique qui utilise un classifieur SVM. L'avantage d'une méthode hybride est qu'elle permet de pallier les inconvénients des deux types de méthodes d'extraction d'information, à savoir le manque de couverture potentiel des méthodes à base de patrons et la dépendance aux données annotées des méthodes statistiques<sup>27</sup>. Nous identifions aussi des relations spécifiques aux patients : (i) son sexe, (ii) son âge et (iii) sa catégorie d'âge (adulte, adolescent, enfant, bébé) en utilisant des patrons construits manuellement. Ces méthodes ont été présentées et évaluées dans (Ben Abacha et Zweigenbaum, 2011c ; Ben Abacha et Zweigenbaum, 2011a).

### 5.1.3. Annotation en RDF des informations extraites

L'objectif de base de cette étape est d'écrire les annotations ou les informations extraites (*i.e.* entités médicales et relations sémantiques) sous forme de triplets RDF conformes à l'ontologie utilisée.

## 5.2. Recherche sémantique et classement des réponses

Les questions des utilisateurs, posées en langage naturel, sont analysées suivant le procédé indiqué dans la section 4.3 afin de construire une ou plusieurs représentations SPARQL initiales de la question selon le nombre de types de réponses attendues. Ces requêtes SPARQL sont ensuite relaxées en éliminant au fur et à mesure certaines informations. Par exemple, pour la question « *What is the best treatment for oral thrush in healthy infants?* »<sup>28</sup>, la requête SPARQL la plus précise est celle décrite dans la figure 5. Ensuite d'autres requêtes sont générées automatiquement. La figure 6 présente la requête la moins précise.

Nous définissons trois niveaux de relaxation pour la requête initiale, chacun contenant une ou plusieurs requêtes. Les requêtes sont triées dans chaque niveau suivant leur précision :

1) **niveau 1** : ce niveau regroupe d'abord la requête initiale elle-même, puis une forme relaxée de cette requête obtenue en supprimant les valeurs des entités médicales (ici : ?value1 ; ?value2 et ?value3). Par exemple, en plus de chercher une phrase qui contient *healthy infants*, on cherchera une phrase qui contient une entité médicale ayant comme concept *Infant* sans restriction sur la valeur exacte. Cette relaxation

27. L'extraction d'une relation se fait suivant l'influence (ou poids) de chaque méthode. Par exemple, le nombre d'exemples d'entraînement est utilisé pour calculer l'influence de l'approche statistique.

28. Cette question fait partie des questions utilisées pour l'évaluation. Les requêtes SPARQL présentées font partie de celles générées automatiquement par le système MEANS au cours de l'évaluation.

```

SELECT ?concept1 ?file ?line WHERE {

?concept1 <http://limsi.fr/means#file> ?file .
?concept1 <http://limsi.fr/means#line> ?line .

?concept2 <http://limsi.fr/means#file> ?file .
?concept2 <http://limsi.fr/means#line> ?line .

?concept3 <http://limsi.fr/means#file> ?file .
?concept3 <http://limsi.fr/means#line> ?line .

?concept1 <http://limsi.fr/means#concept_category> "Treatment" .

?concept2 <http://limsi.fr/means#concept_category> "Problem" .
?concept2 <http://limsi.fr/means#umls_semanticType> "Disease or Syndrome" .
?concept2 <http://limsi.fr/means#umls_concept> "Oral candidiasis" .
?concept2 <http://limsi.fr/means#value> "oral thrush" .

?concept3 <http://limsi.fr/means#concept_category> "Patient" .
?concept3 <http://limsi.fr/means#umls_semanticType> "Age Group" .
?concept3 <http://limsi.fr/means#umls_concept> "Infant" .
?concept3 <http://limsi.fr/means#value> "healthy infants" .

?concept1 <http://limsi.fr/means#treats> ?concept2 .
?concept3 <http://limsi.fr/means#patient_has> ?concept2 .

}

```

**Figure 5.** La requête SPARQL la plus précise construite automatiquement pour la question : « What is the best treatment for oral thrush in healthy infants? »

peut diminuer légèrement la précision (le fait de garder le concept UMLS garantit de chercher la bonne entité médicale) mais elle augmente beaucoup le rappel ;

2) **niveau 2** : supprimer les entités médicales une à une (en gardant la réponse attendue et le focus<sup>29</sup>) ;

3) **niveau 3** : supprimer la ou les relations principales<sup>30</sup>.

Ces requêtes SPARQL sont ensuite exécutées dans l'ordre pour interroger les triplets RDF générés dans la phase d'annotation de corpus. Les réponses sont ainsi triées par le tri des requêtes. Un tri supplémentaire est effectué sur les réponses dans le cas des questions factuelles. Pour ces questions, une réponse à une requête est privilégiée si elle a plus de justifications que les autres. Deux réponses sont jugées identiques

29. Un focus est identifié comme étant une entité médicale reliée à la réponse attendue par une relation.

30. Une relation principale est identifiée comme étant une relation ayant la réponse attendue pour objet ou sujet.

```

SELECT ?concept1 ?file ?line WHERE {
?concept1 <http://limsi.fr/means#file> ?file .
?concept1 <http://limsi.fr/means#line> ?line .

?concept2 <http://limsi.fr/means#file> ?file .
?concept2 <http://limsi.fr/means#line> ?line .

?concept1 <http://limsi.fr/means#concept_category> "Treatment" .

?concept2 <http://limsi.fr/means#concept_category> "Problem" .
?concept2 <http://limsi.fr/means#umls_concept> "Oral candidiasis" .
}

```

**Figure 6.** La requête SPARQL la moins précise construite automatiquement pour la question : « What is the best treatment for oral thrush in healthy infants? »

si les entités retournées ont le même CUI<sup>31</sup> UMLS (*e.g. the corticosteroid injection, corticosteroid injections* : C2095490). Les justifications sont ensuite groupées par CUI différent et comptées.

Le tableau 4 présente quelques exemples de modèles de questions et de leurs réponses.

Questions	Exemples	Formes simplifiées	Réponses
Definition	What is X	X	X is ...
Yes/No	Can X REL Y	X REL Y	Yes/No
Factuelle ou liste : TRA et REL connus	How can you REL Y ou What is X REL Y	ANSWER? REL Y (on supprime X)	X'? REL Y avec X' is a X
Factuelle ou liste : TRA et REL inconnus	What is X REL Y	X REL Y    ANSWER? REL Y	X'? REL Y avec X' is a X

**Tableau 4.** Questions vs réponses

La nature des réponses retournées par le système **MEANS** varie suivant le type de la question :

- question définition : réponse = une phrase ;
- question booléenne : réponse = oui ou non ;
- question factuelle : réponse = entité médicale (*e.g. un traitement*) ou une information précise (*e.g. la dose d'un médicament*) ;

31. Concept Unique Identifier

– question liste : réponse = une liste d’entités médicales (*e.g.* liste de symptômes).

En plus de la réponse, un extrait de quatre phrases entourant cette réponse est précisé. Nous utilisons le moteur de recherche sémantique Jena<sup>32</sup> pour la lecture et l’interrogation des annotations RDF.

## 6. Évaluation du système de questions-réponses MEANS

Dans cette section, nous présentons les critères qui entrent en jeu pour évaluer les SQR en domaine ouvert et en domaine médical. Ensuite nous présentons les données d’évaluation ainsi que les résultats obtenus par le système **MEANS**.

### 6.1. Critères et mesures d’évaluation : performances vs rapidité

Plusieurs campagnes d’évaluation de SQR ont été menées en domaine ouvert. Citons par exemple TREC<sup>33</sup> (anglais), CLEF<sup>34</sup> (multilingue), NTCIR<sup>35</sup> (japonais) et Quaero<sup>36</sup> (français, anglais). Pour le domaine médical, les défis de SQR anglais sont rares. La tâche Genomics du défi TREC peut être citée comme une piste exploitable pour les tâches de QR en anglais, bien que non officiellement introduite comme telle. Nous avons néanmoins pu collecter un corpus « standard » préparé par des experts du domaine pour la tâche de recherche de réponses à des questions en anglais<sup>37</sup> (cf. section 6.2).

La performance des SQR est souvent évaluée en mesurant le MRR (*Mean Reciprocal Rank*) et potentiellement la précision et le rappel des réponses retournées, en supposant qu’un système renvoie une liste ordonnée de réponses, éventuellement avec un nombre maximal fixé de réponses :

- le MRR : la moyenne des réciproques du rang de la première bonne réponse (1 si la première réponse est correcte, 0,5 si la deuxième est la première correcte, etc.) ;
- le rappel : mesure la proportion des réponses correctes retournées par le système parmi l’ensemble des réponses correctes ;
- la précision : mesure la proportion de réponses correctes trouvées parmi toutes les réponses trouvées par le SQR.

Ces mesures demandent de bien définir ce qu’est une réponse « correcte », et de quels critères la réponse doit tenir compte.

32. <http://jena.sourceforge.net/>

33. <http://trec.nist.gov>

34. <http://clef.isti.cnr.it>

35. <http://research.nii.ac.jp/ntcir>

36. <http://www.quaero.org>

37. Nous remercions Dina Demner-Fushman de nous avoir indiqué cette piste.

La réponse à ces questions peut passer par les deux points clés suivants :

1) **la spécificité de la réponse** : en domaine ouvert : à une question comme « *Où se trouve le musée de Louvre ?* », plusieurs réponses sont possibles : (i) *Le musée du Louvre se trouve en France*, (ii) *Le musée du Louvre se situe en plein centre de Paris*, (iii) *Le musée du Louvre se trouve dans un ancien palais royal...* On peut ainsi distinguer plusieurs évaluations des réponses qui peuvent être jugées, par exemple, correctes, complètes, incomplètes ou fausses. Le point d'ambiguïté dans cet exemple étant surtout que le type de réponse attendue n'est pas précis, il peut aussi bien s'agir du nom d'un quartier, d'une ville, d'un pays ou d'une adresse complète. En domaine médical, le problème se pose aussi pour les questions ouvertes (e.g. « *To what extent can we say that cell phones are harmful?* ») mais est plus restreint pour les questions fermées (e.g. factuelles, liste), où le type de réponse attendue est explicite, ou pour les questions booléennes ;

2) **la justification de la réponse** : la justification est l'extrait textuel (e.g. phrase) contenant la réponse extraite. Deux cas sont possibles : (i) la réponse peut être correcte avec une justification fautive, mais aussi (ii) la réponse peut être fautive aujourd'hui même si la justification est correcte. Par exemple « *Qui est le président de la France* » ? la réponse « Jacques Chirac » est fautive même si la justification « Discours de M. Jacques Chirac, président de la République française, devant l'université de Beïda. » est correcte.

En plus de la performance, un autre critère est important dans le cadre des SQR : **la rapidité**. Ainsi, dans le domaine médical, Takeshita *et al.* (2002) ont montré que les médecins<sup>38</sup> ont besoin d'accéder à l'information en moins de trente secondes et abandonnent la recherche au-delà.

## 6.2. Données d'évaluation

Pour évaluer notre système, nous utilisons le corpus de questions-réponses construit par Mollá et Santiago-Martínez (2011)<sup>39</sup>. Ce corpus (*Corpus for Evidence Based Medicine Summarisation*) est une collection de résumés d'articles regroupés par question et provenant de la section des questions cliniques du *Journal of Family Practice*<sup>40</sup>. Il est écrit au format XML et annoté avec les éléments suivants : (i) la question clinique, (ii) la ou les réponses à la question, (iii) les passages justifiant les réponses extraits du *Journal of Family Practice* et (iv) la référence aux articles par leurs identifiants PubMed.

38. Cette étude a été effectuée avec un groupe de cinq médecins de famille, cinq résidents et cinq médecins internes.

39. Ce corpus d'évaluation est différent de celui utilisé pour évaluer notre méthode d'analyse de questions (Ben Abacha et Zweigenbaum, 2012b) où nous avons utilisé uniquement des questions médicales extraites du JFP et qui n'ont pas de réponses associées.

40. Le corpus est téléchargeable à cette adresse : <http://sourceforge.net/projects/ebmsumcorpus>

Nous évaluons un ensemble de cinquante questions médicales de la collection de Mollá et Santiago-Martínez (2011). Cet ensemble correspond aux vingt questions booléennes et aux dix-neuf questions factuelles qui ont une sémantique qui peut être exprimée avec notre ontologie. Les onze questions restantes emploient des entités médicales et/ou des relations qui ne sont pas couvertes par notre système. Les réponses (extraits de texte), qui n'existent pas dans les articles et qui étaient préparées manuellement à partir des différentes justifications trouvées dans les articles, nous ont servi pour évaluer manuellement les réponses retournées par le système MEANS. Une évaluation automatique n'est pas possible car il faut évaluer les justifications retournées par le système (les phrases qui contiennent les réponses) manuellement. Une réponse est considérée correcte si l'entité médicale ou la valeur booléenne retournée est correcte et si la justification est correcte aussi.

### 6.3. Questions booléennes

Deux éléments entrent en jeu pour l'évaluation des questions booléennes : la valeur de la réponse : oui ou non et sa justification. Nous mesurons donc la précision de notre SQR pour ces questions par rapport à la valeur de la réponse (oui/non) et par rapport à la pertinence de sa justification.

Le tableau 5 présente les résultats obtenus pour les vingt questions booléennes traitées. *N1*, *N2* et *N3* désignent les niveaux de relaxation employés. Pour les questions booléennes, nous avons évalué les réponses obtenues par le niveau 1 puis celles obtenus par les trois niveaux en même temps.

Nous désignons par « autres » les questions booléennes dont les relations principales sont différentes de celles que nous traitons, par exemple *have a role* (*Does routine amniotomy have a role in normal labor?*), *tolerated* (*Are any oral iron formulations better tolerated than ferrous sulfate?*).

La précision et le rappel sur les questions booléennes sont ainsi de **45%** si l'on applique uniquement le niveau 1 de relaxation et de **60%** si l'on applique le niveau 3. La précision et le rappel ont la même valeur ici car si le système retourne 'non' à une question et que la réponse correcte était 'oui' il n'y a pas de justifications à évaluer et cela compte pour une erreur et dans le cas où le système retourne correctement la valeur 'oui' nous n'avons pas eu de justifications erronées.

Nous distinguons quatre principaux types d'erreurs :

- **T1** : la réponse n'existe pas dans le corpus d'articles dont nous disposons (quelques fichiers ne contiennent parfois que le titre de l'article) ;
- **T2** : les réponses ou les justifications sont sur deux phrases ou plus (*e.g.* description des résultats d'une expérimentation) ;
- **T3** : une entité ou une relation importante n'a pas été reconnue. C'est le cas, par exemple, des questions suivantes :

- *Does reducing smoking in the home protect children from the effects of second-hand smoke?*

- *What is the appropriate use of sunscreen for infants and children?*

- *Do preparticipation clinical exams reduce morbidity and mortality for athletes?*

– **T4** : des questions nécessitent des connaissances externes ou des inférences, par exemple :

- *Does heat or cold work better for acute muscle strain?* (*cold* n'est pas un problème ici, mais plutôt « *Cold Therapy* » ou « *Cryotherapy* », pareil pour *heat* (« *heat therapy* »))

- *Does psychiatric treatment help patients with intractable chronic pain?* (les traitements cités dans les articles peuvent être des sous types de « *psychiatric treatment* », par exemple : « *Cognitive therapy* »)

- *Do antiarrhythmics prevent sudden death in patients with heart failure?* (les médicaments cités dans les articles peuvent être des sous-types, par exemple : « *beta blockers* » sont des « *antiarrhythmic drugs* »)

Types	Nombre de questions	N1	N1,N2,N3	Types d'erreurs
<i>Treats</i>	9	5	5	T1, T2, T4
<i>Prevents</i>	4	0	3	T3, T4
<i>Diagnoses</i>	1	1	1	–
<i>Autres</i>	6	3	4	T3, T4
<i>Total</i>	20	9	12	–

**Tableau 5.** Nombre de questions booléennes auxquelles il a été répondu correctement et types d'erreurs par catégorie

Pour certaines questions la réponse est indiquée comme existante mais les articles qui la contiennent sont fournis uniquement avec leur titre sans leur contenu. Ce problème a influencé les résultats des questions de type « *treats* » même si elles étaient, en général, correctement analysées. Aussi, pour certaines questions de type « *treats* », le problème était que la réponse était étalée sur plusieurs lignes. Par exemple pour la question « *Does yoga speed healing for patients with low back pain?* », la justification<sup>41</sup> dans le corpus était : « *In a case series, 21 women aged >60 years (mean age, 75) with hyperkyphosis, participated in twice-weekly 1-hour sessions of hatha yoga for 12 weeks. Measured height increased by a mean of 0.52 cm, forward curvature diminished, patients were able to get out of chairs faster, and they had longer functional reach. Eleven patients (48%) reported increased postural awareness/improvement and improved well-being; 58% perceived improvement in their physical functioning.* »

Pour les erreurs de type T3, la relaxation des requêtes a amélioré les résultats (cas des questions de type *prevents* ou autres). En effet, elle a permis de trouver davantage

41. Cette justification a été prise du fichier de référence qui contient les réponses et les justifications, pour nous, le fichier en question (12356608) comporte uniquement : « *Yoga for women with hyperkyphosis: results of a pilot study.* »

d'éléments qui répondaient aux requêtes SPARQL et qui menaient de ce fait à des réponses positives (*oui*). Ces réponses positives se trouvaient être correctes alors que sans relaxation le système répondait *non* par absence d'information. Nous n'avons pas eu de bruit car la relaxation conserve au moins deux entités de la question (une hypothèse qui est peut-être un peu stricte pour certaines questions mais qui garantit un minimum de précision).

Dans certains cas, même si la question est analysée correctement et que la requête qui lui est associée est correcte, le système n'a pas trouvé la bonne réponse car des inférences ou des connaissances externes étaient nécessaires. Par exemple, pour la question « *Do antiarrhythmics prevent sudden death in patients with heart failure?* », les réponses étaient par exemple : (i) *Beta-blockers to reduce mortality in patients with systolic dysfunction: a meta-analysis*, (ii) *Beta-blockers are particularly effective in people with a high sympathetic drive (i.e., high pulse rates) to lower blood pressure and reduce cardiovascular risk*. Ainsi, pour retrouver la réponse (i) il fallait inférer que « *heart failure* » est un type de « *systolic dysfunction* », que les « *Beta-blockers* » sont des « *antiarrhythmics* » et que les relations « *reduce mortality* » et « *prevent (sudden) death* » sont suffisamment similaires.

#### 6.4. Questions factuelles

Une réponse à une question factuelle est jugée correcte si la bonne entité médicale est retournée (CUI correct) avec une justification correcte. Ainsi, les entités médicales incomplètes (*e.g.* un mot en moins) ou qui comportent du bruit ont été considérées comme fausses et les réponses correctes avec de fausses justifications ont été également considérées comme fausses. Le tableau 6 présente les résultats obtenus sur dix-neuf questions factuelles du corpus de référence. *N1*, *N2* et *N3* désignent les niveaux de relaxation 1, 2 et 3.

Types	Nombre de questions	N1		N1 + N2 + N3	
		MRR	P@5	MRR	P@5
<i>Treats</i>	8	0,625	70,58	1	62,5
<i>Prevents</i>	1	0	–	1	60
<i>Diagnoses</i>	5	0	–	0,432	25
<i>Causes</i>	1	0	–	1	20
<i>Manages</i>	3	0,66	100	0,5	80
<i>2 TRAs ou plus</i>	1	0	–	1	66,66
<b>Total</b>	<b>39</b>	<b>0,42</b>	<b>85,71</b>	<b>0,77</b>	<b>57,47</b>

**Tableau 6.** Questions factuelles : MRR et précision à cinq réponses (en %)

La précision finale sur toutes les questions est de 85,71% sans relaxation avec un MMR de 0,42. Le MRR a augmenté de 0,35 avec la relaxation qui a aussi permis de

retrouver plus de réponses. Cependant la précision s'est dégradée avec la relaxation, ce qui était attendu, mais la perte de 0,28 point de précision au total reste acceptable au vu de l'augmentation du nombre de réponses et surtout au vu de l'augmentation du MRR. Les résultats obtenus pour chaque question factuelle sont détaillés dans le tableau 7.

Pour un exemple concret, nous présentons les résultats obtenus pour la question Q1 du tableau 7 : « *What is the best treatment for oral thrush in healthy infants?* ». Sans relaxation, deux réponses correctes sont obtenues pour cette question « *Nystatin* » et « *Nystatin Suspension* »<sup>42</sup>. Avec le niveau 1 de relaxation (suppression des valeurs textuelles exactes) nous obtenons une réponse supplémentaire et correcte « *Fluconazole* ». Le niveau 2 de relaxation (ici, suppression de l'entité « *healthy infants* ») poursuit dans le même sens en permettant d'obtenir une nouvelle réponse correcte « *Gentian Violet* ». Le niveau 3 de relaxation (suppression de la relation principale, ici *treats*) apporte cinq nouvelles réponses dont une seule correcte « *Miconazole Gel* ». Notons que le tableau détaillé 7 s'arrête à cinq réponses et ne montre donc pas toutes les réponses correspondant à cet exemple, mais uniquement les cinq premières qui sont ici obtenues dès la première réponse du niveau 3 de relaxation et qui est ici une réponse correcte.

## 6.5. Discussion

Les résultats montrent l'intérêt de la relaxation pour l'amélioration des performances du système **MEANS** : augmentation de la précision pour les questions booléennes et du nombre de réponses et du MRR pour les questions factuelles. La précision a augmenté pour les questions booléennes car l'absence d'information due à la forme initiale, fortement contrainte, de la requête est interprétée comme une réponse négative (non) alors que les réponses positives et correctes sont présentes dans le corpus et accessibles avec la relaxation. Les premières observations des cas d'erreur pour les questions booléennes mettent en évidence qu'il est nécessaire de prendre en compte plus efficacement la présence de négations dans les phrases car leur impact est décisif pour les questions booléennes (e.g. « *Another Cochrane review found no added benefit in function from combining deep transverse friction massage with ultrasound or a placebo ointment* »). Aussi déterminer le niveau de certitude joue un rôle important dans la sélection des bonnes justifications. Par exemple, l'annotation de la phrase « *There's insufficient evidence to support specific physiotherapy methods or orthoses (braces), shock wave therapy, ultrasound, or deep friction massage* » a mené à une mauvaise justification car le niveau de certitude n'a pas été pris en compte.

Les principales observations pour les questions factuelles sont les défaillances détectées dans les cas de comparaison qui ne sont pas encore pris en compte par le système **MEANS**. Par exemple, la question « *What are the most effective treatments*

42. Nystatin et Nystation suspensions ont deux CUI UMLS différents, donc il s'agit bien de deux entités médicales différentes.

Catégorie	Q	P@5 (x)		RR		P@5/catégorie		MRR/catégorie	
		N1	N1, 2, 3	N1	N1, 2, 3	N1	N1, 2, 3	N1	N1, 2, 3
Treats	Q1	100 (3)	100 (5)	1	1	70,58	62,5	0,625	1
	Q2	na (0)	40 (5)	1	1				
	Q3	60 (5)	60 (5)	1	1				
	Q4	na (0)	100 (5)	0	1				
	Q5	na (0)	60 (5)	0	1				
	Q6	100 (2)	60 (5)	1	1				
	Q7	80 (5)	80 (5)	1	1				
	Q8	na (0)	100 (5)	0	1				
Diagnoses	Q9	na (0)	20 (5)	0	0,33	na (0)	0,25	0	0,432
	Q10	na (0)	0 (5)	0	0				
	Q11	na (0)	40 (5)	0	0,5				
	Q12	na (0)	20 (5)	0	1				
	Q13	na (0)	50 (4)	0	0,33				
Prevents	Q14	na (0)	60 (5)	0	1	na (0)	60	0	1
Manages	Q15	na (0)	na (0)	0	0	100	80	0,66	0,5
	Q16	100 (3)	80 (5)	1	1				
	Q17	100 (3)	80 (5)	1	0,5				
Causes	Q18	na (0)	20 (5)	0	1	na (0)	20	0	1
2 TRA	Q19	na (0)	66,66 (3)	0	1	na (0)	66,66	0	1

**Tableau 7.** Résultats détaillés par question et catégorie.  $P@5(x)$ : précision à cinq réponses et  $x$  nombre de réponses trouvées (nombre maximal fixé à cinq). *na* : « no answer », la précision ne peut pas être calculée.  $N_i$  : niveau de relaxation  $i$ .

for *PB*? » qui cherche les meilleurs traitements à une maladie *PB* ne peut pas avoir comme réponse un traitement *T1* avec la justification « *T1 is less effective than T2* ». *T1* peut être accepté comme réponse correcte si des justifications contraires existent car parfois des études aboutissent à des résultats contradictoires (c'est un cas que nous avons eu dans notre évaluation).

## 7. Conclusion

Nous avons présenté le SQR médical MEANS. Le système se fonde sur une analyse sémantique des corpus et des questions. Il construit plusieurs requêtes SPARQL pour représenter la question de l'utilisateur et les trie par ordre de précision (spécificité) décroissante avant de les exécuter. Les résultats obtenus montrent que ce tri au niveau des requêtes permet d'ordonner efficacement les réponses retournées. Les résultats montrent aussi que la diminution de la précision pour les questions factuelles due à la relaxation reste acceptable comme petite contrepartie de l'apport de nouvelles réponses et de l'amélioration du MRR. Cependant, malgré la généralité de la méthode d'analyse de questions proposée, il reste des questions auxquelles nous n'avons pas

répondu car elles ont des types de réponses attendues non encore traités. Dans ce cadre, nous envisageons d'améliorer la couverture du système en traitant de nouveaux types de questions. Au niveau de la recherche de réponses, nous travaillons sur l'interrogation de bases de connaissances RDF locales ou distantes et la combinaison des résultats avec les réponses obtenues à partir des corpus textuels utilisés.

## 8. Bibliographie

- Athenikos S. J., Han H., « Biomedical question answering: A survey », *Computer Methods and Programs in Biomedicine*, vol. 99, n° 1, p. 1-24, 2010.
- Belleau F., Nolin M.-A., Tourigny N., Rigault P., Morissette J., « Bio2RDF: Towards a mashup to build bioinformatics knowledge systems », *Journal of Biomedical Informatics*, vol. 41, n° 5, p. 706 - 716, 2008.
- Ben Abacha A., Zweigenbaum P., « A Hybrid Approach for the Extraction of Semantic Relations from MEDLINE Abstracts », *Computational Linguistics and Intelligent Text Processing, 12th International Conference, CICLing 2011*, vol. 6608 of *Lecture Notes in Computer Science*, Tokyo, Japan, p. 139-150, 2011a.
- Ben Abacha A., Zweigenbaum P., « Medical Entity Recognition: A Comparison of Semantic and Statistical Methods », *Actes BioNLP 2011 Workshop*, Association for Computational Linguistics, Portland, Oregon, USA, p. 56-64, 2011b.
- Ben Abacha A., Zweigenbaum P., « Une approche hybride pour la détection automatique des relations sémantiques entre entités médicales », *Journées francophones d'informatique médicale (JFIM)*, Tunis, Tunisie, 2011c.
- Ben Abacha A., Zweigenbaum P., « Analyse et transformation des questions médicales en requêtes SPARQL », *CORIA (Conférence en Recherche d'Informations et Applications)*, Bordeaux, 2012a.
- Ben Abacha A., Zweigenbaum P., « Medical Question Answering: Translating Medical Questions into SPARQL queries », *ACM SIGHIT International Health Informatics Symposium (IHI 2012)*, Miami, FL, USA, January, 2012b.
- Ben Abacha A., Zweigenbaum P., « Une étude comparative empirique sur la reconnaissance des entités médicales », *Traitement Automatique des Langues (TAL)*, vol. 53, n° 1, p. 39-68, 2012c.
- Cairns B. L., Nielsen R. D., Masanz J. J., Martin J. H., Palmer M. S., Ward W. H., Savova G. K., « The MiPACQ Clinical Question Answering System », *AMIA Annu Symp Proc*, p. 171-180, 2011.
- Demner-Fushman D., Lin J., « Knowledge Extraction for Clinical Question Answering: Preliminary Results », *Actes AAAI 2005 Workshop on Question Answering in Restricted Domains*, AAAI, 2005.
- Demner-Fushman D., Lin J. J., « Answer Extraction, Semantic Clustering, and Extractive Summarization for Clinical Question Answering », *ACL*, 2006.
- Ely J. W., Osherooff J. A., Gorman P. N., Ebell M. H., Chambliss M. L., Pifer E. A., Stavri P. Z., « A taxonomy of generic clinical questions: classification study », *British Medical Journal*, vol. 321, p. 429-432, 2000.

- Embarek M., Un système de question-réponse dans le domaine médical : le système Esculape, Thèse de doctorat, Université Paris-Est, 2008.
- Engelbrecht R., « Expert systems for medicine—functions and developments », *Zentralbl Gynakol*, vol. 119, n° 9, p. 428-434, 1997.
- Green Jr. B. F., Wolf A. K., Chomsky C., Laughery K., « Baseball: an automatic question-answerer », *Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference*, IRE-AIEE-ACM '61 (Western), ACM, p. 219-224, 1961.
- Hotvedt M. O., « Continuing medical education: actually learning rather than simply listening », *JAMA*, 275:1638, vol. 275, n° 21, p. 1637-1638, 1996.
- Huang X., Lin J., Demner-Fushman D., « Evaluation of PICO as a knowledge representation for clinical questions », *AMIA Annu Symp Proc*, p. 359-363, 2006.
- Jacquemart P., Zweigenbaum P., « Towards a Medical Question-Answering System: a Feasibility Study », in R. Baud, M. Fieschi, P. Le Beux, P. Ruch (eds), *Actes Medical Informatics Europe*, vol. 95 of *Studies in Health Technology and Informatics*, IOS Press, Amsterdam, p. 463-468, 2003.
- Katz B., « From sentence Processing to Information Access on the World Wide Web », *AAAI Spring Symposium on Natural Language Processing for the World Wide Web*, 1999.
- Katz B., Felshin S., Yuret D., Ibrahim A., Lin J. J., Marton G., McFarland A. J., Temelkuran B., « Omnibase: Uniform Access to Heterogeneous Data for Question Answering », *NLDB*, p. 230-234, 2002.
- Lopez V., Uren V. S., Sabou M., Motta E., « Is Question Answering fit for the Semantic Web?: A survey. », *Semantic Web*, vol. 2, n° 2, p. 125-155, 2011.
- Moldovan D., Paca M., Harabagiu S., Harabagiu A., Surdeanu M., « Performance Issues and Error Analysis in an Open-Domain Question Answering System », *ACM Trans. Information Systems*, 21(2), p. 133-154, 2003.
- Mollá D., Santiago-Martínez M. E., « Development of a Corpus for Evidence Medicine Summarisation », *Australasian Language Technology Workshop (ALTA 2011)*, Australia, 2011.
- Niu Y., Hirst G., « Analysis of semantic classes in medical text for question answering », *Proceedings of the ACL-2004 Workshop Question Answering in Restricted Domains*, 2004.
- Niu Y., Hirst G., McArthur G., Rodriguez-Gianolli P., « Answering clinical questions with role identification », *Proceedings of the ACL 2003 workshop on Natural language processing in biomedicine - Volume 13*, BioMed '03, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 73-80, 2003.
- Popescu A., Etzioni O., Kautz H., « Towards a theory of natural language interfaces to databases », *Proceedings of the International Conference on Intelligent User Interfaces (IUI'03)*, p. 149-157, 2003.
- Rinaldi F., Dowdall J., Schneider G., « Answering Questions in the Genomics Domain », *Proc. ACL04 Workshop on Question Answering in Restricted Domains*, 2004.
- Sackett D. L., Straus S. E., Richardson W. S., Rosenberg W., Haynes R. B., *Evidence-Based Medicine: How to Practice and Teach EBM*, Churchill Livingstone, Edinburgh, 2000.
- Takeshita H., Davis D., Straus S. E., « Clinical evidence at the point of care in acute medicine: a handheld usability case study », *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, p. 1409-1413, 2002.

Terol R. M., Martínez-Barco P., Palomar M., « A knowledge based method for the medical question answering problem », *Computers in Biology and Medicine*, vol. 37, n° 10, p. 1511-1521, 2007.

Woods W. A., « Progress in natural language understanding: an application to lunar geology », *Proceedings of the June 4-8, 1973, national computer conference and exposition*, AFIPS '73, ACM, New York, NY, USA, p. 441-450, 1973.