

## Questions-réponses dans le domaine médical : une approche sémantique

Asma Ben Abacha

LIMSI-CNRS BP 133 91403 Orsay Cedex - France.

Contact : asma.benabacha@limsi.fr

---

### Résumé

Ce travail porte sur l'étude de méthodes permettant de répondre à des questions formulées en langue naturelle dans le domaine médical. En premier lieu nous étudions dans quelle mesure les méthodes utilisées en domaine ouvert sont transposables à ce domaine. Dans un second lieu, nous proposons une nouvelle approche, plus adaptée à ce domaine de spécialité. Cette approche se base sur (i) des ressources sémantiques disponibles en domaine médical (le métathésaurus et le réseau sémantique d'UMLS) et (ii) une analyse syntaxique et sémantique de la question et des corpus médicaux afin de les représenter sous forme de graphes sémantiques. Les avantages soulignés dans cette première étude sont les apports en termes d'expressivité et d'évolutivité pour le système de questions-réponses mais aussi l'amélioration de la précision des réponses.

### Abstract

We study how to answer questions in natural language in the medical domain. At first, we analyse the methods used in open domain and how much they comply with the specificities of the medical domain. In a second step, we propose a new approach, more adapted to this field. Our approach is based on (i) semantic knowledge bases (UMLS's Methathesaurus and Semantic Network) and (ii) a semantic and syntactic analysis of the question and the medical corpora from which the answer will be extracted. As a result of this analysis we construct semantic graphs representing the question and the documents. The advantages we underline in this first study are the enhancements in expressivity and evolutivity for the question-answering system as well as in the answers' precision.

**Mots-clés :** systèmes de questions-réponses, domaine médical, graphes sémantiques.

**Keywords:** question-answering systems, medical domain, semantic graphs.

---

### 1. Introduction

Le domaine médical est un domaine où les connaissances augmentent de façon considérable d'année en année ; elles doubleraient même tous les cinq ans [1]. Certaines études montrent que les médecins se posent en moyenne 3,2 questions pour chaque dix patients [2] mais ne cherchent la réponse qu'à une minorité de leurs questions (35%) [3]. L'obstacle le plus important à cette recherche est la défaillance de la ressource utilisée à fournir une réponse (26%) [4].

Des outils de recherche automatiques et efficaces sont donc de plus en plus indispensables pour le domaine médical. Cependant des outils généralistes tels les moteurs de recherche actuels ont le plus souvent une granularité de l'ordre du document et sont donc inappropriés pour des besoins immédiats en information précise. Les systèmes de questions-réponses (SQR) visent une granularité plus fine et tentent de répondre en quelques mots ou en une ou quelques phrases à des questions formulées en langue naturelle. Ces systèmes ont été jusqu'ici principalement développés en domaine ouvert dans le but de répondre à des questions générales en exploitant un corpus fermé (e.g. articles de presse) ou le Web.

Ce travail porte sur l'étude de méthodes permettant de rechercher des réponses à des questions en anglais dans une base de documents en anglais dans le domaine médical : dans quelle mesure les méthodes utilisées dans des systèmes en domaine ouvert sont-elles transposables à ce domaine ?

Les conditions différentes rendent-elles possibles ou nécessaires la conception de méthodes nouvelles ?

Le présent article s'articule autour de sept sections. Nous présentons les SQR dans la deuxième section. Dans la section 3, nous nous intéressons au domaine médical, nous exposons les spécificités de ce domaine et la problématique qui en découle ainsi qu'un petit aperçu sur l'existant. La section 4 est consacrée à la présentation de notre approche qui se fonde sur la représentation sémantique de la question et des documents interrogés sous forme de graphes. La cinquième section décrit la méthode que nous proposons pour la construction automatique des graphes sémantiques. Dans la sixième section nous présentons le processus d'évaluation, en cours, de notre approche. Enfin, dans la section 7 nous tirons les conclusions en soulignant les avantages des méthodes proposées pour ce domaine de spécialité.

## 2. Système de Questions-Réponses (SQR)

Un SQR est un système capable de répondre à des questions posées en langage naturel. L'architecture d'un tel système, plus complexe que celle d'un moteur de recherche classique, est composée en général de 3 modules : (i) analyse de la question, (ii) recherche de documents et (iii) extraction de la réponse (cf. figure 1).

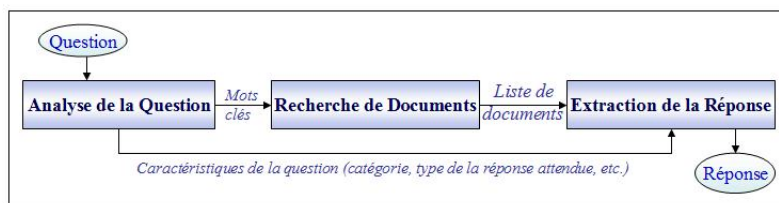


FIG. 1 – Architecture classique d'un système de questions-réponses

L'analyse de la question s'intéresse à déduire à partir de la question toutes les informations utiles pour les modules qui suivent. Il s'agit, entre autres, d'extraire des mots clés de la question ainsi que déterminer sa catégorie (par rapport à une classification déjà établie) et le type de la réponse attendue (personne, date, etc.). Certains travaux cherchent à déterminer une autre caractéristique de la question, appelée *focus*, désignant le mot ou l'expression de la question qui a le plus de chances d'être présent(e) dans la réponse (e.g. La question " *Qui a inventé le télescope ?* " attend une réponse de type Personne et a comme focus le mot *télescope*).

Le module de recherche fouille une collection de documents pour extraire les documents ou les passages susceptibles de contenir la réponse. Enfin, le troisième module utilise les informations déduites de l'analyse de la question pour extraire, depuis les documents retrouvés, des réponses candidates (habituellement courtes : quelques mots, une phrase ou un passage). Une étape finale consistera à trier les réponses candidates suivant leur pertinence par rapport à la question.

## 3. SQR pour le domaine médical : problématique et état de l'art

Pour mettre en évidence la problématique des SQR en domaine médical, nous examinons les différences entre un SQR en domaine ouvert et un SQR spécialisé pour des questions médicales. Nous présentons ensuite quelques travaux réalisés dans le domaine médical.

### 3.1. Problèmes spécifiques posés par le domaine médical pour un SQR

Nous avons testé le SQR généraliste FIDJI [5], développé au LIMSI, sur une collection de questions médicales (la liste *EQueR* médicale) et un corpus médical. Nous avons ensuite analysé les résultats fournis, les questions et des extraits du corpus. En se basant sur ces analyses, nous avons relevé plusieurs différences entre le domaine ouvert et le domaine médical que nous classons en quatre catégories.

**Entités nommées (EN).** Il s'agit d'unités textuelles désignant un objet précis. Les entités nommées classiques désignent les noms de personnes, de lieux, d'organisations, les dates et les unités monétaires. Pour le domaine médical, les EN correspondent aux termes du domaine, à savoir les noms de maladies, de médicaments, etc. La reconnaissance des EN, qui consiste en leur identification ainsi que leur catégorisation, est une étape essentielle lors de l'analyse de la question et l'extraction de la réponse.

**Classification de la question et détermination du type de la réponse attendue.** En domaine ouvert, la classification se base principalement sur le pronom interrogatif (qui, quand, etc.) et sur les EN de la question. Elle permet de déterminer le type de la réponse attendue. Par exemple, la question « *Quand se déroule le bac 2009 ?* » attend une réponse de type *Date*. Cependant, une classification des questions basée sur le pronom interrogatif n'est pas appropriée au domaine médical. Une question avec *Quand*, par exemple, peut avoir plusieurs types de réponses attendues. Dans la question « *Quand devrais-je prendre de l'Hiconcil ?* », le pronom *Quand* réfère à des conditions d'utilisation d'un médicament et dans « *Quand devrais-je prendre mon antibiotique ?* », *Quand* réfère à un temps relatif (une heure après le repas, etc.). D'autre part, les questions médicales portent souvent sur des connaissances médicales ; des diagnostics, des symptômes, etc. Par exemple, le type de la réponse attendue pour la question « *Comment traiter l'obésité chez les enfants ?* » est *Traitement* et celui de la question « *Comment diagnostiquer la maladie d'Alzheimer ?* » est *Examen*.

**Informations contextuelles.** Les questions médicales contiennent souvent des informations relatives au patient (e.g. *Comment traiter la bronchiolite chez les enfants de moins de 2 ans ?*). Il est indispensable de tenir compte de ces informations pour trouver la bonne réponse. En domaine ouvert, d'autres types d'informations contextuelles sont importants à repérer dans une question, comme le contexte temporel (e.g. *Quels rois ont régné en France entre le XVe et le XVIe siècle ?*).

**Recherche de documents et analyse sémantique.** En domaine ouvert, un SQR peut se baser sur la fréquence d'occurrence de l'information dans des grandes bases de données (ou sur le web) et sur des méthodes classiques de recherche d'information. En domaine médical, des sources d'information plus fiables et crédibles sont nécessaires. D'un autre côté, se baser sur la redondance des informations n'est pas suffisant vu que les questions médicales nécessitent des interprétations sémantiques plus profondes. Par exemple, dans une question comportant les deux termes T1 de type *médicament* et T2 de type *maladie*, il est important de connaître la relation sémantique entre T1 et T2 car chercher si T1 *complicque* T2 est différent de chercher si T1 *traite* T2.

En résumé, le domaine médical a besoin de nouvelles approches à différents niveaux. En effet, (1) de nouvelles méthodes sont nécessaires pour la *reconnaissance des EN* vu que leurs types diffèrent de ceux en domaine ouvert ; (2) de nouvelles *classifications des questions* doivent être élaborées ; (3) de nouveaux principes pour la *détermination du type de la réponse attendue* et (4) des techniques appropriées pour *l'extraction de la réponse* doivent être mis en place. En revanche, le domaine médical a une terminologie très évolutive. Il est donc nécessaire de passer par l'utilisation de *bases terminologiques spécialisées et évolutives* pour la réalisation d'un SQR.

Par exemple, pour l'analyse de la question, ces bases aident à identifier le type de la réponse attendue, elles pourront aussi être utilisées pour l'expansion des requêtes (déduire d'autres termes liés aux mots clés ; synonymes, acronymes, hyperonymes, variantes morphologiques, etc.).

### 3.2. État de l'art

L'existant des SQR médicaux peut être classifié en deux principales approches.

**Approche surfacique - syntaxique.** Il s'agit de rechercher des passages de textes pouvant contenir la réponse puis l'extraire. La recherche passe généralement par une étape d'indexation qui consiste à identifier les mots ou expressions présentant le mieux le contenu d'un document. La méthode la plus simple pour réaliser cette indexation est de considérer le document comme un « sac de mots » (considérer que les mots sont indépendants et ne pas tenir compte des relations entre les termes). Des traitements linguistiques comme une analyse morphologique et/ou syntaxique des documents peuvent aussi enrichir cette indexation. Dans cette approche, l'extraction de la réponse se fonde sur quelques techniques issues du Traitement Automatique des Langues (TAL) qui n'impliquent pas forcément une analyse sémantique de la question et des documents interrogés.

**Approche profonde - sémantique.** Cette approche procède typiquement à une analyse sémantique de la question et des documents et produit une représentation formelle du sens. Comme exemple de travaux qui se fondent sur cette approche, citons [6] qui dérivent des représentations logiques des questions et des documents. Quand un appariement se produit entre la représentation logique de la question et celle d'un document, les phrases initiant l'appariement sont extraites comme réponses candidates. D'un autre côté, [7] utilisent le format « PICO » qui permet de représenter les concepts clés d'une question médicale typique (*Problème, Intervention, Comparaison et Résultat*). Un cas particulier de cette approche, appelé **approche à base de modèles (template-based approach)**, exploite une collection de modèles (ou types génériques) de questions créés manuellement et utilise des techniques sophistiquées de TAL pour la classification des questions. Comme exemple de travaux, [8] ont classifié des questions médicales en se basant sur des modèles sémantiques prédéterminés.

En synthèse de l'existant, nous considérons que les approches profondes - sémantiques sont les plus appropriées au domaine médical. En effet, ces approches manipulent des types plus avancés de questions qui impliquent la compréhension de leurs contextes et se fondent sur des techniques qui permettent une analyse sémantique des questions et des documents interrogés. Cependant, les travaux qui se fondent sur le format PICO pour représenter une question médicale présentent un inconvénient majeur qui est le manque d'expressivité de ce format par rapport à la sémantique de la question. Aussi, les approches à base de modèles présentent un autre inconvénient non négligeable, à savoir que les modèles utilisés pour classifier les questions ne couvrent pas toutes les questions possibles.

#### 4. Approche proposée

Afin de mettre en place un SQR spécialisé dans le domaine médical, nous proposons une approche fondée sur des graphes sémantiques. 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 toute question, même si elle ne correspond pas à un type de question prédéterminé.

Le prototype que nous proposons, décrit dans la figure 2, suit une architecture de base semblable à celle des SQR généralistes (cf. figure 1). Comme point de départ, nous analysons le corpus médical pour extraire les index et construire les graphes sémantiques correspondant aux différentes unités textuelles des documents du corpus. La construction des graphes sera décrite dans la section 5. Nous nous intéressons ensuite à la question, que nous analysons pour (i) déterminer les mots clés (termes médicaux et verbes de la question) et (ii) construire le graphe sémantique correspondant (cf. section 5.2.3). Le module « Recherche de documents » utilise ensuite les index des documents et les mots clés de la question pour extraire un premier sous-ensemble de documents avec des techniques de recherche d'information classiques. Cette première recherche réduit la taille des documents à fouiller sémantiquement. Enfin, le module « Extraction de la réponse » recherche des appariements entre le graphe sémantique de la question et les graphes correspondant aux documents extraits pour récupérer les unités textuelles pertinentes et donc extraire la/les réponse(s).

#### 5. Construction automatique d'un graphe sémantique à partir d'une phrase

Dans cette section, nous expliquons la méthode que nous proposons pour la construction automatique d'un graphe sémantique à partir d'une phrase du corpus ou d'une question médicale.

##### 5.1. Exemple représentatif

Prenons un exemple de phrase d'un document médical :

(1) *A new class of drugs, called aromatase inhibitors, is being used to treat women with advanced breast cancer and some with early breast cancer.*<sup>1 2</sup>

Ce qui nous intéresse dans une telle phrase est de pouvoir récupérer les informations suivantes : (i) les entités nommées médicales de la phrase, ici : *aromatase inhibitors* et *breast cancer*, (ii) les types

<sup>1</sup> <http://www.cancer.org>

<sup>2</sup> Une nouvelle catégorie de médicaments, appelée « aromatase inhibitors » est utilisée pour traiter des femmes atteintes d'un cancer du sein avancé et certaines à un stade peu avancé.

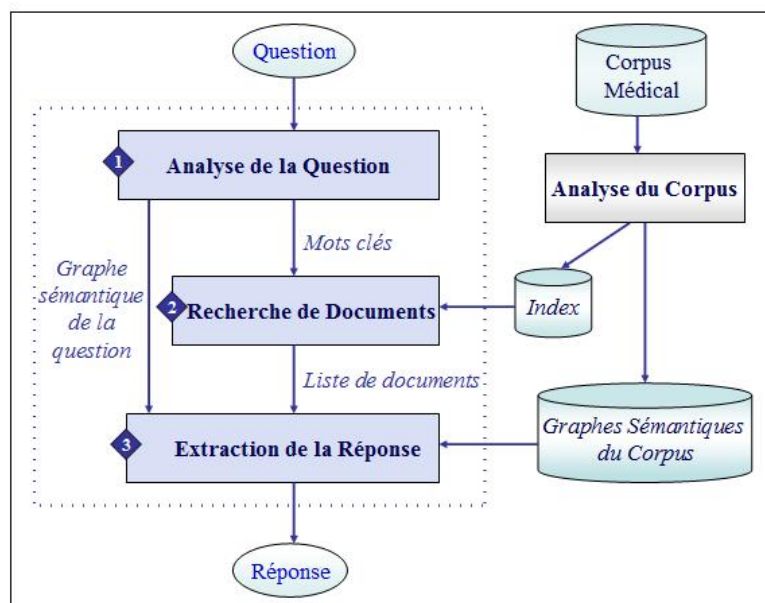


FIG. 2 – Architecture du prototype proposé

de ces EN (« Médicament » pour *aromatase inhibitors* et « Maladie » pour *breast cancer*) et (iii) la relation sémantique qui lie ces EN, qui est « *treats* » pour cette question. Une fois ces informations récupérées, nous cherchons à construire un graphe dont les nœuds correspondent aux EN, étiquetées par leurs types. Les arcs correspondent aux relations sémantiques entre les EN médicales (cf. figure 3).

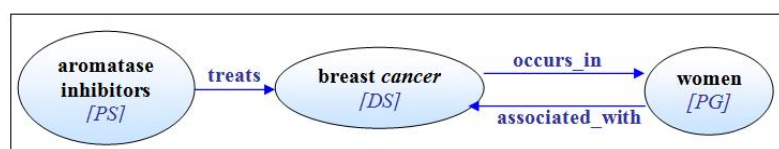


FIG. 3 – Un graphe sémantique possible pour la phrase (1)

## 5.2. Méthode proposée pour la construction automatique des graphes sémantiques

Pour construire automatiquement un graphe sémantique à partir d'une phrase, nous procédons en trois étapes : (i) le repérage des EN médicales présentes dans la phrase, (ii) l'étiquetage de ces EN par leurs types et (iii) l'identification des relations sémantiques les reliant.

### 5.2.1. Repérage et étiquetages des entités nommées médicales

Les EN médicales correspondent aux instances des concepts génériques du domaine médical (maladie, médicament, etc.). Afin de faciliter leur reconnaissance, il est possible d'utiliser des ressources terminologiques spécialisées. La ressource terminologique la plus développée en domaine médical est UMLS<sup>3</sup>, composé d'un métathésaurus, d'un réseau sémantique et d'un lexique spécialisé. Le métathésaurus intègre une centaine de thésaurus et de terminologies biomédicales. Il est organisé autour de la notion de concept possédant un terme préférentiel et éventuellement des termes synonymes, des abréviations et des variantes lexicales. Par exemple, le terme *infections*

<sup>3</sup> <http://www.nlm.nih.gov/research/umls/>

est associé au concept *Infection* (terme préférentiel). Le réseau sémantique organise les concepts du métathésaurus avec (i) des types sémantiques (135 types), hiérarchisés, catégorisant les concepts (e.g. le concept *Infection* est associé au type sémantique *Disease or Syndrome*) et (ii) des relations sémantiques entre ces types (54 relations). Nous considérons les types sémantiques du réseau sémantique de l'UMLS comme types d'EN médicales possibles et donc comme concepts génériques du domaine médical.

Pour la reconnaissance des EN médicales, nous avons choisi d'utiliser l'analyseur MetaMap [9]. Il s'agit d'un outil qui permet de détecter le vocabulaire médical à partir de documents en anglais et de déterminer les concepts du métathésaurus de l'UMLS correspondant aux termes repérés ainsi que leurs types sémantiques. La figure 4a présente un extrait du résultat de MetaMap associé à la phrase « This lady has a pneumonia.<sup>4</sup> ».

<p><b>Meta Candidates (3):</b>            1000 Pneumonia [Disease or Syndrome]            907 Lung [Body Part, Organ, or Organ Component]            907 Lung (Entire lung) [Body Part, Organ, or Organ Component]</p> <p><b>Meta Mapping (1000):</b>            1000 Pneumonia [Disease or Syndrome]</p>	<p><b>Meta Candidates (6):</b>            861 Infections (Infection) [Disease or Syndrome]            827 Infection (Communicable Diseases) [Disease or Syndrome]            ...            694 Enterococcus [Bacterium]</p> <p><b>Meta Mapping (888):</b>            694 Enterococcus [Bacterium]            861 Infections (Infection) [Disease or Syndrome]</p>
(a) « This lady has a pneumonia. »	(b) « for enterococcus infection »

FIG. 4 – Extraits de résultats de MetaMap

Malgré ses performances, MetaMap ne fournit pas toujours les bonnes EN médicales et les bons concepts. L'objet de la question : « What is the drug of choice for enterococcus infections?<sup>5</sup> » est l'EN *enterococcus infections* qui désigne une maladie. MetaMap fournit les informations présentées dans la figure 4b pour le syntagme nominal *for enterococcus infections*. Le terme *enterococcus* est associé au concept *Enterococcus* de type *Bacterium* et le terme *infections* est associé au concept *Infection* de type *Disease or Syndrome* alors que le terme *enterococcus infections* n'a pas été détecté (ce qui nous fait perdre une information importante). Pour traiter ce genre d'erreurs nous avons établi des patrons, écrits manuellement, que nous appliquons sur les résultats fournis par MetaMap.

### 5.2.2. Identification des relations sémantiques entre les termes médicaux

Après avoir repéré les EN médicales d'une phrase et les avoir étiquetées par leurs types, nous utilisons le réseau sémantique de l'UMLS ainsi que le métathésaurus pour déterminer la ou les relation(s) sémantique(s) liant ces EN. En effet, nous exploitons les relations du métathésaurus liant les concepts associés aux EN repérées et les relations du réseau sémantique liant les types sémantiques associés à ces concepts (e.g. entre les types sémantiques *Therapeutic or Preventive Procedure* et *Disease or Syndrome* il existe cinq relations : *treats*, *prevents*, *complicates*, *affects* et *associated\_with*). Nous récupérons alors les relations sémantiques possibles entre les EN de la phrase. Si plusieurs relations existent entre deux EN, nous déterminons la bonne relation sémantique en nous basant sur les informations syntaxiques provenant principalement des verbes de la phrase. Nous appliquons l'analyseur syntaxique XIP [10] pour identifier les syntagmes verbaux. Puis, nous utilisons des patrons lexico-syntaxiques, écrits manuellement, pour déduire la bonne relation sémantique parmi les relations possibles. Par exemple, pour les phrases de la forme « X [Therapeutic or Preventive Procedure] used to treat Y [Disease or Syndrome] », on peut déduire que la relation sémantique entre X et Y est *treats*. Si ces patrons ne permettent pas de sélectionner une relation unique, nous gardons toutes les relations possibles.

D'un autre côté, les relations présentes dans le réseau sémantique ou le métathésaurus de l'UMLS sont parfois vagues, redondantes ou ambiguës. Nous effectuons alors plusieurs traitements sur les résultats fournis par ces ressources. Par exemple, s'il existe deux relations R1 et R2 possibles entre deux termes et ces relations sont telles que R1 est un sous type de R2 (e.g. *occurs\_in* est un

<sup>4</sup> Cette dame a une pneumonie.

<sup>5</sup> Quel est le médicament de choix pour des infections à enterococcus ?

sous type de *associated\_with* dans le réseau sémantique de l’UMLS) alors nous ne gardons que la relation la plus spécialisée (R1, dans ce cas).

### 5.2.3. Construction d’un graphe sémantique associé à la phrase analysée

Les nœuds de ce graphe correspondent aux entités nommées identifiées et les arcs correspondent aux relations sémantiques entre ces termes.

(2) : *Nasogastric intubation is used to treat gastric atony.*<sup>6</sup>

La figure 5 présente le graphe sémantique associé à la phrase (2). Les entités nommées « nasogastric intubation » et « gastric atony » forment les nœuds du graphe. L’arc les reliant correspond à la relation sémantique « treats ».

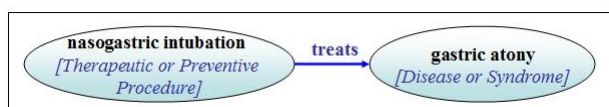


FIG. 5 – Graphe sémantique correspondant à la phrase (2)

Pour construire le graphe sémantique d’une question, nous avons besoin, en plus des EN médicales, des informations suivantes : (i) le focus de la question et le concept générique correspondant et (ii) le type de la réponse attendue désigné par un concept générique (traitement, maladie, etc.). Pour déterminer ces informations, nous utilisons des patrons construits manuellement. Ces patrons s’appuient, entre autres, sur le pronom interrogatif et les EN de la question. Nous différencions également deux types de relations sémantiques entre les EN d’une question : (i) relations sémantiques entre le concept générique correspondant au focus et le type de la réponse attendue et (ii) autres relations sémantiques possibles entre les EN de la question. Le premier type de relation est plus prioritaire lors de la recherche de la réponse car il correspond au cœur de la question alors que le deuxième type correspond au *contexte sémantique* de la question.

(3) « *45-year-old woman with dysfunctional uterine bleeding. What is the treatment ?* »<sup>7</sup>

Dans le graphe sémantique associé à la question (3) (cf. figure 6), la relation *treats* lie la réponse attendue et le focus de la question (*dysfunctional uterine bleeding*). D’autres relations existent entre les EN de la question et correspondent au contexte. Un des avantages des graphes sémantiques est qu’ils nous permettent de tenir compte des informations contextuelles présentes dans une question médicale.

## 6. Processus d’évaluation

Dans cette section, nous décrivons les données de test et les mesures d’évaluation utilisées.

**Evaluation globale.** Pour évaluer les résultats du SQR, nous avons besoin de couples de questions/réponses (questions avec leurs bonnes réponses) en anglais. Vu que ces données ne sont

<sup>6</sup> L’intubation nasogastrique est utilisée pour traiter l’atonie gastrique.

<sup>7</sup> Femme âgée de 45 ans avec saignement utérin dysfonctionnel. Quel est le traitement ?

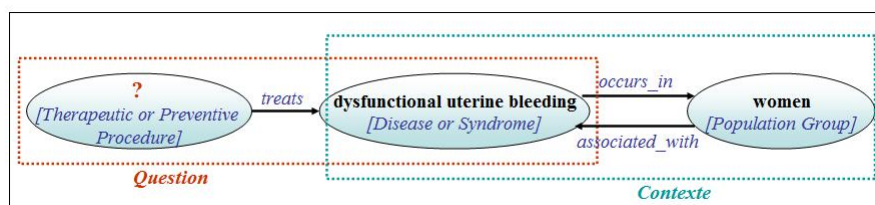


FIG. 6 – Graphe sémantique correspondant à la question (3).

pas disponibles, nous avons prévu de construire ces couples à partir de la base Medline, une base de résumés d'articles scientifiques en anglais dans le domaine médical (comportant plus de quatorze millions de références).

**Evaluation unitaire.** La précision des réponses retournées par le système est directement dépendante de l'exactitude des graphes sémantiques construits. Il est donc important d'évaluer les résultats intermédiaires des différentes phases de construction des graphes sémantiques, à savoir, la reconnaissance des EN médicales et l'identification des relations sémantiques. Nous utilisons deux corpus de test, le premier comporte des questions médicales et le deuxième des textes médicaux. Nous associons à chaque question et à chaque phrase la liste des EN médicales contenues, leurs types ainsi que les relations sémantiques qui les relient.

**Mesures d'évaluation.** Nous évaluons (i) les réponses fournies par le SQR et (ii) la reconnaissance des EN et l'identification des relations en nous basant sur les deux mesures d'évaluation classiques en recherche d'information : le rappel (nombre de résultats corrects trouvés/nombre de résultats corrects total) et la précision (nombre de résultats corrects trouvés/nombre de résultats trouvés). Ces expérimentations, en cours, mettent à l'épreuve les patrons et les règles que nous avons développés pour améliorer les résultats fournis par MetaMap et déterminer la bonne relation sémantique entre les EN médicales reconnues.

## 7. Conclusion

Nous avons présenté dans cet article une approche sémantique pour la réalisation d'un SQR spécialisé dans le domaine médical. Cette approche se fonde sur la modélisation de la question ainsi que des documents interrogés sous forme de graphes sémantiques. Nous avons présenté également la méthode que nous proposons pour la construction automatique de ces graphes. Cette méthode s'appuie entre autres sur des outils et ressources spécialisés disponibles ainsi que sur des patrons et des règles conçus pour améliorer la précision de l'extraction des EN et des relations sémantiques et donc améliorer la qualité des réponses retournées. Nous avons aussi proposé d'utiliser la notion de contexte sémantique de la question pour utiliser le voisinage des réponses candidates dans les textes. Enfin nous avons proposé un benchmark que nous utilisons actuellement pour évaluer notre approche.

## Bibliographie

1. Engelbrecht R. Expert systems for medicine - functions and developments. *Zentralbl Gynakol*, 119(9), 1997.
2. Ely J., Osheroff J., Ebell M., Bergus G., Levy B., Chambliss M., and Evans E. Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319, 1999.
3. Covell D., Uman G., and Mannig P. Information needs in office practice. Are they being met? *Ann. Intern. Med*, 103, 1985.
4. Ely J., Osheroff J., Ebell M., Chambliss M., Vinson D., Stevermer J., and Pifer E. Obstacles to answering doctors' questions about patient care with evidence : qualitative study. *BMJ*, 324, 2002.
5. Moriceau V., Tannier X., and Grau B. Utilisation de la syntaxe pour valider les réponses à des questions par plusieurs documents. *CORIA*, 2009.
6. Dowdall J. Schneider G. Rinaldi, F. and A. Persidis. Answering questions in the genomics domain. *ACL 2004 Workshop on Question Answering in Restricted Domains*, 2004.
7. Demner-Fushman D. and Lin J. Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33 :63–103, 2007.
8. P. Jacquemart and P. Zweigenbaum. Towards a medical question answering system : a feasibility study. *Studies in Health Technology and Informatics*, 2003.
9. Aronson A.R. Effective mapping of biomedical text to the UMLS Metathesaurus : the MetaMap program. pages 17–21, 2001.
10. Aït-Mokhtar S., Chanod J-P., and Roux C. Robustness beyond shallowness : incremental dependency parsing. *Special issue of the NLE Journal*, 2002.