



**HAL**  
open science

## Extraction de relations dans des comptes rendus hospitaliers

Anne-Lyse Minard, Anne-Laure Ligozat, Brigitte Grau

► **To cite this version:**

Anne-Lyse Minard, Anne-Laure Ligozat, Brigitte Grau. Extraction de relations dans des comptes rendus hospitaliers. IC 2011, 22èmes Journées francophones d'Ingénierie des Connaissances, May 2012, Chambéry, France. pp.491-506. hal-00746720

**HAL Id: hal-00746720**

**<https://hal.science/hal-00746720>**

Submitted on 29 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Extraction de relations dans des comptes rendus hospitaliers

Anne-Lyse Minard<sup>1,2</sup>, Anne-Laure Ligozat<sup>1,3</sup>, Brigitte Grau<sup>1,3</sup>

<sup>1</sup> LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France

<sup>2</sup> Université Paris-Sud, 91400 Orsay, France

<sup>3</sup> ENSIIE, 1 square de la résistance, 91000 Évry, France  
prenom.nom-@-limsi.fr

**Résumé** : L'extraction d'informations en domaine de spécialité amène à se poser différents problèmes, liés aux types d'information recherchés. Dans cet article, nous nous intéressons à l'identification de relations entre concepts dans des compte-rendus médicaux, tâche évaluée dans la campagne i2b2 en 2010. Les relations étant exprimées par des formulations très variées en langue, nous avons procédé à l'analyse des phrases en extrayant des traits qui concourent à la reconnaissance de la présence d'une relation et nous avons considéré l'identification des relations comme une tâche de classification multi-classes, chaque catégorie de relation étant considérée comme une classe, pour laquelle nous avons utilisé un SVM. Cette approche a permis d'obtenir une F-mesure de 0,70 environ, classant le système parmi les meilleurs lors de l'évaluation.

**Mots-clés** : Apprentissage semi-supervisé, extraction de relations, domaine médical

## 1. Introduction

Les systèmes informatiques médicaux ont connu une grande évolution depuis ces deux dernières décennies et permettent de stocker de l'information, d'y accéder en vue de découvrir de nouvelles informations ou de fournir une aide à la décision pour l'amélioration de la qualité des soins. Dans ce cadre général, l'information exploitée concerne principalement la littérature médicale et les dossiers médicaux de patients, comme les compte-rendus cliniques et les compte-rendus de consultation, qui contiennent de nombreuses informations sur le suivi médical des patients. L'information à exploiter est pour une grande part sous forme textuelle, et il s'agit alors de pouvoir l'extraire

automatiquement des textes. Le besoin de convertir toute cette information sous forme structurée est donc un enjeu majeur et constitue le point de départ du développement et de la mise au point d'outils d'interrogation adaptés, et de traitement automatique de cette information.

L'extraction de ces informations amène à se poser différents problèmes, liés aux types d'information recherchés : la reconnaissance des termes du domaine dans les textes, des concepts qui leur sont liés, ainsi que l'identification des types de relations qui les lient dans les documents. L'analyse terminologique des documents permet d'en construire des index sémantiques exploités pour leur recherche (Jonquet *et al.*, 2010), qui permettent de s'abstraire de la formulation en langue. L'identification de relations entre les concepts du domaine en donne une représentation plus structurée et constitue une étape nécessaire à la recherche d'informations précises dans ces textes via par exemple des moteurs de question-réponse en langue naturelle (Sang *et al.*, 2005; Embarek & Ferret, 2010).

Dans cet article <sup>1</sup>, nous nous intéressons à l'identification de relations dans des compte-rendus médicaux, tâche qui a fait l'objet d'une campagne d'évaluation i2b2 en 2010 <sup>2</sup>. L'objectif était de reconnaître différents types de relations entre des concepts (de type traitement, test et problème) de ces documents. Les relations y sont exprimées par des formulations très variées en langue, s'appuyant sur des pivots qui peuvent être des verbes ou des prépositions, et leur reconnaissance nécessite d'analyser les énoncés portant sur ces concepts. L'incomplétude des bases de connaissances sémantiques alliée à la difficulté d'associer toutes les formulations en langue à leur représentation conceptuelle est un frein à la réalisation d'une analyse profonde des phrases qui mettrait en évidence les relations entre concepts.

Aussi, nous avons considéré l'identification des relations comme une tâche de classification multi-classes, chaque relation étant considérée comme une classe. Cette approche permet de travailler sur l'extraction de traits qui permettent la reconnaissance de la présence d'une relation, en s'appuyant sur des analyses de surface et l'identification de certains termes.

---

1. Ce travail a été partiellement financé par OSEO dans le cadre du programme Quæro.

2. <https://www.i2b2.org/NLP/Relations/>

## 2. L'extraction de relations en domaine biomédical

Parmi les approches utilisées en extraction de relation, celles qui reposent sur la définition manuelle de patrons sont peu robustes, et généralement efficaces surtout en précision. On peut citer dans le domaine médical le système SemRep (Rindfleisch *et al.*, 2000) conçu pour extraire des relations de branchement artériel dans des compte-rendus opératoires et également appliqué à des relations entre des problèmes médicaux et leurs traitements (Srinivasan & Rindfleisch, 2002). Le système MedLEE, extrait des relations de compte-rendus radiologiques, des interactions biomoléculaires (Friedman *et al.*, 2001) et des relations gène-phénotype (Chen & Friedman, 2004).

D'autres travaux utilisent l'apprentissage afin de combiner automatiquement les informations pertinentes pour détecter et typer les relations. (Uzuner *et al.*, 2010) utilisent des SVM (machines à vecteurs de support) pour typer des relations entre des problèmes, des tests et des traitements dans des compte-rendus médicaux. Pour classer les relations, ils utilisent des attributs de surface (l'ordre des concepts, la distance, etc.), des attributs lexicaux (des trigrammes lexicaux, les mots qui forment les concepts, etc.), des attributs morpho-syntaxiques et syntaxiques (les verbes, des bigrammes syntaxiques, les dépendances syntaxiques entre les concepts, etc.). Pour plusieurs relations, ils obtiennent des F-mesures entre 0,60 et 0,85, mais pour les relations pour lesquelles il y a peu d'exemples dans le corpus d'apprentissage les F-mesures sont nulles. (Roberts *et al.*, 2008) utilisent également des SVM pour extraire des relations dans le corpus du projet CLEF (the Clinical E-Science Framework project). Ils extraient des relations entre des entités (ex : condition, médicament, résultat) et des modificateurs (ex : marqueur de négation) dans des dossiers de patients atteints d'un cancer. Les relations sont de sept types, et deux types d'entités (ou une entité et un modificateur) ne peuvent être reliées que par une relation (sauf entre une investigation et une condition où la relation peut être de deux types). La tâche est modélisée comme une classification binaire, c'est-à-dire que les SVM sont entraînés pour une décision entre une classe et les non-relations. Ils utilisent aussi des attributs lexicaux, morpho-syntaxiques et sémantiques.

Des travaux similaires ont été conduits en domaine général ; par exemple (Zhou *et al.*, 2005) ont travaillé sur le corpus ACE. Ils utilisent des SVM pour identifier des relations entre des personnes, des organisations, des lieux, etc.

Le système que nous décrivons ici utilise également des SVM pour classer des relations avec une granularité fine. Il utilise des attributs classiques,

ainsi que des attributs liés à la spécificité du corpus. En effet nous utilisons les types sémantiques de l'UMLS<sup>3</sup>, des listes d'abréviations spécifiques aux comptes rendus hospitaliers et nous traitons la coordination des concepts, structure très présente dans ce corpus.

### 3. Relations considérées

L'objectif de la tâche était d'annoter dans les documents les relations existant entre deux concepts. Les concepts considérés étaient les suivants :

- Les problèmes médicaux, définis comme les observations des patients ou des cliniciens concernant ce qui n'allait pas ou semblait être causé par une maladie. Cette catégorie comprend notamment les maladies, syndromes, les observations sur l'état mental du patient etc.
- Les traitements, définis comme les procédures, interventions, substances et médicaments donnés à un patient pour tenter de résoudre un problème.
- Les tests, comprenant les procédures et examens effectués sur un patient ou un fluide corporel pour vérifier ou infirmer la présence d'un problème, ou pour avoir plus d'informations sur un problème.

Entre ces trois types de concepts, des relations peuvent exister : un test comme un examen est par exemple prescrit pour analyser un problème. Ce sont ces relations que nous avons cherché à identifier dans les documents. Afin d'étudier la reconnaissance des relations, les concepts étaient annotés<sup>4</sup>, et il s'agissait de déterminer si, étant donné deux concepts, ils étaient en relation, et si oui, laquelle. Huit relations ont été définies (cf. tableau 3.).

L'accord inter-annotateur pour chaque relation est donné dans le tableau 2. Cet accord a été calculé par Knowtator et fourni par les organisateurs du challenge. L'accord ajusté est obtenu après discussion des cas posant problème. On peut voir que l'accord inter-annotateur est faible pour la relation TrWP (le traitement aggrave le problème) et TrIP (le traitement améliore le problème).

### 4. Corpus

Les corpus, fournis par les organisateurs de la tâche i2b2, sont composés de comptes rendus hospitaliers provenant de plusieurs centres médicaux aux

---

3. Unified Medical Language System (<http://www.nlm.nih.gov/research/umls/>)

4. Dans les exemples présentés dans l'article, les problèmes seront balisés <pb>, les traitements <treat> et les tests <test>.

TrIP	le traitement améliore le problème
	<pb> hypertension </pb> was controlled on <treat> hydrochlorothiazide </treat>
TrWP	le traitement aggrave le problème
	<pb> the tumor </pb> was growing despite the available <treat> chemotherapeutic regimen </treat>
TrCP	le traitement cause le problème
	<treat>Bactrim</treat> could be a cause of <pb>these abnormalities</pb>
TrAP	le traitement est administré en raison du problème
	<treat>antibiotic therapy</treat> for presumed <pb>right forearm phlebitis</pb>
TrNAP	le traitement n'est pas administré en raison du problème
	<treat>Relafen</treat> which is contraindicated because of <pb>ulcers</pb>
TeRP	le test révèle le problème
	<test>an echocardiogram</test> revealed <pb>a pericardial effusion</pb>
TeCP	le test est conduit en raison du problème
	<test>an VQ scan</test> was performed to investigate <pb>pulmonary embolus</pb>
PIP	un problème en indique un autre
	<pb>Azotemia</pb> presumed secondary to <pb>sepsis</pb>

TABLE 1: Description des huit relations

États-Unis. Ces documents avaient été anonymisés et annotés manuellement pour constituer une référence. Un premier corpus a été fourni avant l'évaluation, composé de 350 documents. Ce premier corpus a été divisé en trois : corpus d'entraînement, de développement, et de test. Puis, les organisateurs d'i2b2 ont fourni le corpus d'évaluation, qui comporte 477 documents. Le nombre de relations dans chaque corpus est indiqué dans le tableau 2.

Le corpus est composé de phrases courtes (en moyenne 17 mots/phrasedans le corpus d'entraînement). En effet les documents de type compte rendu hospitalier sont composés de fragments de phrase (1) et d'énumérations (2).

- (1) <pb>C5-6 disc herniation</pb> with <pb>cord compression</pb> and <pb>myelopathy</pb>.
- (2) Revealed <pb>icteric sclerae</pb>, <pb>the oropharynx with extensive thrush</pb>, and <pb>an ulcer under his tongue</pb>.

Relation	pré-éval	<i>entraînement</i>	<i>développement</i>	<i>test</i>	évaluation	IAA stric
TrIP	107	74	18	15	198	0,44
TrWP	56	39	10	7	143	<b>0,30</b>
TrCP	296	237	23	36	444	0,50
TrAP	1423	1013	229	181	2487	0,68
TrNAP	106	71	15	20	191	0,44
PIP	1239	855	224	160	1986	<b>0,35</b>
TeRP	1734	1236	226	272	3033	0,70
TeCP	303	196	49	58	588	0,43
Toutes	5264	3721	794	749	9070	0,56

TABLE 2: Nombre de relations par catégorie dans chaque corpus et accord inter-annotateur (IAA)

## 5. Méthodes

### 5.1. Prétraitement du corpus

Les fichiers du corpus ont été prétraités et normalisés. Tout d'abord, les abréviations connues ont été remplacées par leur forme complète, grâce à une liste. La liste a été constituée pour la campagne d'évaluation i2b2 2009<sup>5</sup> à partir de la liste d'abréviations biomédicales formée par Berman<sup>6</sup> à laquelle ont été ajoutés les exemples du corpus du challenge i2b2 2009. Ainsi, *h.o.* a été converti en *history of* et *p.r.n.* en *as needed*. Puis, les données d'anonymisation (différentes pour chaque corpus) ont été remplacées par des balises *NAME*, *DATE* et *AGE*. *NUM* remplace toutes les valeurs numériques présentes dans les fichiers (principalement des dosages). Enfin, les textes ont été étiquetés par le TreeTagger (Schmid, 1994).

### 5.2. Classification

Nous avons utilisé comme classifieur les SVM, avec la bibliothèque LIBSVM. La classification a été faite en multi-classes avec une approche « un-contre-un ». Nous l'avons également testée par couple de concepts, c'est-à-dire que nous avons entraîné un classifieur pour les relations entre test et problème,

5. <https://www.i2b2.org/NLP/Medication/> (page consultée en février 2011)

6. <http://www.julesberman.info/abtwo.htm> (page consultée en février 2011)

un pour les relations entre traitement et problème et un pour la relation problème problème. Les résultats que nous avons obtenus étaient globalement moins bons qu'en multi-classes.<sup>7</sup> Les attributs choisis pour la classification automatique prennent en compte des informations de surface, sur les distances entre mots par exemple, des informations lexicales, comme les mots formant les concepts, des informations syntaxiques, les catégories morpho-syntaxiques des mots, et des informations sémantiques grâce au typage des concepts. Ils sont calculés automatiquement, en utilisant éventuellement des outils et des ressources externes. Ils sont modélisés par des identifiants numériques uniques qui prennent pour valeur un si l'attribut est présent et zéro sinon.

### 5.2.1. Attributs de surface

- L'ordre des concepts : permet de savoir si le test/traitement est indiqué avant ou après le problème, ce qui influera sur la façon dont est exprimée la relation. Dans l'exemple (3) le test (*an MRI*) est donné avant les problèmes qu'il a révélés. L'exemple (4) montre un cas où le problème est cité avant le test, ce dernier étant suivi des résultats obtenus (<NUM>).
  - (3) She had <test>a workup</test> by her neurologist and <test>an MRI</test> revealed <pb>a C5-6 disc herniation</pb> with <pb>cord compression</pb> and <pb>a T2 signal change</pb> at that level.
  - (4) The patient was <pb>thrombocytopenic</pb> with <test>a platelet count</test> of <NUM> on the <NUM>.
- La distance entre les deux concepts en termes de nombre de mots<sup>8</sup> : dans le corpus d'entraînement, il n'y a jamais plus de 65 mots entre deux concepts en relation, mais deux concepts qui ne sont pas en relation peuvent être séparés par 205 mots au maximum<sup>9</sup>.
- La présence d'autres concepts entre les deux concepts étudiés : dans le corpus d'entraînement, 80% des paires de concepts séparées par d'autres

---

7. Nous avons utilisé le paramètre *c* avec la valeur 16 et le paramètre *gamma* avec la valeur 0,03125. Ces valeurs ont été déterminées par l'outil python *grid* fourni dans la bibliothèque LIBSVM.

8. Le découpage en mots est effectué par le TreeTagger. Les mots incluent aussi les signes de ponctuation

9. Cet attribut prend pour valeur une distance, contrairement aux autres attributs qui prennent pour valeur un ou zéro.



concepts ne sont pas en relation. Dans l'exemple (5), cinq relations sont envisagées, chaque problème peut être en relation avec chaque traitement, et les problèmes peuvent être également en relation. Mais le fait qu'il y ait un traitement entre le premier traitement et le deuxième problème élimine la possibilité d'avoir une relation entre ces derniers.

(5) Take with food < treat >MILK OF MAGNESIA (MAGNESIUM HYDROXIDE)< /treat > < NUM > MILLILITERS per oral twice daily as needed < pb >Constipation< /pb > < treat >PERCOCET< /treat > < NUM > tablet per oral Q4H as needed < pb >PAIN< /pb >

### 5.2.2. Attributs lexicaux

- Les mots, et leurs radicaux <sup>10</sup>, qui désignent les concepts, et le mot tête de chaque concept <sup>11</sup>. Nous avons considéré les radicaux de manière à regrouper les variantes flexionnelles et dérivationnelles. Les mots qui forment les concepts peuvent être déclencheurs d'une relation : dans l'exemple (6) la relation entre le traitement et le problème est de type TrWP (le traitement aggrave le problème), relation qui est déclenchée par l'adjectif *recurrent* à l'intérieur du deuxième concept.

(6) He has had < NUM > week courses of < treat >antibiotics< /treat > with < pb >recurrent bacteremia< /pb >.

- Les radicaux des trois mots à gauche et à droite des concepts. La taille de la fenêtre a été choisie après des tests. Avec une fenêtre plus grande ou plus petite, la précision augmente légèrement mais le rappel diminue.
- Les radicaux des mots entre les concepts : permet de tenir compte de tous les mots entre les concepts ; c'est ici qu'est située l'information la plus utile à la classification.
- Les radicaux des verbes dans une fenêtre de trois mots avant et après chaque concept, et entre eux. Le verbe marque souvent la relation : par exemple dans (7) la relation TeRP (le test révèle le problème) est exprimée par le verbe *reveal*.

(7) < test >Chest x-rays< /test > **revealed** evidence of < pb >congestive heart failure< /pb >.

10. Pour obtenir les radicaux des mots nous utilisons le module PERL *lingua* : `:stem`.

11. La tête d'un concept correspond au mot précédant une préposition ou le dernier mot du concept, comme défini dans (Zhou *et al.*, 2005).

- Les prépositions entre les concepts. Dans l'exemple (8) la préposition *for* marque une relation de type TrAP (le traitement est administré pour le problème).

(8) She was treated with <reat>IVF</reat> **for** <pb>her ARF</pb>.

### 5.2.3. Attributs syntaxiques

Les attributs syntaxiques permettent d'utiliser des informations relatives à la structure syntaxique de surface des phrases pour repérer les relations. Nous prenons en compte la succession des catégories syntaxiques et la présence d'élément d'une certaine catégorie, indifféremment de leur réalisation lexicale.

- La catégorie morpho-syntaxique des mots dans une fenêtre de trois mots à gauche et à droite des concepts : les catégories proviennent de l'étiquetage du TreeTagger.
- La présence d'une préposition entre les concepts, peut importe la préposition.
- Un attribut marque la présence d'un signe de ponctuation lorsqu'il est le seul présent entre les deux concepts. Cet attribut permet de tenir compte des énumérations.

### 5.2.4. Attributs sémantiques

Ces attributs permettent de généraliser l'information portée par certains mots des phrases et concernent les concepts du domaine d'une part et les classes de verbes d'autres part.

- Le type sémantique (issu de l'UMLS) des mots dans une fenêtre de trois mots à gauche et à droite de chaque concept. Dans l'exemple (3), le terme *neurologist* a pour type sémantique *professional or occupational group* ; un attribut l'indiquera pour la classification des paires contenant le concept *an MRI*.
- Les types des concepts (problème, test ou traitement) : c'est l'attribut le plus important car les relations ne sont pas les mêmes entre un test et un problème, un traitement et un problème et entre deux problèmes.
- Les classes de VerbNet<sup>12</sup> (une extension des classes de Levin) des verbes dans une fenêtre de trois mots à gauche et à droite des concepts, et entre

---

12. <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (page consultée en février 2011)

les concepts. Par exemple le verbe *reveal* fait partie de la classe *indicate-78-1-1* qui contient également les verbes *show*, *prove*, *demonstrate*, etc. Nous observons que dans les exemples (7) et (9) les verbes *reveal* et *show* sont utilisés pour marquer des relations du même type.

- (9) <test>Recent chest x-ray</test> **shows** <pb>resolving right lower lobe pneumonia</pb>.

### 5.2.5. Coordination

Comme nous l'avons signalé dans la section 4., les phrases du corpus sont souvent formées d'énumérations. Deux concepts en relation peuvent être séparés par une liste d'autres concepts sans qu'ils apportent une information sur la relation. Suite à cette observation nous avons traité les phrases (de façon différente pour chaque couple de concepts) avant d'extraire les attributs décrits précédemment. Nous avons supprimé les concepts en coordination avec les concepts étudiés. Trois attributs ont alors été rajoutés : le nombre de concepts entre les deux concepts en relation avant la réduction de la phrase, les mots déclencheurs de la suppression des concepts (*or*, *and*, *,* et *,* *and*) et un dernier qui indique que la phrase a été réduite.

Dans le corpus d'entraînement, une réduction de la phrase a été faite par suppression des concepts coordonnés pour 23% des couples de concepts possiblement en relation (3819 couples sur 16437).

- (10) A/P : 48yo man with history of <pb>HCV</pb>, <pb>bipolar DO</pb>, history of suicide attempts, a/w <pb>overdose</pb> of <treat>Inderal</treat>, <treat>Klonopin</treat>, <treat>Geodon</treat>, s/T Jackson stay with <treat>intubation</treatment> [...]

Dans l'exemple (10), pour la classification de la relation entre les concepts *overdose* et *Geodon*, le segment de phrase entre les concepts après réduction est : a/w <pb>overdose</pb> of <treat>Geodon</treat> (les segments avant et après ne changent pas).

### 5.2.6. Sélection des attributs

Pour évaluer l'utilité de chaque attribut, nous avons suivi la même méthode que (Roberts *et al.*, 2008). Nous avons mesuré les performances du système sur le corpus de test en utilisant des attributs de base et en ajoutant une série d'attributs à la fois. Les résultats sont donnés dans le tableau 3.

Les attributs ont été regroupés en classes selon les informations qu’ils décrivent. Les classes d’attributs utilisées par le système initial *base* sont les radicaux et les catégories morpho-syntaxiques des trois mots à gauche et à droite des concepts et les radicaux des mots entre les concepts. Sont ensuite ajoutées les classes d’attributs *dist* (distance entre les deux concepts), *conc* (autres concepts), *dir* (position du premier concept par rapport au deuxième), *verb* (radicaux des verbes et classes de Levin), *prep* (préposition entre les concepts), *intra* (mots qui forment les concepts et mot tête du concept) et *types* (types sémantiques). Les résultats de la dernière colonne du tableau correspondent aux résultats du système final. Dans le système de base utilisé pour ces tests, nous n’avons pas utilisé les attributs liés à la gestion de la coordination des concepts. Nous avons évalué ces attributs séparément, en les utilisant dans notre système final nous améliorons la F-mesure de 0,002.

L’ajout de certaines classes d’attributs peuvent améliorer la classification d’une relation mais détériorer celle d’une autre relation. Quand on ajoute la série d’attributs de la classe *intra*, la F-mesure de la classification de la relation TrCP et TrAP augmente alors que la F-mesure pour PIP et TeCP diminue. Nous gardons cette classe d’attribut car elle permet d’augmenter la F-mesure générale. De même l’ajout de *verb* augmente d’environ 0,04 la F-mesure pour la relation TrNAP, mais fait chuter de 0,09 la F-mesure pour la relation TrIP. La F-mesure générale reste stable, mais si nous retirons cette classe d’attribut du système final, la F-mesure passe de 0,727 à 0,712.

Relation	base	+dist	+conc	+dir	+verb	+prep	+intra	+types
<b>TrIP</b>	0,333	0,333	0,333	0,333	0,235	0,235	0,235	0,235
<b>TrWP</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
<b>TrCP</b>	0,366	0,370	0,405	0,411	0,424	0,441	<b>0,526</b>	0,517
<b>TrAP</b>	0,620	0,638	<b>0,708</b>	0,721	0,708	0,706	0,737	0,726
<b>TrNAP</b>	0,620	0,620	0,620	0,620	<b>0,666</b>	0,666	0,666	0,620
<b>PIP</b>	0,611	0,613	0,664	0,664	0,654	0,671	0,618	<b>0,659</b>
<b>TeRP</b>	0,790	0,792	0,833	0,843	0,850	0,848	0,866	0,866
<b>TeCP</b>	0,253	0,253	<b>0,373</b>	0,351	0,373	0,351	0,333	0,285
<b>Toutes relations</b>	0,647	0,652	<b>0,704</b>	0,712	0,711	0,713	0,724	<b>0,727</b>

TABLE 3: Variation de la F-mesure selon les attributs utilisés sur le corpus de test. Les F-mesures en gras sont significativement meilleurs que la F-mesure les précédant.

## 6. Évaluation

Les résultats obtenus sur le corpus d'évaluation sont présentés dans le tableau 4. Lors du challenge i2b2 2010 nous avons utilisé ce système (sans la gestion des problèmes de coordination) associé à des patrons prioritaires sur l'apprentissage pour l'extraction de quatre relations. Nous avons obtenu une F-mesure de 0,709, ce qui a placé notre système au 3ème rang sur 16 participants. Dans le tableau 4 nous donnons la F-mesure obtenue par les deux premiers systèmes, ainsi que la F-mesure médiane. Pour la relation TeRP, présentant le plus grand nombre d'exemples dans le corpus d'entraînement, la F-mesure atteint 0,852. En revanche, les relations moins bien dotées en exemples sont moins bien reconnues, comme la relation TrCP, la F-mesure n'est que de 0,489.

Relation	Rappel	Précision	F-mesure
<b>TrIP</b>	0,156	0,861	0,264
<b>TrWP</b>	0,000	0,000	0,000
<b>TrCP</b>	0,369	0,725	0,489
<b>TrAP</b>	0,693	0,739	<b>0,715</b>
<b>TrNAP</b>	0,057	0,423	0,101
<b>PIP</b>	0,552	0,787	0,649
<b>TeRP</b>	0,835	0,870	<b>0,852</b>
<b>TeCP</b>	0,238	0,833	0,370
<b>Toutes relations</b>	<b>0,628</b>	<b>0,803</b>	<b>0,705</b>
Médiane			0,664
1er système			0,736
2ème système			0,731

TABLE 4: Rappel, précision et F-mesure obtenus sur le corpus d'évaluation

Pour la classification des relations nulles (ou non-relations), le système obtient un rappel de 0,93, une précision de 0,84 et une F-mesure de 0,89.

## 7. Analyse des erreurs

Pour analyser les erreurs nous avons établi des matrices de confusion. Le tableau 5 présente la matrice de confusion pour les relations entre un traite-

ment et un problème, le tableau 6 pour les relations entre un test et un problème et le tableau 7 pour les relations entre deux problèmes.

La matrice de confusion 5 révèle que les relations mal typées sont le plus souvent classées dans TrAP (le traitement est administré en raison d'un problème) ou non-relation, c'est-à-dire dans les classes les plus représentées. Par exemple 54% des relations TrIP (le traitement améliore le problème) sont classées par erreur comme des non-relations, et 31% sont classées TrAP.

		ESTIMÉES					
		TrIP	TrWP	TrAP	TrNAP	TrCP	Non-relation
RÉELLES	TrIP	<b>25</b>	0	61	0	4	106
	TrWP	1	<b>0</b>	48	0	12	81
	TrAP	2	0	<b>1760</b>	10	21	689
	TrNAP	0	0	62	<b>12</b>	7	108
	TrCP	0	0	90	2	<b>168</b>	181
	Non-relation	1	0	396	7	26	<b>3069</b>

TABLE 5: Matrice de confusion pour la classification des relations entre un traitement et un problème

Il est parfois difficile de classer une relation TrIP ou TrAP car TrIP est une relation spécifique de TrAP. En effet, si un traitement améliore un problème alors le traitement a été administré en raison du problème. Il en est de même pour la relation TrWP qui inclut les cas où un traitement est administré pour un problème mais ne l'améliore pas : cette relation est englobée dans la classe TrAP. Dans la première partie de la phrase de l'exemple (11), nous détectons par erreur une relation de type TrAP entre *multiple wires* et *the total occlusion*. Mais le déclencheur d'une relation TrWP (*unsuccessful*) est dans la deuxième partie. Notre système ne prend en compte que les trois mots suivant le deuxième concept, l'adjectif n'est donc pas utilisé.

(11) <reat>Multiple wires</reat> were used to attempt to cross <pb>the total occlusion</pb>, but were ultimately unsuccessful.

Dans (12), la relation TeRP (un test révèle un problème) n'est pas reconnue entre *pressures* et *hypertension*, alors que la formulation n'est pas très complexe ; le problème doit provenir du faible nombre d'exemples de ce type.

(12) <test>His PA catheter</test> is showing <test>pressures</test> of about <NUM> and I suspect that he has <pb>chronic pulmonary hypertension</pb>, given these numbers.

		ESTIMÉES		
		TeCP	TeRP	Non-relation
RÉELLES	TeCP	<b>140</b>	71	376
	TeRP	24	<b>2559</b>	445
	Non-relation	4	306	<b>2147</b>

TABLE 6: Matrice de confusion pour la classification des relations entre un test et un problème

50% des relations PIP (un problème en indique un autre) ne sont pas repérées (c'est-à-dire qu'elles sont classées non-relation). Dans le corpus d'entraînement il y a un nombre suffisant d'exemples, mais la définition de la relation n'est peut être pas assez claire, comme le montre l'accord inter-annotateur (voir tableau 2). Dans l'exemple (13), une relation PIP a été notée entre *symptoms* et *anxiety*, mais dans l'exemple (14), la relation entre *symptoms* et *dry cough* n'a pas été annotée.

- (13) She was hooked up with support services in Collot Ln, Dugo, Indiana <NUM> for <reat>further counselling</reat> and given <reat>Xanax</reat> for <pb>symptoms</pb> of <pb>anxiety</pb>.
- (14) Pt was o/w in his USOH until <NUM> weeks ago when he developed <pb>a URI</pb> with <pb>symptoms</pb> of <pb>dry cough ;</pb> no <pb>fever</pb>, <pb>chills</pb>, <pb>sore throat</pb>, <pb>myalgias</pb>

		ESTIMÉES	
		PIP	Non-relation
RÉELLES	PIP	<b>992</b>	990
	Non-relation	402	<b>10746</b>

TABLE 7: Matrice de confusion pour la classification des relations entre deux problèmes

Trois types d'erreurs se dégagent de cette analyse :

- La relation est clairement exprimée par un verbe ou une expression, mais cette construction n'est pas présente dans le corpus d'entraînement. Dans (15), la relation entre *pulmonary nodules in his RML* et *fu imaging* a été étiquetée TeRP ; en effet le verbe *reveal* est déclencheur d'une relation TeRP, et le déclencheur d'une relation TeCP est *which need* mais il n'apparaît qu'une seule fois dans le corpus d'entraînement.

(15) <test>CTS chest</test> was negative for <pb>PE</pb>, however it did reveal <pb>pulmonary nodules in his RML</pb> which need <test>fu imaging</test> in <NUM> months.

- La relation ne peut être classée sans l’utilisation de connaissances externes. Dans (16) le système détecte à tort une relation PIP. Pour classer cette relation il faudrait savoir que les incisions n’ont pas de rapport avec l’obésité du patient.

(16) <pb>obese</pb> with <pb>multiple well healed surgical incisions</pb>, positive bowel sounds.

- L’annotation de la relation est discutable, par exemple dans (17) une relation PIP a été annotée entre *lower abdominal pain* et *a symptom*, mais ces deux termes font référence au même concept.

(17) He’d been having <pb>lower abdominal pain</pb> for approximately the past week, <pb>a symptom</pb> for which he’s been admitted in the past.

Pour améliorer la détection des relations comme TrWP ou TriP il faudrait annoter un plus grand corpus. En effet le problème que nous rencontrons est le trop petit nombre d’instances de ces relations dans le corpus. Il n’existe pas à l’heure actuelle de tels corpus annotés et accessibles librement.

## 8. Conclusion

L’extraction de relations entre des concepts dans des comptes-rendus hospitaliers est une tâche qui contribue à améliorer l’accès aux informations contenues dans les documents médicaux. Cette tâche repose sur la reconnaissance des différentes formes que l’expression de relations peut prendre dans des phrases. Cette variabilité étant très importante en terme de vocabulaire et en terme de structure syntaxique, nous en avons tenu compte en définissant des indices de différentes natures, aptes à caractériser ces types de phrase. Nous utilisons des traits spécifiques au corpus, le type des concepts par exemple, ainsi que des attributs classiques. L’ensemble sélectionné et la combinaison que nous en avons faite a permis d’obtenir de très bons résultats. Les attributs sélectionnés sont suffisamment généraux pour qu’ils puissent être utilisés sur des corpus d’autres domaines, avec une adaptation des attributs dépendants du domaine (comme les types sémantiques).

Les résultats sont mauvais pour les relations peu représentées dans le corpus. Pour avoir plus d’exemples similaires de ces relations, il faudrait réduire



la variabilité syntaxique et à cette fin effectuer une simplification des phrases avant d'apprendre les relations. Il serait également intéressant d'étudier si l'utilisation d'autres informations syntaxiques améliorerait la classification.

## Références

- CHEN L. & FRIEDMAN C. (2004). Extracting phenotypic information from the literature via natural language processing. In *Medinfo 2004 : Proceedings Of THE 11th World Congress On Medical Informatics*.
- EMBAEK M. & FERRET O. (2010). Adapter un système de question-réponse en domaine ouvert au domaine médical. In *TALN*.
- FRIEDMAN C., KRA P., YU H., KRAUTHAMMER M. & RZHETSKY A. (2001). Genies : a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics*, **17**.
- JONQUET C., COULET A., SHAH N. H. & MUSEN M. A. (2010). Indexation et intégration de ressources textuelles à l'aide d'ontologies : application au domaine biomédical. In *21èmes Journées Francophones d'Ingénierie des Connaissances, IC'10*, p. 271–282.
- RINDFLESCH T. C., BEAN C. A. & SNEIDERMAN C. A. (2000). Argument identification for arterial branching predications asserted in cardiac catheterization reports. In *AMIA Annu Symp Proc*, p. 704–708.
- ROBERTS A., GAIZAUSKAS R. & HEPPLER M. (2008). Extracting clinical relationships from patient narratives. In *BioNLP2008 : Current Trends in Biomedical Natural Language Processing*, p. 10–18.
- SANG E., BOUMA G. & DE RIJKE M. (2005). Developing offline strategies for answering medical questions. In *Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains, Pittsburgh, PA, USA*, p. 41–45.
- SCHMID H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, p. 44–49.
- SRINIVASAN P. & RINDFLESCH T. (2002). Exploring text mining from medline. In *Proc AMIA Symp*, p. 722–726.
- UZUNER O., MAILLOA J., RYAN R. & SIBANDA T. (2010). Semantic relations for problem-oriented medical records. *Artificial Intelligence in Medicine*, **50**, 63–73.
- ZHOU G., SU J., ZHANG J. & ZHANG M. (2005). Exploring various knowledge in relation extraction. In *Proceedings of the 43rd Annual Meeting of the ACL*, p. 427–434.