

---

## Identification des assertions dans les textes médicaux : application à la relation {patient, problème médical}

Amandine Périnet\* – Natalia Grabar\*\* – Thierry Hamon\*

\*LIM&BIO (EA3969), UFR SMBH Léonard de Vinci, Université Paris 13,  
74, rue Marcel Cachin, 93017 Bobigny Cedex  
amandine.perinet@edu.univ-paris13.fr

thierry.hamon@univ-paris13.fr

\*\*CNRS UMR 8163 STL, Université Lille 3, 59653 Villeneuve d'Ascq  
natalia.grabar@univ-lille3.fr

---

*RÉSUMÉ. Les textes scientifiques regorgent d'assertions (hypothèses, possibilités) qui nuancent le discours et donnent plus ou moins de certitude aux informations. Nous nous intéressons aux assertions portant sur la relation {patient, problème médical} dans les dossiers patients. Notre objectif est d'identifier automatiquement la certitude et les degrés de certitude, ainsi que leur polarité (positive/négative). Nous exploitons des marqueurs morphologiques, contextuels et structurels pour identifier et désambiguïser les assertions. Notre système a été évalué dans le cadre de la compétition internationale I2B2. La détection de certaines catégories d'assertion est satisfaisante alors que d'autres restent difficiles à détecter. Nous étudions également l'apport de chaque marqueur et analysons l'évolution des résultats.*

*ABSTRACT. Scientific texts are full of assertions (i.e., hypothesis, possibility) that shade the discourse and give more or less certainty to information. We focus on the assertions that influence the relation {patient, medical problem} in patients' records. We aim at automatically identifying certainty and the degrees of certainty, as well as their polarity (positive/negative). We exploit morphological, contextual and structural markers for identification and disambiguation of assertions. Our system has been evaluated within the I2B2 international challenge. Several categories of assertion are correctly identified while others remain difficult to detect. We explore also the contribution of each marker and analyse the evolution of the results.*

*MOTS-CLÉS : documents médicaux, extraction d'information, modalité, certitude.*

*KEYWORDS: medical documents, information extraction, modality, certainty.*

---

## 1. Pourquoi s'intéresser aux assertions en traitement automatique des langues ?

Les textes scientifiques sont porteurs d'informations mais ne contiennent pas uniquement des faits énoncés avec une *certitude absolue*. En effet, ils regorgent d'*assertions* (hypothèses, conditions, négations...) qui nuancent le discours et attribuent plus ou moins de fiabilité aux informations. Dans un système de traitement automatique des langues (TAL) dédié à des tâches comme l'indexation, le codage, la recherche ou l'extraction d'information, il est pertinent de prendre en compte les assertions et de chercher à distinguer la certitude associée à un fait ou événement. Ainsi, lorsqu'un fait ou événement est nié, comme *abdominal pain* dans l'exemple (1-a)<sup>1</sup>,

- (1) a. *The patient denies suffering from abdominal pain.*

et lorsque l'information sur la négation a pu être relevée, elle permet de ne pas générer des faux positifs et d'améliorer la qualité de l'information retrouvée. L'information sur la certitude peut aussi conduire à une description plus nuancée et riche des faits. Par exemple, dans la base de données biologiques UniProtKB/SwissProt (Bairoch *et al.*, 2004), lorsque des données issues de résultats de recherche en biologie sont enregistrées, les annotateurs indiquent aussi la certitude associée à ces données. Les assertions permettent ainsi d'annoter plus fidèlement les fonctions moléculaires actuellement connues ou supposées des gènes et protéines. Dans les exemples (2), pour les trois gènes humains *DLG4*, *DIP1* et *HE6* respectivement, les informations sont sûres pour le gène *DLG4*, mais restent hypothétiques pour les gènes *DIP1* et *HE6*.

- (2) a. *DLG4 : Cortical cytoskeleton protein found in a complex involved in maintaining cell polarity and epithelial integrity. Involved in the regulation of mitotic spindle orientation, proliferation, differentiation and tissue organization of neuroepithelial cells.*  
 b. *DIP1 : May negatively regulate cell cycle progression. May act at least in part via inhibition of the cyclin-D1/CDK4 complex, thereby preventing phosphorylation of RB1 and blocking E2F-dependent transcription.*  
 c. *HE6 : Could be involved in a signal transduction pathway controlling epididymal function and male fertility.*

Outre l'identification et l'enregistrement de la fiabilité des énoncés, il a également été proposé d'exploiter la certitude pour le marquage des développements rhétoriques dans les écrits scientifiques (Teufel, 1998 ; Mizuta *et al.*, 2006), pour l'identification de leur structure (Grabar et Hamon, 2009), pour l'atténuation des énoncés (Hyland, 1998) et pour la détection des opinions (Rillof *et al.*, 2003 ; Bethard *et al.*, 2006). Si les marqueurs de certitude sont communs à toutes ces situations, il nous semble néanmoins qu'ils jouent un rôle spécifique dans les textes cliniques. Ils marquent alors non pas les opinions des cliniciens à proprement parler mais leurs incertitudes. Ainsi,

1. Les traductions des exemples de cet article sont proposées dans l'annexe A. Elles ont été réalisées par les auteurs.

ils s'inscrivent dans un raisonnement clinique typique. Par exemple, si le médecin suspecte un problème médical, c'est-à-dire que la présence du problème médical chez le patient n'est pas certain, des tests de laboratoire sont effectués. Une fois infirmé ou confirmé, ce problème médical passe de l'incertain au factuel.

Dans ce travail, nous étudions la certitude dans la langue biomédicale. Mais ce phénomène est également très répandu dans la littérature scientifique et technique dans son ensemble (Hyland, 1998 ; Teufel, 1998 ; Rubin, 2007), de même que dans la langue générale, comme en témoigne l'étude de la presse journalistique et de l'Internet (Rillof *et al.*, 2003 ; Bethard *et al.*, 2006 ; Saurí et Pustejovsky, 2007). Dans la langue biomédicale où ce phénomène a été quantifié, il a été observé qu'en fonction du type de document clinique, entre 39 % et 83 % d'observations apparaissent dans des contextes négatifs (Chapman *et al.*, 2001). Ce chiffre peut même atteindre 95 à 99 % dans les rapports d'exams radiologiques (Mutalik *et al.*, 2001). Quant aux assertions, en fonction des catégories ciblées, il a été estimé que les résumés des articles scientifiques contiennent entre 11 % (Light *et al.*, 2004) et 19,44 % (Vincze *et al.*, 2008) de phrases modales. Citons aussi une étude des objectifs de la communication dans les textes scientifiques effectuée grâce à la détection d'indices linguistiques de modalité d'une grande granularité, où les auteurs cherchent à distinguer entre autres la dispute, la confirmation et les citations (Marco et Mercer, 2004). Les auteurs démontrent ainsi que, parmi les situations communicatives ciblées, l'incertitude apparaît le plus fréquemment dans les contextes de citation. Ces contextes applicatifs montrent l'implication sémantique de l'assertion dans la langue et la nécessité de prendre en compte cet aspect dans les applications de TAL. Dans cet article, nous proposons d'étudier les assertions dans les textes médicaux et d'implémenter leur détection automatique au sein d'un système d'extraction d'information. Bien que la tendance soit au développement de systèmes utilisant des méthodes d'apprentissage, surtout dans le cadre de campagnes d'évaluation, nous avons choisi de développer une approche qui exploite des ressources linguistiques afin de disposer d'un système facilement adaptable à un nouveau domaine de spécialité.

Nous présentons d'abord un état de l'art autour de cette question (section 2) et précisons les objectifs de notre travail (section 3). Nous présentons ensuite le matériel et les méthodes (sections 4 et 5) et décrivons les expériences effectuées et nos résultats (section 6). Nous terminons avec une conclusion et des perspectives (section 7).

## 2. État de l'art

Au sein de l'état de l'art, nous allons d'abord cerner la notion d'assertion et mentionner quelques classifications qui ont été proposées (section 2.1). Nous présenterons ensuite les travaux d'identification automatique des assertions (section 2.2).

### 2.1. La notion d'assertion et ses classifications

La notion d'assertion est définie différemment selon les auteurs et possède une portée plus ou moins étendue. Si l'assertion a pour acception la plus réduite l'affirmation d'un énoncé (Riegel *et al.*, 2004), elle peut aussi recouvrir la négation et toute une granularité de nuances, cherchant ainsi à décrire *toute prise en charge d'un énoncé par un énonciateur* (Bouscaren, 1991). Nous adoptons cette notion étendue de l'assertion.

L'assertion est liée à plusieurs autres notions, comme la polarité, la modalité, la certitude, l'incertitude, le degré de certitude, voire parfois même le temps, l'aspect et le sentiment, etc. Les plus étudiées semblent être la polarité et la modalité. La polarité peut être positive, et correspond alors à un événement reconnu comme vrai par l'énonciateur (exemple (2-a)), ou bien négative lorsque l'événement est nié (exemple (1-a)). Quant à la modalité, cette notion plus complexe divise les linguistes et reste toujours ambiguë malgré de nombreux travaux de recherche sur le sujet. Différentes classifications ont été proposées. Certaines adoptent une vision étroite et la restreignent à l'expression du possible et du nécessaire. D'autres proposent une vision plus large qui inclue les notions de permission, d'obligation et de probabilité. Nous présentons brièvement ici trois de ces classifications (Palmer, 1990 ; Culioli, 1984 ; Sauri *et al.*, 2006). Une première classification distingue trois types de modalités (Palmer, 1990) : épistémique (ou le jugement), déontique (ou la permission) et dynamique (ou la possibilité et le potentiel). Il s'agit d'une acception assez large de cette notion, mais qui permet de couvrir essentiellement les modaux. Quant à d'autres formes porteuses de modalité (comme le présupposé, la négation et l'interrogation, les souhaits et les peurs, l'emploi du passé), elles sont considérées à part. Une deuxième classification distingue quatre types de modalités (Culioli, 1984 ; Vion, 2001) : neutre (*i.e.*, affirmation, négation, interrogation, impératif), épistémique (*i.e.*, possibilité, probabilité, certitude), appréciative ou affective (*i.e.*, jugement, implication) et intersubjective (*i.e.*, permission, ordre). Il s'agit d'une acception plus étendue de la modalité. Notons que la modalité neutre est une catégorie qui n'apparaît pas chez tous les auteurs. Enfin, la troisième classification couvre un spectre complexe des types de modalités et vise à les représenter sur un axe de véridicité allant du factuel au contrefactuel (Sauri *et al.*, 2006), comprenant les degrés de possibilité, l'opinion, l'évidentialité (ou la preuve), la prédiction, la tentative et l'ordre. Notons que cette classification monoaxiale sera jugée par la suite trop réductrice et sera abandonnée par les auteurs pour donner la priorité à un seul type de modalité : la modalité épistémique qui semble alors regrouper les trois premiers types (degrés de possibilité, opinion et évidentialité). La modalité épistémique, appelée également *spéculation* ou *hedging*, est la catégorie la plus stable dans ces trois classifications, de même que chez d'autres auteurs (Lakoff, 1972 ; Lyons, 1977 ; Hyland, 1998 ; Vold, 2008 ; Sauri, 2008). Il s'agit également du type le plus usité dans les textes scientifiques, car il correspond aux situations où l'auteur se prononce sur le statut factuel d'une proposition. Trois fonctions sont associées à la modalité épistémique (Hyland, 1998) :

1) exprimer une proposition avec précision en suivant le raisonnement de l'énonciateur, ce qui permet de faire une différenciation entre le réel et le possible ;

2) détacher l'auteur de tout engagement de sa responsabilité personnelle vis-à-vis d'un fait, qui se traduit par l'utilisation des modaux ;

3) développer la relation énonciateur-lecteur : il s'agit de ne pas énoncer de vérité absolue dans les articles scientifiques. Le recours à la modalité permet ainsi d'atténuer et de nuancer le discours.

Plusieurs travaux ont aussi cherché à analyser et à proposer une gradation de la certitude. La finesse varie en fonction des applications et des besoins, et les noms donnés aux degrés sont rarement les mêmes. Ainsi, la gradation se compose d'au moins deux degrés, faible et élevé, (Grabar et Hamon, 2009), ou de trois degrés (Hoye, 1997 ; Light *et al.*, 2004 ; Saurí, 2008 ; Jilani *et al.*, 2008), par exemple, la certitude, la probabilité et la possibilité (Hoye, 1997). Elle peut aussi inclure quatre degrés (certitude absolue, élevée, modérée, faible et incertitude) (Rubin *et al.*, 2006 ; Wilbur *et al.*, 2006 ; Thompson *et al.*, 2008), cinq (Rubin, 2007), en allant jusqu'à dix (Battistelli et Amardeilh, 2009). Dans ces travaux, la finesse de la gradation de la certitude varie également en fonction des besoins, mais également de la capacité des outils de TAL à détecter ces degrés de certitude. Par ailleurs, certaines langues, comme l'anglais, ne sont pas toujours suffisamment précises pour distinguer plusieurs nuances de la certitude (Thompson *et al.*, 2008).

## 2.2. Identification automatique des assertions

En TAL, la plupart des applications se concentrent sur la recherche et l'extraction d'informations factuelles. Jusqu'à présent, peu de tentatives ont été réalisées pour automatiser l'identification des assertions. Ceci était certainement dû à la complexité du recensement de la variété d'expressions linguistiques pour qualifier un énoncé, mais aussi à la difficulté de leur modélisation et de l'identification de la portée. Notons que dans le domaine biomédical, cette recherche a été récemment stimulée par plusieurs campagnes d'évaluation incluant des tâches sur les assertions : BioScope<sup>2</sup>, BioNLP SharedTasks<sup>3</sup> et I2B2<sup>4</sup>. La difficulté dans l'identification de la certitude ressort à travers l'état de l'art, autant parmi les systèmes à base d'apprentissage que dans les approches linguistiques.

Parmi les méthodes d'apprentissage, l'apprentissage supervisé domine. Par exemple, un classifieur SVM et des descripteurs composés de sous-chaînes de caractères et de mots simples, ont été appliqués pour la détection de l'incertitude dans les articles scientifiques (Light *et al.*, 2004). La meilleure précision (0,95) est obtenue avec des sous-chaînes de caractères, mais le rappel n'est pas indiqué pour cette expérience. De plus, la caractérisation des types d'incertitude est grossière et l'utilisation de connaissances linguistiques plus sophistiquées pour affiner le typage et d'obtenir une granularité plus fine, n'est qu'une perspective. Ce travail a été réexploité pour effectuer

2. <http://www.inf.u-szeged.hu/rgai/bioscope> (données accessibles).

3. <http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/> (données non accessibles).

4. <https://www.i2b2.org/NLP/> (données accessibles sous licence).

une classification binaire incertain/certain en utilisant une méthode d'apprentissage semi-supervisé (un modèle probabiliste) et les amorces *suggest* et *likely* (Medlock et Briscoe, 2007). La précision est améliorée mais la catégorisation reste grossière et ne permet pas d'identifier les énoncés affirmatifs. Une des rares expériences à avoir exploité l'apprentissage non supervisé poursuit l'objectif de catégorisation des noms selon leur classe sémantique et leur subjectivité (Rillof *et al.*, 2003). Les auteurs se fondent pour ceci sur le contexte des noms et obtiennent une F-mesure de 76 %.

Le fonctionnement des approches linguistiques est fondé le plus souvent sur des ressources linguistiques comme les marqueurs (mot, abréviation, groupe de mots, préfixe, suffixe, base morphologique, etc.), une analyse syntaxique ou des patrons lexico-syntaxiques (description lexicale et syntaxique des contextes sémantiquement significatifs). Par exemple, une annotation manuelle des textes biomédicaux (Wilbur *et al.*, 2006) s'appuie sur de tels marqueurs et prévoit une caractérisation fine des phrases : centre d'intérêt (générique, scientifique, méthodologie), preuve (de 0 à 3), certitude (de 0 à 3), polarité (positive, négative) et tendance (+, -). Cette annotation est effectuée pour préparer le fonctionnement des systèmes automatiques. Un travail exploitant un algorithme d'analyse descendante et des dépendances syntaxiques, combinés à des marqueurs grammaticaux et des contextes de subordination a été aussi proposé (Saurí et Pustejovsky, 2007). En effet, ces deux derniers sont souvent porteurs des assertions. Les performances de F-mesure obtenues dans ce travail varient entre 0,74 et 0,85. Par ailleurs, le calcul de la portée de l'incertitude et de la négation a été abordé au sein de la compétition BioScope mentionnée plus haut. Les systèmes présentés ont souvent exploité des ressources linguistiques et des règles. Les performances du meilleur système en terme de F-mesure est de 0,24 (Kilicoglou et Bergler, 2009). Une autre compétition, I2B2 2010, a offert la possibilité de comparer différentes approches pour l'identification de six catégories d'assertion dans des textes médicaux. Les systèmes présentés s'appuient généralement sur des méthodes d'apprentissage supervisé ou des approches mixtes (apprentissage et ressources). Lorsqu'elles sont indiquées, les performances des systèmes varient entre 0,79 et 0,93. Nous avons présenté ici les performances de quelques systèmes. Toutefois, la comparaison entre ces performances reste difficile étant donné la différence des tâches poursuivies, du niveau de granularité des assertions à identifier et du matériel traité.

Pour une annotation fine des assertions, notons aussi que la nécessité d'exploiter un ensemble riche de marqueurs est souvent ressentie dans les travaux. En effet, les modaux ne jouent qu'un rôle modeste dans l'expression de la modalité (Thompson *et al.*, 2008), tandis que les verbes lexicaux (*examine, evaluate, analyze, study, investigate*) et les adjectifs en constituent d'autres indices importants (Holmes, 1988 ; Thompson *et al.*, 2008 ; Kilicoglou et Bergler, 2009), de même que la ponctuation (Grabar et Hamon, 2009) ou l'analyse syntaxique (Saurí et Pustejovsky, 2007 ; Kilicoglou et Bergler, 2009). Nous pouvons donc voir que l'assertion est un phénomène linguistique complexe et que son identification automatique demande une gamme d'indices d'autant plus large que la catégorisation ciblée est fine.

### 3. Objectifs scientifiques et contexte du travail

Étant donné la multitude des marqueurs qui permettent d'associer une certitude à une information et la complexité du phénomène, nous proposons d'étudier leur contribution à l'identification des assertions dans des textes médicaux. Nous nous plaçons pour cela dans le contexte du développement d'un système d'extraction d'information et nous étudions l'impact de chaque marqueur sur la qualité du système.

Afin de disposer des données externes d'entraînement et d'évaluation, nous nous positionnons dans le cadre de la compétition TAL internationale I2B2<sup>5</sup> tenue en 2010 et consacrée à l'extraction d'information à partir des dossiers cliniques. Comme il s'agit d'une compétition internationale, les dossiers cliniques sont en anglais. Les exemples de cet article sont extraits de ce corpus (voir section 4.1). Trois types d'informations sont ciblés dans l'édition 2010 de la compétition I2B2 à travers les tâches suivantes :

**tâche 1 :** termes médicaux et leur typage en fonction de trois catégories possibles : problèmes médicaux, tests de laboratoires ou d'imagerie et traitements médicaux ;

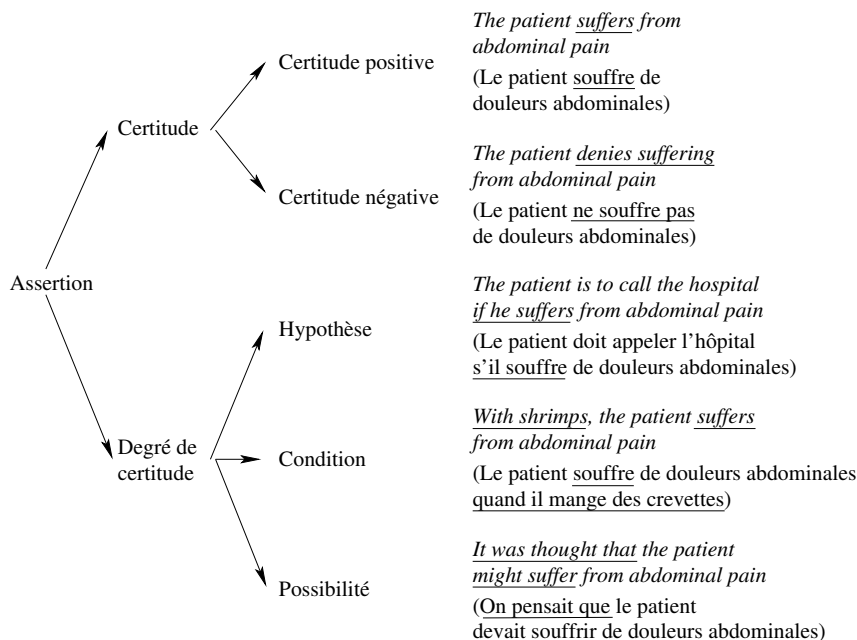
**tâche 2 :** assertions associées aux problèmes médicaux ;

**tâche 3 :** relations existant entre les problèmes médicaux et les traitements, les tests ou d'autres problèmes médicaux.

Il s'agit donc d'un tableau assez complexe de ce que peut représenter un processus clinique de soins et d'investigations. L'objectif fondamental de cette compétition consiste justement à expliciter les entités et les relations médicales et à les structurer afin de les rendre accessibles aux applications et aux utilisateurs humains. Dans ce travail, nous nous intéressons uniquement à l'identification de l'assertion associée au problème médical (tâche 2), celui-ci étant fourni en entrée du système par la tâche 1. L'ensemble des assertions se répartit en cinq catégories. Les plus fréquemment rencontrées dans les documents cliniques d'observation des patients et de leurs soins sont les suivantes : présent, hypothétique, conditionnel, possible et absent. À ceci s'ajoute une autre catégorie, également fréquente dans les documents cliniques, lorsqu'un problème médical est associé à une autre personne. Elle n'est pas directement liée aux assertions mais joue plutôt un rôle de filtre, et intervient comme catégorie prioritaire dans l'ordre défini dans la compétition que nous présentons dans le paragraphe suivant. Ainsi, les problèmes médicaux qui ne sont pas concernés par cette catégorie sont associés au patient et participent à la relation *{patient, problème médical}*. Étant donné les différentes classifications que nous avons présentées dans l'état de l'art et les catégories ciblées dans notre travail, nous proposons de distinguer au sein de l'assertion la polarité de la certitude et les degrés de certitude (figure 1). Comme le montre cette figure, pour la polarité nous distinguons deux catégories classiques : la certitude positive et la certitude négative. Les degrés de certitude correspondent à l'hypothèse, la condition et la possibilité. Nous donnons aussi des exemples afin de rendre la distinction entre ces catégories plus explicite. Une explication plus détaillée est proposée à la

5. *Informatics for Integrating Biology & the Bedside*, <https://www.i2b2.org/NLP/Relations>

section 5. Afin de pouvoir étudier en détail l'apport de chaque indice ou marqueur de l'assertion, nous proposons de mettre en application un système automatique à base de marqueurs et de ressources sémantiques, et plus particulièrement, d'effectuer une étude incrémentale de ces marqueurs et de leur influence sur les résultats.



**Figure 1.** Structuration des catégories d'assertion pouvant être rattachées aux problèmes médicaux associés à un patient : la polarité et les degrés de certitude. La catégorie associée à quelqu'un d'autre reste en dehors de cette structuration.

En plus des types d'assertion visés dans notre travail, le contexte de cette compétition nous impose deux autres contraintes. Un ordre de priorité prédéfini doit être respecté. Celui-ci est utile lorsque plusieurs indices d'assertion sont associés à un même problème médical. L'assertion assignée est alors celle qui apparaît en premier dans l'ordre imposé : non associé au patient, hypothétique, conditionnel, possible, absent et présent. Ainsi, l'importance est d'abord donnée au rapport avec le patient : tout problème lié à une autre personne est écarté. Puis viennent les degrés de certitude et enfin la certitude : l'expression d'une hypothèse, d'une condition ou d'un doute émis sur le problème médical amène à ne pas considérer la polarité positive ou négative. Prenons par exemple la phrase en (3-a), où deux marqueurs sont présents dans le contexte gauche du terme *contamination* : *probably* signifiant la possibilité et *not* signifiant la certitude négative. En accord avec l'ordre de priorité imposé par la compétition, le problème médical *contamination* doit être annoté comme *possible* : la catégorie *possible* est en effet prioritaire sur la catégorie *absent*. Ainsi, quand plusieurs indices de degrés de certitude sont présents, l'incertitude la plus élevée est prioritaire.



- (3) a. *The patient remained afebrile, and it was felt that it was probably not a contamination.*

Une autre contrainte imposée par la compétition porte sur l'évaluation. Les résultats produits sont évalués par rapport aux données de référence fournies dans le cadre de la compétition. Trois métriques classiques sont utilisées : précision  $P$  (pourcentage d'assertions correctes générées par le système au sein de l'ensemble des assertions générées), rappel  $R$  (pourcentage des assertions correctes générées par le système au sein des assertions de référence attendues) et F-mesure  $F_1$  (moyenne harmonique de la précision et du rappel) :  $F = \frac{(1+\beta^2) \times P \times R}{\beta^2 \times P + R}$ , ici  $\beta = 1$ .

#### 4. Identification des assertions en corpus

##### 4.1. Description du corpus de documents cliniques

Le corpus fourni par la compétition comporte 1 653 documents cliniques, divisés en deux sous-ensembles : les données d'entraînement (1 176 documents, 772 209 occurrences) et les données de test (477 documents, 396 172 occurrences). Les données d'entraînement sont disponibles pour une durée de trois mois. Ces données initiales nous permettent de développer notre méthode fondée sur des ressources linguistiques (essentiellement des marqueurs lexicaux et morphologiques). Nous avons également introduit un mécanisme de désambiguïsation des marqueurs, décrit à la section 4.2. Il est exploité lorsque plusieurs marqueurs sont présents de manière concurrente. Au sein du corpus d'entraînement, seulement 349 documents sont annotés, tandis que 827 documents restent bruts. Les documents sont écrits en langue anglaise. Pour garantir la confidentialité des patients, les documents ont été anonymisés par les organisateurs : toutes les traces permettant d'identifier le patient, l'établissement ou le médecin ont été remplacées par des noms fictifs ou cachés (Sibanda et Uzuner, 2006).

Le corpus fourni comporte trois types de documents cliniques (comptes-rendus hospitaliers, résumés de sortie et notes d'évolution des patients), qui correspondent à des genres littéraires différents bien que proches. Ces documents proviennent de trois centres hospitaliers américains, ce qui conduit à des différences de contenu et de structure. De plus, les services qui fournissent les documents couvrent presque l'ensemble des services pouvant exister dans un hôpital (par exemple, maternité et néonatalité, urgences, cardiologie, chirurgie et chirurgie plastique, urologie, hématologie, gynécologie, orthopédie, médecine interne, neurologie...). Tout ceci fait que les documents sont très variés, autant par leur contenu et leur taille que par leur structure, tout en sachant que la structure reste implicite. La figure 2 présente un extrait de document clinique. Dans ce document, les séquences comme *HISTORY OF PRESENT ILLNESS*, *PAST MEDICAL HISTORY* ou *Vital signs* correspondent aux titres des sections. Par ailleurs, dans la section *MEDICATIONS ON ADMISSION*, les noms et dosages des médicaments apparaissent sous forme d'une liste. Ceci n'est bien sûr pas le cas dans tous les documents du corpus. Nous constatons également quelques particularités de

**HISTORY OF PRESENT ILLNESS :**

*The patient is an 82-year-old male with a history of hypertension, remote tobacco abuse, chronic renal insufficiency, and a history of a TIA who was in his usual state of health until five nights prior to admission when he awoke at midnight with severe epigastric pain radiating to both arms. He also had bilateral arm numbness...*

**PAST MEDICAL HISTORY :**

*TIA in 1998.*

*COPD with 2 liters dependence of home O<sub>2</sub>.*

*BPH.*

*Chronic renal insufficiency.*

*Gastritis.*

*Asthma.*

**SOCIAL HISTORY :**

*He is very active.*

*He uses a cane p.r.n. for walking.*

*He denied any falls at home...*

**ALLERGIES :**

*The patient has no known drug allergies.*

**MEDICATIONS ON ADMISSION :**

*1. Aspirin 325 mg q.d.*

*2. Metoprolol 50 mg b.i.d.*

*3. Lipitor 10 mg q.d.*

*4. MDI p.r.n.*

**PHYSICAL EXAMINATION :**

*Vital signs :*

*Heart rate 91, regular, blood pressure 169/81 from a right groin A line, respirations 12, oxygen saturation of 97 percent on 2 liters nasal cannula.*

*General :*

*He was a well appearing male in no acute distress.*

**HEENT :**

*Crusting around the right eyelid but the pupils were equal, round, and reactive to light.*

*He was edentulous...*

**Cardiovascular :**

*His heart was regular with a normal S1 and S2.*

*He had no murmurs, rubs, or gallops...*

**LABORATORY DATA :**

*On admission, the CBC revealed a white count of 7.4, hematocrit 39.7, and platelets of 217,000...*

*He had a bedside echocardiogram done in the Catheterization Laboratory which showed normal LV wall thickness and normal LV cavity size...*

**Figure 2.** Extrait d'un document clinique

la langue médicale clinique, telles qu'une terminologie abondante, un usage abondant d'abréviations (*p.r.n.*, *TIA*, *COPD*, *BPH*, *MDI*, *CBC*, ...) et des données numériques.

#### 4.2. Système d'identification des assertions

Pendant l'étape de prétraitement, les documents sont convertis au format XML, ce qui est requis par la plate-forme d'annotation Ogmios (Hamon et Nazarenko, 2008) qui permet entre autres l'agencement des outils du TAL. Le format XML est utilisé en particulier pour encoder et rendre plus explicite l'information sur la structure des documents : les sections et les listes y sont marquées. Les sections, les titres de sections et les listes numérotées (par exemple, la liste des médicaments dans la section *MEDICATIONS ON ADMISSION* de la figure 2) sont identifiés à l'aide de marqueurs, de la casse, de la ponctuation et de l'apprentissage endogène. Quant aux listes syntaxiques, comme *He had no murmurs, rubs, or gallops...* de la figure 2, elles sont identifiées grâce à une analyse syntaxique (Klein et Manning, 2003). Le format XML est également utilisé pour enregistrer l'ensemble des annotations qui sont réalisées. Au travers de la plate-forme, nous effectuons la reconnaissance et l'annotation des entités nommées, ainsi que l'étiquetage morphosyntaxique (Tsuruoka *et al.*, 2005). C'est également au travers de la plate-forme que sont projetées les ressources sémantiques, que nous décrivons dans la section 5.

La décision sur les assertions est fondée sur l'exploitation de marqueurs (dont la constitution est décrite dans la section 5), d'informations sur la structure des documents et de l'ordre de priorité des catégories d'assertion. L'identification des assertions prend en compte les marqueurs au sein de la phrase, mais elle peut en dépasser les limites s'il s'agit des titres de sections. Lorsqu'une relation  $\{patient, problème\}$  peut recevoir plusieurs catégories d'assertion ( $ast_i$ ), nous calculons, pour chaque catégorie d'assertion concurrente  $ast_i$ , la somme pondérée du nombre d'occurrences de ses marqueurs :

$$weight(ast_i) = \sum_{j \in \{lex, base, m-p, m-s, context, struct\}} \alpha_j \times nbocc_j(ast_i)$$

où les poids assignés aux marqueurs sont :  $\alpha_{lex} = 1$  (lexème),  $\alpha_{adj} = 0,8$  (adjectif),  $\alpha_{m-p} = 0,5$  (préfixe),  $\alpha_{m-s} = 0,5$  (suffixe),  $\alpha_{base} = 0,25$  (base morphologique),  $\alpha_{context} = 1$  (marqueur contextuel) et  $\alpha_{struct} = 1$  (marqueur structurel).  $nbocc_j(ast_i)$  est le nombre d'occurrences de marqueurs de type  $j$  pour la catégorie d'assertion  $ast_i$ . En cas de catégories concurrentes, la catégorie attribuée est alors celle qui a le poids le plus élevé  $weight(ast_i)$ . Enfin, en dernier recours, s'il n'est toujours pas possible de choisir une seule catégorie, nous prenons en compte l'ordre de priorité défini par la compétition. Nous avons choisi de donner moins de poids aux marqueurs morphologiques car, selon nous, leur fiabilité sémantique est plus faible que celle des marqueurs lexicaux et adjectivaux (voir par exemple les phrases en (5-e) et (5-f) par rapport aux phrases en (5-a), (5-b) ou (5-c)). De plus, au sein des marqueurs morphologiques, la granularité est assez fine car nous considérons que ces différents

types de marqueurs ont des implications sémantiques différentes : les préfixes privatifs comme *a-* ou *dys-* ont une valeur sémantique de négation fiable, alors qu'un mot contenant la chaîne *free* ne se rapporte pas toujours à la négation. Cependant, dans la catégorie *absent*, les marqueurs morphologiques autres que préfixes (base et suffixe) sont extrêmement rares. En suivant le même raisonnement, certains indices, comme les marqueurs contextuels et structurels, présentant le moins d'ambiguïté sont privilégiés grâce à l'assignation de poids élevés. Les valeurs des poids ont été définies empiriquement à partir d'observations en corpus, mais ces valeurs pourront être affinées dans un travail futur. Nous envisageons, notamment, d'évaluer l'apport d'une approche par apprentissage pour ajuster ces poids. Enfin, dans les listes syntaxiques, la portée des marqueurs est propagée à tous les éléments de cette liste.

## 5. Constitution de ressources pour l'identification des assertions

Afin d'identifier les assertions, nous avons constitué un ensemble de marqueurs. Nous avons également préparé et exploité des ressources terminologiques et sémantiques existantes. En ce qui concerne les indices d'assertion, nous avons observé qu'ils peuvent être de trois types :

1) certains indices se trouvent au sein des termes comme dans les exemples (5-e) et (5-f). Nous les appelons marqueurs *intratermes*. Cette situation est surtout propre à la polarité négative. Ce type de marqueurs peut correspondre aux préfixes (souvent des préfixes privatifs), suffixes, lexèmes en entier ou toute base morphologique ;

2) d'autres indices se trouvent dans le voisinage des termes à caractériser, comme dans les exemples (5-a) à (5-c) pour la négation. Ces marqueurs que nous appelons *contextuels* se trouvent dans la même phrase que le terme, à sa gauche ou à sa droite ;

3) enfin, d'autres indices s'appuient sur la structure des documents, comme les titres de sections. Ces indices sont appelés *structurels*. Par exemple, si un problème médical se trouve dans la section *FAMILY HISTORY*, il est fort probable qu'il s'agisse d'un problème médical lié à une autre personne que le patient.

Dans la suite de cette section, nous présentons la constitution de la ressource des marqueurs pour les six catégories d'assertion qui doivent être reconnues. Les marqueurs peuvent appartenir aux trois types mentionnés (intraterme, contextuel et structurel). Étant donné la nature internationale de la compétition I2B2, les exemples sont en anglais. Une traduction des exemples est disponible à l'annexe A.

### 5.1. Certitude à polarité positive : problème médical présent

La certitude positive correspond à tous les problèmes médicaux qui sont affirmés, déclarés certains ou reconnus comme existants bel et bien chez le patient, comme dans l'exemple (4-a). Il s'agit de la catégorie par défaut, qui est associée aux problèmes médicaux lorsque aucune autre catégorie d'assertion n'a pu y être associée. Cette catégorie ne comporte donc pas de marqueur spécifique.

- (4) a. *The patient had a stroke.*

### 5.2. Certitude à polarité négative : problème médical absent

La négation peut être marquée par différentes catégories grammaticales, comme les déterminants, les verbes, les adverbes, les expressions adjectivales et les préfixes dans les exemples en (5).

- (5) a. *No aortic regurgitation present.*  
 b. *Review of systems shows that the patient denies any orthopnea.*  
 c. *The patient has continued to feel well without any symptoms of shortness of breath.*  
 d. *Lower extremity dopplers negative for deep venous thrombosis.*  
 e. *ABDOMEN : soft, nontender, nondistended.*  
 f. *She is afebrile with stable vital signs.*

La marque de la négation se situe surtout dans le contexte gauche du terme, comme les locutions *no history of* ou *no findings suggestive of* dans les exemples en (6).

- (6) a. *No history of abnormal pap smears.*  
 b. *There were no findings suggestive of an accompanying hollow viscous perforation.*

La certitude négative peut aussi porter sur le prédicat et non sur le groupe nominal, ce qui est un mécanisme très courant (Saurí, 2008, p.44). Ainsi, dans les exemples en (7), les événements sont infirmés au travers de la négation de leurs prédicats.

- (7) a. *She [was] not [felt to be] acutely suicidal.*  
 b. *The baby did not [turn] blue at that time.*  
 c. *She did not have any further episodes of blurry vision or chest pain.*

Un cas particulier apparaît lorsqu'il s'agit d'un problème médical soigné. Dans ce cas, les marqueurs se situent le plus souvent après le problème médical. Ces marqueurs comportent principalement des verbes *de guérison* (*resolve, improve, clear*).

- (8) a. *Bradycardia is resolved after beta blockers and calcium channel blockers were stopped and norvasc was started.*  
 b. *By the following morning all mental status changes had cleared.*  
 c. *Lower extremity edema improved with diuresis.*

Nous exploitons un ensemble de 158 marqueurs de polarité négative. Une partie de ces marqueurs provient du travail précédent (Chapman *et al.*, 2001), que nous

avons adapté aux données observées en corpus. La ressource contient 25 marqueurs intratermes, 118 marqueurs qui précèdent les entités niées et 15 qui les suivent.

### 5.3. Hypothèse : problème médical hypothétique

Lorsque la catégorie *hypothétique* est associée à un problème médical, il s'agit d'un problème qui n'est pas constaté dans le présent mais qui peut se développer dans le futur. Dans notre matériel, l'hypothèse est surtout marquée par *if*, *with* et *for* quand ceux-ci introduisent des propositions subordonnées avec des structures précises :

- (*if/with/for*) (*there is/there are*) problème médical
- (*if/with/for*) pronom-personnel verbe problème médical
- (*if/with/for*) problème médical

De plus, ces structures peuvent être nuancées par le quantifieur *any* (très souvent ambigu (Larrea et Rivière, 1999)), lorsqu'il apparaît avant le problème médical, ou par un modal situé avant le verbe. Ces structures subordonnées accompagnent souvent une proposition principale exprimant une injonction ou bien une principale au futur (*will*) ou au conditionnel (*would*). L'injonction est souvent exprimée à l'aide de : (1) verbes *de prise de contact* comme *call*, *contact*, *notify*, (2) verbes *de déplacement* comme *return*, *go to*, ou (3) verbes *de communication* comme *tell*, *advise*, *ask*, *inform*, *instruct*. Ces différents cas sont exemplifiés en (9).

- (9) a. *He was to contact our office if there is any drainage.*  
 b. *If any abdominal pain, fever or dizziness, go to the emergency room.*  
 c. *Return immediately for worsening signs of infection.*  
 d. *He was advised to return to the emergency department, or call his physician for any new symptoms.*

Notons que certaines phrases ont un niveau d'incertitude très élevé. La phase en (10-a) comporte un *if* et une suite de *should* dans la principale et dans la subordonnée.

- (10) a. *If his heart should stop or he should have a lethal arrhythmia, we should not do any chest compressions.*

Les problèmes médicaux susceptibles de se développer devant être surveillés, l'hypothèse est aussi exprimée par les verbes *de prévention* comme *detect*, *prevent*, *monitor* ou *watch*.

- (11) a. *Watch for BZD withdrawal, agitation.*  
 b. *Wound care consult given anasarca to prevent decub.*  
 c. *Monitor wounds for infection - redness, drainage, or increased pain.*

Un autre cas particulier concerne la prescription de médicaments : des médicaments peuvent en effet être prescrits au cas où ils seraient nécessaires. Dans ces cas, des expressions de *besoin* comme *as necessary*, *as needed* (12-b) et (12-c) ou *prn* (expression latine *pro re nata* signifiant *autant que nécessaire*) (12-a) sont utilisées.

- (12) a. *Ativan 0.5 mg PO q4 hours prn for anxiety.*  
 b. *She was given Duoneb as needed for any shortness of breath.*  
 c. *Docusate Sodium 100 mg Capsule Sig : One (1) Capsule PO BID (2 times a day) : take as necessary for constipation.*

Nous exploitons un ensemble de 45 marqueurs d'hypothèse : 1 marqueur intraterme, 42 marqueurs dans le contexte gauche des problèmes médicaux et 2 dans le contexte droit.

#### 5.4. Condition : problème médical apparaissant sous certaines conditions

Le statut *conditionnel* est associé avec un problème médical lorsque celui-ci apparaît seulement dans certaines situations et dans des conditions particulières. Il s'agit par exemple, de symptômes qui apparaissent suite à un effort physique (exemples (13-a) à (13-d)) ou à une allergie (exemples (13-e) à (13-h)). Dans l'exemple (13-h), l'indice de structure (*ALLERGIES* : est reconnu en réalité comme nom de section) permet de statuer sur la catégorie d'assertion.

- (13) a. *Patient reports shortness of breath upon climbing stairs.*  
 b. *He was also short of breath with ambulation.*  
 c. *He can climb approximately three flights of stairs before becoming short of breath.*  
 d. *He reports severe dyspnea on exertion.*  
 e. *Penicillin causes a rash.*  
 f. *The patient is allergic to Ciprofloxacin, morphine sulfate and Droperidol.*  
 g. *Allergies included penicillin and iodine.*  
 h. *ALLERGIES : TOMATO (causing anaphylaxis) and TETANUS SHOT.*

Les aliments et les médicaments peuvent également être la cause de problèmes médicaux, comme dans les exemples (14).

- (14) a. *The patient gets very somnolent with Ativan.*  
 b. *She also has some shortness of breath and throat constriction with shrimp.*

Nous avons constitué un ensemble de 24 marqueurs de condition : 4 marqueurs intratermes, 13 marqueurs dans le contexte gauche, 2 dans le contexte droit, et 5 mar-

queurs structurels. Malgré la diversité des situations décrites et relatives à cette catégorie, la ressource est relativement réduite. En effet, cette catégorie bénéficierait surtout d'une analyse syntaxique venant appuyer l'utilisation de marqueurs.

### 5.5. Possibilité : problème médical suspecté

Lorsque la catégorie *possible* est en jeu, il s'agit d'un problème médical que le médecin suspecte chez un patient mais pour lequel il n'existe pas de preuves, comme dans l'exemple en (15).

- (15) a. *This is very likely to be an asthma exacerbation.*  
 b. *Questionable pneumonia.*

L'incertitude sur le problème médical peut être exprimée avec différentes catégories sémantiques et grammaticales : adjectifs (15-b), noms (16-c), adverbes (15-a) et (16-a), participes (16-b), et bien sûr les modaux (exemples (16-d) à (16-f)).

- (16) a. *A lucency and irregularity of the glenoid, possibly representing a fracture, was also seen.*  
 b. *Florinef 0.1 mg po q day for presumed adrenal insufficiency.*  
 c. *He was found on physical exam to have an asymmetric prostate with prominence on the left and a question of a nodule.*  
 d. *He could have a baseline high and increased creatinine.*  
 e. *Slight irregularity of the origin of the right vertebral artery may be atherosclerotic in nature.*  
 f. *The patient was already on an aspirin and plavix should this represent a small stroke.*

Le médecin peut aussi exprimer le diagnostic sous forme d'une suspicion (exemples (17-a) et (17-b)) ou comme une précaution (exemples (17-c) et (17-d)).

- (17) a. *We suspected aspiration or a chemical pneumonitis.*  
 b. *Still suspicion that she had a stroke while off her coumadin.*  
 c. *The acute onset of chest pain was initially very concerning for a PE.*  
 d. *The patient was started on Florinef for concern of adrenal insufficiency.*

Lorsque le problème médical est suspecté, le recours à la preuve (un contrôle (18-a) à (18-d) ou un examen (18-e) et (18-f)) est souvent nécessaire. Dans ce cas, les marqueurs *de preuve* (*make sure, monitor, check, determine, rule out*) et les verbes *d'épreuve* (*go through, undergo*) des exemples (18-g) et (18-h) marquent la possibilité.

- (18) a. *We will make sure there is no associated B12 deficiency as well.*  
 b. *We will monitor for signs of alcohol withdrawal.*



- c. *She was admitted to rule out an acute cardiac event.*
- d. *Please r/o acute process.*
- e. *Troponins were checked for a chest pain.*
- f. *A magnetic resonance imaging study will be scheduled as an outpatient in three months to rule out a small vascular malformation if responsible for the hemorrhage.*
- g. *He underwent lower extremity study to rule out a DVT in both legs.*
- h. *He then underwent a CT scan to rule out a PE.*

D'autres marqueurs peuvent également conduire à l'identification de la possibilité, mais ils sont en général plus ambigus et fluctuants lorsque intégrés dans un système automatique. C'est le cas des prépositions ou conjonctions comme *versus* (exemples (19-a) et (19-b)) ou *whether* (19-c), ou du point d'interrogation dans (19-d) et (19-e).

- (19) a. *A repeat chest x-ray done for dyspnea showed persistent pleural effusion with atelectasis versus pneumonia.*
- b. *Benign positional vertigo vs labyrinthitis.*
- c. *Discussion was had at this time with the patient regarding whether or not this represented an infected abscess vs. post-surgical change.*
- d. *DM ? Arthritis.*
- e. *Bleed ?*

Nous avons constitué un ensemble de 151 marqueurs de possibilité : 120 marqueurs en contexte gauche et 31 en contexte droit. Le nombre important de marqueurs résulte de l'observation d'une grande diversité de situations.

### 5.6. Autre personne : problème médical associé à une autre personne

Si un document clinique contient essentiellement les informations relatives au patient, il peut aussi contenir des informations relatives à d'autres personnes, notamment aux membres de sa famille ou au personnel soignant. Comme nous l'avons indiqué plus haut (section 3), cette catégorie sert de filtre et permet de détecter les problèmes médicaux qui ne concernent pas le patient. La catégorie *associé à une autre personne* est identifiable à l'aide de marqueurs de structure (20-b), de marqueurs de relations de parenté (20-a) et (20-d), ou d'autres relations sociales (20-c) et (20-d).

- (20) a. *The patient's father died of a myocardial infarction at age 40.*
- b. *Family history : not significant for coronary artery disease.*
- c. *Avoid individuals who may be ill.*
- d. *The family and other caregivers should be considered for immunization against influenza.*

Nous avons établi un ensemble de 7 marqueurs (4 lexicaux et 3 structurels).

## 6. Expériences en identification et rattachement des assertions

Nous présentons les résultats obtenus avec tous les marqueurs que nous avons relevés sur le corpus d'entraînement. Cette ressource a été appliquée aux corpus d'entraînement et de test avec le système décrit dans la section 4, puis évaluée. Les entrées de notre système sont les documents cliniques et l'ensemble des problèmes médicaux fournis par la tâche 1. Nous présentons et discutons d'abord les résultats globaux obtenus sur les corpus d'entraînement et de test (section 6.1) ; nous décrivons ensuite l'étude incrémentale de l'influence des marqueurs sur le corpus d'entraînement (section 6.2) ; enfin, nous montrons comment cette étude incrémentale a permis d'améliorer nos résultats (section 6.3).

### 6.1. Résultats globaux

Le tableau 1 indique le nombre et le pourcentage d'assertions dans les corpus d'entraînement et de test, ainsi que la répartition par catégorie. Les données de test sont légèrement plus volumineuses, globalement et pour chaque catégorie. Mais la distribution par catégorie reste stable. Les deux catégories les plus volumineuses sont *présent* et *absent*. Les catégories *possible* et *hypothétique* sont moyennement fréquentes, alors que les catégories *conditionnel* et *autre personne* sont très peu fréquentes.

Catégories	Entraînement	%	Test	%
Autre personne	92	0,77	145	0,78
Hypothétique	651	5,44	717	3,86
Conditionnel	103	0,86	171	0,92
Possible	535	4,47	883	4,76
Absent	2 536	21,19	3 609	19,45
Présent	8 052	67,27	13 025	70,22
Total	11 969	100	18 550	100

**Tableau 1.** Nombre et pourcentage d'assertions par catégorie et globalement, dans les corpus d'entraînement et de test

Le tableau 2 indique les performances globales obtenues lors de l'identification et l'association des assertions aux problèmes médicaux dans les corpus d'entraînement et de test. Les performances sont indiquées en termes de précision, de rappel et de F-mesure. Ainsi, pour les assertions en général, la valeur de la F-mesure est égale à 0,80 points, de même que les valeurs de la précision et du rappel. Une telle distribution des valeurs est due au fait qu'il s'agit d'un ensemble fermé de catégories et de problèmes médicaux : un faux positif dans une catégorie correspond à un faux négatif dans une autre catégorie. Les valeurs globales des trois mesures d'évaluation sont donc identiques. En revanche, si nous observons les valeurs obtenues pour chaque catégorie séparément, des différences sont à remarquer. Ainsi, trois catégories (*présent*, *absent* et *autre personne*) sont les mieux identifiées, avec des valeurs de F-mesure entre 0,79

Catégories	Entraînement			Test		
	P	R	F	P	R	F
Autre personne	0,96	0,80	0,88	0,84	0,74	0,79
Hypothétique	0,71	0,31	0,43	0,63	0,24	0,35
Conditionnel	0,08	0,40	0,14	0,08	0,33	0,12
Possible	0,46	0,57	0,51	0,51	0,47	0,49
Absent	0,92	0,75	0,82	0,87	0,75	0,81
Présent	0,86	0,90	0,88	0,84	0,87	0,86
Assertions	0,82	0,82	0,82	0,80	0,80	0,80

**Tableau 2.** Précision (*P*), rappel (*R*) et *F*-mesure (*F*) obtenus sur les corpus d'entraînement et de test par catégorie d'assertion et globalement

pour *autre personne* et 0,86 pour *présent*. La catégorie *possible* montre une *F*-mesure de 0,49. Les valeurs de la précision et du rappel pour ces quatre catégories sont comparables. Les deux dernières catégories sont beaucoup plus difficiles à identifier : la *F*-mesure pour *hypothétique* est de 0,35 et elle atteint seulement 0,12 pour *conditionnel*. Ces deux catégories montrent aussi des différences entre la précision et le rappel : 0,63 et 0,24 pour *hypothétique*, et 0,08 et 0,21 pour *conditionnel* respectivement.

Les résultats de l'identification des catégories *présent* et *absent* s'avèrent ainsi être les meilleurs. Ceci n'est pas surprenant car il s'agit des catégories les plus fréquentes : plus de 8 000 annotations *présent* et plus de 2 500 annotations *absent* dans le corpus d'entraînement. En ce qui concerne les erreurs autour de ces deux catégories, nous constatons qu'elles sont principalement dues à la mauvaise identification de la portée de la négation. Les erreurs observées sont de deux types principalement :

1) le système ne prend pas en charge toutes les structures syntaxiques. Ainsi pour l'exemple (21-a), selon l'annotation de référence des problèmes médicaux, (1) le terme *symptoms of Flolan overdose* est séparé en deux problèmes médicaux : *symptoms* et *Flolan overdose*, (2) ces deux termes ont des assertions différentes : *symptoms* est *absent*, tandis que *Flolan overdose* est *présent* chez le patient. Notre système déduit que chacun de ces problèmes est *absent* chez le patient, ce qui est erroné.

(21) a. *Without precipitating symptoms of Flolan overdose.*

Ce point peut être amélioré en utilisant une analyse syntaxique (Vincze *et al.*, 2008 ; Saurí, 2008). Celle-ci devrait permettre d'identifier la plus grande unité syntaxique située à droite du marqueur lorsqu'il s'agit d'un adjectif, d'un verbe ou d'un adverbe. Ainsi, un adjectif attribut aura une portée limitée au groupe nominal qui suit, alors que dans le cas d'un adjectif épithète, l'assertion portera sur toute la phrase ;

2) un marqueur d'assertion se situe parfois dans un contexte éloigné du terme sur lequel il porte : il s'agit des phrases contenant des énumérations juxtaposées ou coordonnées. Ainsi, dans l'exemple (22-a), le système a identifié *dizziness* comme

*présent*, alors que ce terme fait partie de l'énumération. Il est donc annoté comme *absent* dans les données de référence. Actuellement, l'identification correcte de la portée des assertions est une difficulté que nous n'avons pas complètement résolue.

- (22) a. *She denied any associated nausea, vomiting, diaphoresis, lightheadedness, or dizziness.*

Les trois catégories relatives au degré de certitude (*possible*, *hypothétique* et *conditionnel*) sont peu fréquentes dans les corpus. Comme nous l'avons montré, elles présentent une grande variation de situations et de marqueurs. Elles sont de plus partiellement ambiguës entre elles, ce qui les rend plus difficiles à identifier correctement. Comme le suggère le faible rappel de *possible*, la présence de mots exprimant la preuve médicale ou l'expression de suspicion dans le contexte du terme ne suffit pas à identifier la possibilité. Quant à la précision, elle traduit notamment les problèmes dus à l'ambiguïté des modaux (section 6.2). La catégorie *hypothétique* montre le rappel le plus faible, alors que la précision est la plus élevée au sein des degrés de certitude. Le faible rappel est essentiellement dû au faible nombre d'exemples disponibles pour établir les ressources pour cette catégorie. Ces performances reflètent également la difficulté à déterminer les indices d'hypothèses dans certains contextes. Ainsi, selon les données de référence, dans l'exemple (23-a), le problème médical *staphylococcus coag-negative* est *hypothétique*, mais l'indice est difficile à situer. Une connaissance médicale est alors nécessaire pour le traiter : il peut s'agir du fait que si la bactérie *staphylococcus coag-negative* est sensible aux deux médicaments, *gentamicin* et *clindamycin*, elle a été probablement traitée au moyen de ces médicaments et éliminée. Par ailleurs, une analyse supplémentaire des résultats pour la catégorie *hypothétique*, nous a permis d'observer des erreurs de portée, comme pour la catégorie *absent*.

- (23) a. *Further identification revealed staphylococcus coag-negative, which was sensitive to gentamicin and clindamycin.*

Les résultats pour l'identification de la *condition* sont les moins bons. Comme nous l'avons vu précédemment (section 5), la catégorie *conditionnel* s'exprime principalement à l'aide de propositions subordonnées. Pour identifier ces subordonnées, les seuls indices récurrents sont des prépositions et des conjonctions. Nous ne les avons pas utilisées car elles génèrent beaucoup trop de bruit. Nous n'avons pas pu non plus générer des indices plus extensifs sous forme de groupes prépositionnels : ce calcul serait trop coûteux et trop contraint. En revanche, nous envisageons dans l'avenir l'utilisation des fonctions syntaxiques pour une meilleure identification de la *condition*. Par contre, comme ce type de subordonnées peut être aussi bien un indice de *condition* qu'un indice d'*hypothèse*, nous nous attendons à l'apparition d'autres ambiguïtés et pensons donc, en plus, combiner des fonctions syntaxiques avec des marqueurs.

Par ailleurs, des 385 marqueurs seulement 8 sont ambigus. L'ambiguïté porte le plus souvent sur les catégories *hypothétique* et *possible*. Et il s'agit le plus souvent des verbes modaux (comme *would*, *should*, *may* ou *could*) qui peuvent apparaître dans les

deux contextes. Mais là encore, la priorité des marqueurs imposée par la compétition résout les ambiguïtés : comme la catégorie *hypothétique* est prioritaire sur la catégorie *possible*, c'est la catégorie *hypothétique* qui sera retenue pour les problèmes médicaux en question (comme dans les exemples en (10-a) ou en (16-d) à (16-f)). Ceci veut aussi dire que la catégorie *possible* peut rester déficitaire pour certaines occurrences de problèmes médicaux.

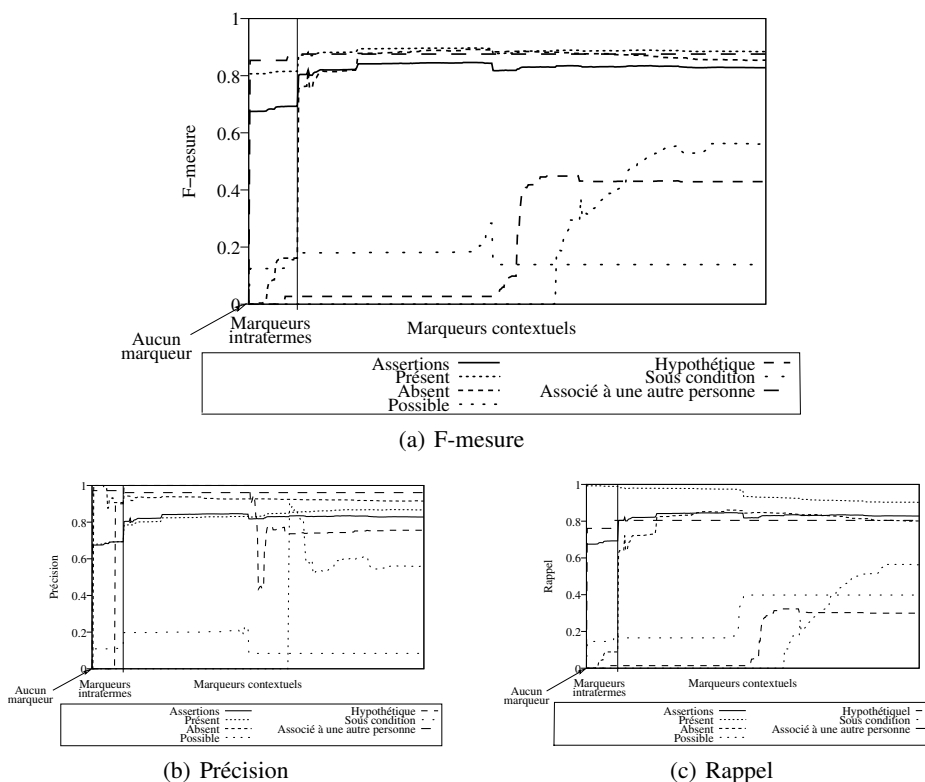
Nous avons également testé l'influence de deux prétraitements :

- identification des sections testée avec et sans sa prise en compte ;
- étiquetage morphosyntaxique testé en remplaçant Genia Tagger par TreeTagger.

Si ces pré-traitements n'influent que faiblement la F-mesure globale, leur absence ou leur qualité peut dégrader la détection de certains types d'assertion. Ainsi, l'absence de l'identification des sections dégrade les résultats dans deux situations : (1) lorsqu'il s'agit d'un problème médical *associé à une autre personne* : nous perdons alors 0,47 sur le corpus d'entraînement et 0,30 sur le corpus de test ; (2) lorsque les problèmes ont le statut *conditionnel* : nous perdons alors respectivement 0,05 et 0,18 sur les corpus d'entraînement et de test. Quant à l'étiqueteur morphosyntaxique, sa qualité montre une influence relativement faible sur la F-mesure lors de la détection des trois catégories d'assertion (les chiffres sont indiqués pour les corpus d'entraînement et de test, respectivement) : la catégorie *possible* perd 0,025 et 0,013 ; la catégorie *conditionnel* perd 0,024 et 0,03 ; et la catégorie *hypothétique* perd 0,005 et 0,034. Ces deux étiqueteurs diffèrent principalement par le type de corpus sur lequel a été défini leur modèle d'apprentissage : TreeTagger a été entraîné sur des corpus de la langue générale, tandis que Genia Tagger a été spécifiquement adapté aux corpus de textes médicaux. Les données sur lesquelles nous travaillons comportent en effet deux spécificités qui rendent la tâche difficile pour TreeTagger : (1) les termes biomédicaux utilisés présentent des spécificités lexicales, et un étiquetage morphosyntaxique correct nécessite des outils adaptés (Codon *et al.*, 2005) ; (2) les textes comportent un nombre important de termes *non traditionnels* (mal formés) ou avec des particularités syntaxiques qui demandent également des outils adaptés.

## 6.2. Influence des marqueurs et évolution des résultats

Nous avons observé l'évolution des performances de notre système sur le corpus d'entraînement, en fonction des marqueurs utilisés. L'analyse de ces résultats permet d'étudier l'influence de chaque marqueur et de distinguer ceux qui sont les plus intéressants, de ceux qui détériorent l'identification des assertions. La figure 3 présente cette évolution en termes de F-mesure, de précision et de rappel. Celle-ci est identique sur le corpus de test. Nous allons plus particulièrement analyser la figure 3(a), qui montre l'évolution de la F-mesure. Notons toutefois qu'il existe une grande similitude dans l'évolution de la F-mesure (fig. 3(a)) et du rappel (fig. 3(c)). La précision montre souvent des différences par rapport à ces deux mesures (fig. 3(b)).



**Figure 3.** Évolution des résultats (*F-mesure*, *précision* et *rappel*) dans le corpus d'entraînement

Dans notre travail sur l'étude incrémentale des marqueurs, nous les avons d'abord regroupés par type (structurels, intratermes et contextuels), puis par catégorie d'assertion, sans pour autant les ordonner au sein de chaque catégorie. Toutefois, nous observons que leur ordre peut avoir une influence, mais uniquement si des marqueurs de catégories d'assertion différentes sont concurrents, comme dans l'exemple (3-a).

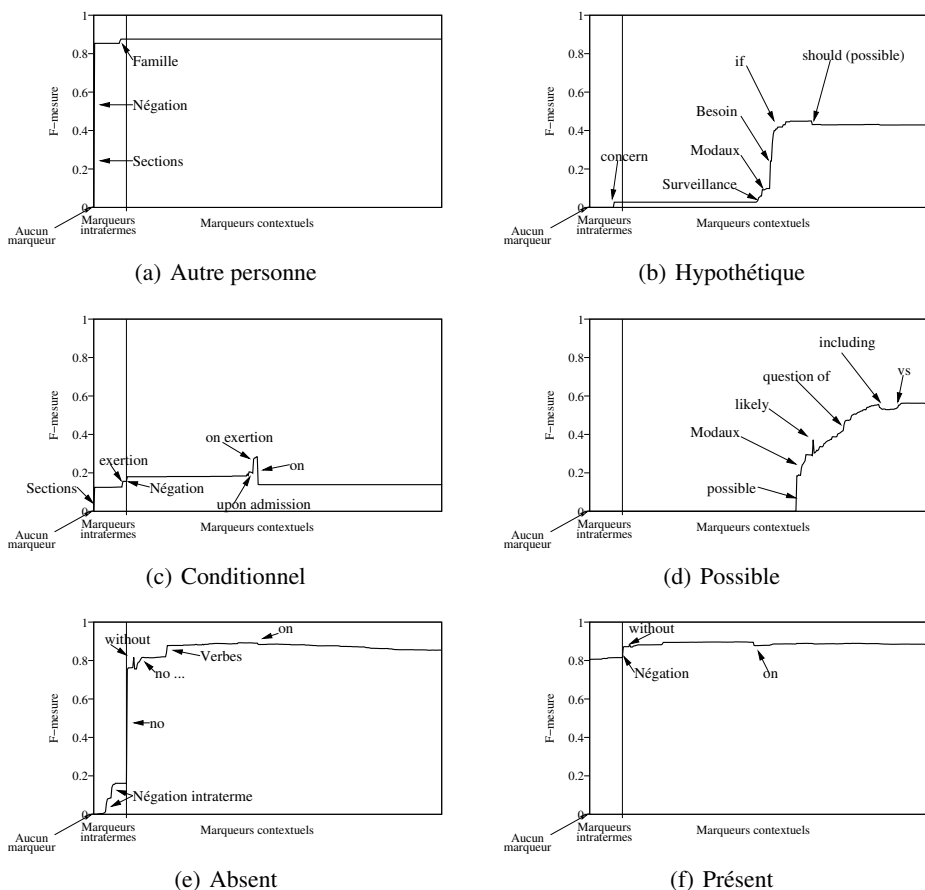
Les figures 3 et 4 montrent l'évolution de l'identification des assertions de manière générale (*Assertions*) et par catégorie. Les types de marqueurs exploités sont distingués sur ces figures : lors de cette étude incrémentale, ce sont d'abord les marqueurs structurels qui sont exploités, puis les marqueurs intratermes et enfin les marqueurs contextuels. Le nombre total de marqueurs est 385. Le point zéro correspond aux résultats minimaux : aucun marqueur et aucune information ne sont utilisés et toutes les assertions associées aux problèmes médicaux ont le statut *présent*. Dans ce cas, la *F-mesure* du système est de 0,672 sur le corpus d'entraînement. L'exploitation des

marqueurs devra donc améliorer ces performances. Sur ces trois figures nous pouvons aussi observer les tendances globales du système, comme l'activation de l'identification de différentes catégories. Ainsi, la catégorie *possible* reste assez longtemps à zéro et n'est prise en charge que par la deuxième moitié des marqueurs contextuels. Une observation similaire est vraie pour la catégorie *hypothétique*, hormis le fait qu'une influence au moment de la projection des marqueurs intratermes est aussi observable. Ceci veut dire que certains marqueurs sont spécifiques aux catégories. En revanche, d'autres marqueurs vont influencer l'identification de l'ensemble des catégories : ce fait est observable avec la prise en compte des marqueurs contextuels. Une autre tendance générale concerne le fait que certaines courbes sont saccadées car peu de marqueurs influencent les catégories correspondantes, alors que d'autres courbes montrent une évolution systématique (croissante ou décroissante). Ce dernier fait témoigne surtout de l'imbrication complexe et de l'impact de ces différentes catégories d'assertion et des marqueurs.

Pour décrire en détail l'évolution de chaque catégorie et l'influence des marqueurs, nous analysons les graphiques de la figure 4. Ces graphiques correspondent à ceux de la figure 3(a) et présentent donc l'évolution de la F-mesure. Seule la catégorie abordée est présentée sur ces graphiques individuels. Nous présentons les catégories dans l'ordre de leur priorité définie dans le cadre de la compétition I2B2 (section 3).

Le graphique 4(a) montre l'évolution de l'identification de la catégorie *autre personne*. Nous pouvons voir que peu de marqueurs influencent cette catégorie. Trois accroissements sont observables. Ils apparaissent avec la prise en compte des sections (augmentation de la F-mesure de 0 à 0,7839), de la négation (F-mesure 0,7868) et des relations de parenté (F-mesure 0,7883). Il existe effectivement des sections spécifiques (*FAMILY HISTORY* le plus souvent) qui ne contiennent que des informations liées à d'autres personnes que le patient. La prise en compte de ce marqueur de structure a donc un impact favorable sur la reconnaissance de cette catégorie. Il en est de même pour les relations de parenté (*twin, mother...*) qui permettent de proposer une association entre les problèmes médicaux et les membres de la famille du patient, et non avec le patient lui-même. Il s'agit de deux types de marqueurs spécifiques à la catégorie *autre personne*. En revanche, bien qu'ils ne soient pas spécifiques à cette catégorie, les marqueurs de négation influencent quand même de manière favorable son identification.

Pour la catégorie *hypothétique* (graphique 4(b)), nous pouvons distinguer six moments principaux : le marqueur *concern* (F-mesure 0,0247), les verbes de surveillance comme *watch for, call for* (F-mesure 0,0355), les modaux de l'*hypothétique* (F-mesure 0,1104), les marqueurs de besoin comme *as needed* ou *p.r.n.* (F-mesure 0,3241), le marqueur *if* et d'autres marqueurs de prévention (F-mesure 0,3805), et enfin après une zone de stabilité, une dégradation qui survient suite à la prise en compte des modaux du *possible* (F-mesure 0,3524). L'identification de cette catégorie évolue très peu par la suite et ne subit que peu l'influence des marqueurs de la catégorie *possible*. Au sein des marqueurs les plus saillants, nous avons d'une part, les marqueurs de besoin et, d'autre part, les marqueurs de surveillance des symptômes par le patient (le marqueur



**Figure 4.** Analyse de l'influence de différents marqueurs sur l'évolution de la F-mesure dans différentes catégories

intraterme *concern*, qui représente les formes comme *concerning* ou *concerned*, et les verbes de prévention comme *watch for* ou *call for*) parfois couplés avec *if*. Les marqueurs de besoin apparaissent souvent dans les prescriptions de médicaments, qui doivent être associés avec des symptômes précis (exemple (12)). Les modaux jouent aussi un rôle important mais sont ambigus avec la catégorie *possible* : le décroissement de la F-mesure de quelques points survient lorsque ces derniers commencent à être exploités. Nous pouvons d'ailleurs remarquer une amélioration parallèle, bien qu'instantanée, dans la reconnaissance du statut *possible*.

L'évolution de la catégorie *conditionnel* est présentée sur le graphique 4(c). Nous pouvons distinguer les moments cruciaux suivants : la prise en compte des sections (la F-mesure passe de 0 à 0,1684), le marqueur *exertion* au niveau intraterme (F-mesure



0,1931), la négation (F-mesure 0,2533), le marqueur *upon admission* qui cause une première dégradation (F-mesure 0,2511), le marqueur *on exertion* (F-mesure 0,3288) et finalement le marqueur *on* qui cause la dégradation définitive (F-mesure 0,1244). Ensuite, l'identification de cette catégorie n'évolue plus. L'acceptation de la *condition* dans le cadre de la compétition suit une logique médicale : il s'agit de problèmes médicaux subvenant sous certaines conditions, comme par exemple *on exertion (avec l'effort)*, *on palpation (avec la palpation)*, *status postexercice (après un exercice)*. C'est donc pour cette raison que les marqueurs qui traduisent ces situations sont très efficaces et améliorent la reconnaissance de cette catégorie d'assertion. En revanche, les conditions liées au temps ou au lieu – *upon admission (au moment de l'admission)*, *morning, in the past* – dégradent les résultats. Le marqueur *on*, qui a été utilisé pour généraliser la reconnaissance de ces situations dégrade beaucoup les résultats du fait de son ambiguïté. D'ailleurs, il dégrade aussi la reconnaissance de deux autres catégories (*présent* et *absent*). Comme pour la catégorie *autre personne*, la prise en compte des informations sur la structure des documents a un effet bénéfique : les sections sur les allergies entrent dans ce cas de figure. La négation a également un effet bénéfique sur la *condition*, sans doute parce qu'elle permet de désambiguïser certains contextes.

Sur le graphique 4(d), nous observons l'évolution de la catégorie *possible*. Six moments importants sont à distinguer : le marqueur *possible* et d'autres marqueurs de possibilité (F-mesure 0,1222 et 0,1994 respectivement), les modaux (F-mesure 0,2367), le marqueur *likely* lorsqu'il précède un problème médical (F-mesure 0,3154), le marqueur *question of* et d'autres marqueurs de suspicion (F-mesure 0,4674), le marqueur *including* qui détériore les résultats (F-mesure 0,4586), et finalement les marqueurs d'opposition de type *vs* (F-mesure 0,4880). La courbe évolue ensuite progressivement jusqu'à atteindre 0,4909 de F-mesure. Globalement, les marqueurs du *possible* sont assez spécifiques de cette catégorie, à l'exception des modaux : comme nous l'avons noté auparavant lorsque les modaux du *possible* interviennent, une dégradation des résultats est observée dans la catégorie *hypothétique*. Une dégradation est notée au sein de la catégorie *possible* avec *including*, qui apparaît fréquemment comme un marqueur d'énumération. Enfin, *likely* améliore les résultats lorsqu'il précède un problème médical. Mais sa présence dans le contexte droit d'un problème engendre des catégorisations erronées qui se traduisent par un pic dans la courbe.

Avec le graphique 4(e), nous passons à l'analyse de l'évolution de la polarité négative. Là aussi, six moments principaux sont à distinguer : la négation intraterme (F-mesure 0,1789), la négation contextuelle avec entre autres le marqueur *no* (F-mesure 0,7400), un pic autour du marqueur *without* (F-mesure 0,7954 qui décroît immédiatement à 0,7370), les locutions avec *no* (F-mesure 0,7923), les verbes de négation (F-mesure 0,8236) et une dégradation avec le marqueur conditionnel *on*. Ensuite la courbe de cette catégorie diminue progressivement de quelques points avec l'exploitation des marqueurs du *possible*. Parmi les faits marquants, notons l'importance des marqueurs de négation qui peuvent se positionner au niveau contextuel (*no, not, without*) et au niveau intraterme (*afebrile, nondistended, uncomplicated*). Quant au marqueur *without*, il produit des catégorisations correctes lorsqu'il apparaît avant un problème médical, mais génère des erreurs lorsqu'il apparaît après. Notons aussi l'im-

portance des verbes de négation comme *deny* et l'influence négative du marqueur conditionnel *on*.

La catégorie *présent* est illustrée avec le graphique 4(f). Cette catégorie par défaut ne reçoit pas de marqueurs spécifiques mais se trouve associée aux problèmes médicaux qui n'ont pas pu être caractérisés par d'autres catégories. Trois influences sont observées : la négation (F-mesure 0,8550), une augmentation ponctuelle autour du marqueur *without* (F-mesure 0,8657) et une dégradation avec le marqueur conditionnel *on* (F-mesure 0,8562). De plus, si l'on considère tous les problèmes médicaux comme *présents*, la F-mesure est de 0,672. La prise en compte et l'affinement des marqueurs dans d'autres catégories permettent donc de mieux identifier la catégorie *présent* mais surtout d'identifier les assertions propres à d'autres catégories.

### 6.3. Amélioration des résultats

L'observation de l'évolution des performances de notre système et l'implication de différents marqueurs dans cette évolution, présentées dans la section précédente, nous ont conduits à l'ajustement de nos ressources. Effectivement, c'est ainsi que nous avons pu détecter que certains marqueurs (comme *on*, *including* ou *upon admission*) détériorent systématiquement les résultats. Avec la nouvelle version des ressources, les performances obtenues sur le corpus d'entraînement passent de 0,82 à 0,86, et celles obtenues sur le corpus de test passent de 0,80 à 0,87. Globalement, nous améliorons la F-mesure sur le corpus d'entraînement de 0,04 points, et de 0,07 points sur le corpus de test. Le bénéfice apparaît être plus positif sur le corpus de test, ce qui est une constatation intéressante. Plus particulièrement la reconnaissance des catégories *conditionnel* et *hypothétique* a été sensiblement améliorée, tandis que la reconnaissance des catégories déjà correctement traitées (*autre personne*, *présent* et *absent*) n'est que légèrement améliorée. La seule catégorie qui reste insensible à cet ajustement de la ressource est la catégorie *possible*, sans doute parce qu'elle a la priorité la plus faible. D'autres stratégies seront nécessaires pour améliorer sa reconnaissance.

## 7. Conclusion et perspectives

Notre étude sur les assertions nous a amenés à nous intéresser à l'expression de la certitude et des degrés de certitude contenus dans les documents cliniques. Ce travail a été évalué sur les données fournies dans le cadre de la compétition internationale de TAL I2B2 2010, ce qui nous a conduits à nous concentrer sur cinq catégories d'assertion fréquentes dans les documents cliniques : la certitude positive, la certitude négative, ainsi que trois degrés de certitude décrivant la condition, l'hypothèse et la possibilité. De plus une sixième catégorie visée concerne les cas où le problème médical n'est pas associé au patient mais à une autre personne (quelqu'un de sa famille, le personnel médical...). L'identification automatique de ces niveaux de certitude est associée à un système d'extraction d'information, où la relation  $\{\textit{patient}, \textit{problème médical}\}$  est ciblée, les problèmes médicaux étant fournis par la tâche 1 de

la compétition. Grâce à l'observation en corpus de cette relation, nous avons constaté que l'assertion peut être identifiée avec trois niveaux d'indices : au sein du terme, dans le contexte du terme ou bien au travers de la structure du document. Face au peu de ressources existantes et à leur très faible disponibilité, nous avons constitué nos propres ressources pour chaque type d'assertion visé.

Nous avons implémenté une méthode fondée sur les ressources. L'évaluation des résultats générés par notre méthode montre des performances globales de 0,80 en termes de F-mesure. Les catégories les mieux reconnues sont la *certitude positive*, la *certitude négative* et *associé à une autre personne* : la F-mesure pour ces trois catégories varie entre 0,79 et 0,86 points. Les trois autres catégories (*conditionnel*, *hypothétique* et *possible*) restent difficiles à identifier : 0,49 pour la *possibilité*, 0,35 pour l'*hypothèse* et seulement 0,12 pour la *condition*. Grâce aux expériences décrites dans cet article et après l'amélioration des ressources, les performances de notre système passent à 0,86 sur le corpus d'entraînement et à 0,87 sur le corpus de test, en montrant une amélioration de 0,04 et 0,07 points respectivement. Notre système est loin d'être le plus performant de la campagne I2B2 2010. Cependant, avec l'ajustement des ressources, il ne montre que 0,06 points de moins que les meilleurs systèmes s'appuyant généralement sur des méthodes d'apprentissage ou des approches mixtes et se trouve au-dessus de la moyenne de la campagne. Par ailleurs, quantitativement nos résultats sont assez comparables avec d'autres travaux de l'état de l'art réalisés en dehors de cette campagne d'évaluation, mais les aspects qualitatifs (catégories d'assertion, type et taille des données) varient.

Les erreurs les plus fréquentes de notre système sont dues au mauvais calcul de la portée et à la non-prise en compte de certaines structures syntaxiques. Nous pensons que le recours à un analyseur syntaxique, comme le Stanford Parser (Klein et Manning, 2003), aidera à identifier les fonctions syntaxiques associées aux prépositions et aux conjonctions caractéristiques des catégories d'assertion *conditionnel* et *hypothétique*. De même, le rattachement au problème médical pourra être calculé à l'aide des dépendances produites par l'analyseur. Par ailleurs, afin d'améliorer le rappel et la précision, les marqueurs et les catégories sémantiques seront à enrichir à l'aide de synonymes.

Trois des catégories des assertions traitées ici (certitude positive, certitude négative et *possible*) sont répandues dans la langue en général. À notre avis, les ressources utilisées pour les identifier pourront être exploitées dans d'autres contextes et sur d'autres types de documents. En revanche, les trois autres catégories (*conditionnel*, *hypothétique* et *associé avec une autre personne*) suivent une logique médicale et clinique précise. Il est possible que ces catégories ne soient pas portables sur d'autres contextes. Quant aux ressources proprement dites, elles ont été constituées dans ce contexte de compétition et sur des documents assez spécifiques. Nous pensons que leur réutilisation pourrait nécessiter une adaptation supplémentaire. En ce qui concerne le système, moyennant l'ajustement des ressources, il est portable sur d'autres types de textes, qu'ils soient spécialisés ou non, comme par exemple l'étude du risque chimique ou des sujets de controverse de manière générale. Sur de nouveaux domaines, l'intérêt

d'un système à base de règles est d'autant plus grand que des données annotées n'y existent pas.

Cette expérience a montré également que s'il est possible de faire une description linguistique précise des degrés de certitude et d'en constituer des ressources, leur traitement et leur reconnaissance automatiques avec des outils de TAL peuvent néanmoins être problématiques, surtout lorsque certaines des catégories ciblées sont très proches l'une de l'autre. Dans notre expérience, c'est le cas des catégories *possible* et *hypothétique*. Dans ce sens, cette constatation semble corroborer avec les observations précédentes (Thompson *et al.*, 2008), mais nous voudrions plutôt mettre l'accent sur la capacité des systèmes de TAL à détecter ces catégories que sur les moyens dont dispose une langue pour les exprimer.

Dans ce travail, une attention spéciale a été consacrée à l'étude de l'évolution des performances du système grâce à une exploitation incrémentale des ressources. Nous avons ainsi commencé par catégoriser les problèmes médicaux dans la catégorie par défaut (présent) pour ensuite ajouter un à un les marqueurs de nos ressources. Ce qui ressort de cette évaluation est que certains marqueurs ont une portée locale sur une seule catégorie, mais d'autres, étant polycatégoriels ou ambigus, influencent l'identification de plusieurs catégories d'assertion. C'est par exemple le cas de la structure des documents, de la négation, des modaux et du marqueur *on*. L'influence de ce dernier est, au final, plutôt négative et conduit la plupart du temps à des erreurs de catégorisation. Cette étude incrémentale nous a permis d'ajuster nos ressources, et, par là même, d'améliorer nos résultats. La poursuite de cet aspect de notre travail nous conduit à la définition d'une méthode générique d'évaluation et de sélection des marqueurs et des ressources. Enfin, nous envisageons d'ajuster les poids des types de marqueurs à l'aide d'un système d'apprentissage.

## 8. Bibliographie

- Bairoch A., Boeckmann B., Ferro S., Gasteiger F., « Swiss-Prot : Juggling between evolution and stability », *Briefings in Bioinformatics*, vol. 5, n° 1, p. 39-55, 2004.
- Battistelli D., Amardeilh F., « Knowledge Claims in Scientific Literature, Uncertainty and Semantic Annotation : A Case Study in the Biological Domain », *SAKM-2009*, Los Angeles, California, 2009.
- Bethard S., Yu H., Thornton A., Hatzivassiloglou V., Jurafsky D., « Extracting Opinion Propositions and Opinion Holders using Syntactic and Lexical Cues », *Computing Attitude and Affect in Text : Theory and Applications*, vol. 20, Springer Netherlands, p. 125-141, 2006.
- Bouscaren J., *Linguistique anglaise - Initiation à une grammaire de l'énonciation*, Ophrys, Paris, 1991.
- Chapman W., Bridewell W., Hanbury P., Cooper G., Buchanan B., « Evaluation of negation phrases in narrative clinical reports », *AMIA 2001*, Washington, DC, p. 105-109, 2001.
- Coden A., Pakhomov S., Ando R., Duffy P., Chute C., « Domain-specific language models and lexicons for tagging », *Journal of Biomedical Informatics*, vol. 38, n° 6, p. 422-30, 2005.

- Culioli A., *Notes du séminaire de DEA, 1983-84*, Université Paris VII, Département de Recherches Linguistiques, Paris, 1984.
- Grabar N., Hamon T., « Exploitation of speculation markers to identify the structure of biomedical scientific writing », *AMIA 2009*, San Francisco, p. 203-207, 2009.
- Hamon T., Nazarenko A., « Le développement d'une plate-forme pour l'annotation spécialisée de documents web : retour d'expérience », *TAL*, vol. 49, n° 2, p. 127-154, 2008.
- Holmes J., « Doubt and certainty in ESL textbooks », *Applied Linguistics*, vol. 13, n° 2, p. 21-44, 1988.
- Hoye L., *Adverbs and modality in English*, Longman, London, New York, 1997.
- Hyland K., *Hedging in scientific research articles*, John Benjamins B.V., Amsterdam, 1998.
- Jilani I., Grabar N., Meneton P., Jaulent M.-C., « Classification of Biomedical Knowledge According to Confidence Criteria », *Stud Health Technol Inform*, IOS Press, Göteborg, Sweden, p. 199-204, 2008.
- Kilicoglou H., Bergler S., « Syntactic dependency based heuristics for biological event extraction », *BIONLP-2009*, Boulder, Colorado, p. 119-127, 2009.
- Klein D., Manning C. D., « Accurate Unlexicalized Parsing », *ACL-2003*, Sapporo, Japan, p. 423-430, 2003.
- Lakoff G., *Linguistics and Natural Logic*, G. Harman and D. Davidson, Reidel, Dordrecht, p. 545-665, 1972.
- Larreja P., Rivière C., *Grammaire explicative de l'anglais*, Longman, Essex, 1999.
- Light M., Qiu X. Y., Srinivasan P., « The Language of Bioscience : Facts, Speculations and Statements in Between », *ACL Workshop on Linking biological literature, ontologies and databases*, Boston, Massachusetts, p. 17-24, 2004.
- Lyons J., *Semantics 2*, Cambridge University Press, Cambridge, 1977.
- Marco C. D., Mercer R., « Hedging in scientific articles as a means of classifying citations », *AAAI-2004*, Stanford University, California, p. 50-54, 2004.
- Medlock B., Briscoe T., « Weakly supervised learning for hedge classification in scientific literature », *ACL-2007*, Prague, Czech Republic, p. 992-999, 2007.
- Mizuta Y., Korhonen A., Mullen T., Collier N., « Zone analysis in biology articles as a basis for information extraction », *International Journal of Medical Informatics*, vol. 75, n° 6, p. 468-487, 2006.
- Mutalik P. G., Deshpande A., Nadkarni P. M., « Use of general-purpose negation detection to augment concept indexing of medical documents : a quantitative study using the UMLS », *Journal of the American Medical Informatics Association*, vol. 8, n° 6, p. 598-609, 2001.
- Palmer F., *Modality and the English Modals*, Longman, London, 1990.
- Riegel M., Pellat J.-C., Rioul R., *Grammaire méthodique du français*, PUF, 2004.
- Riloff E., Wiebe J., Wilson T., « Learning subjective nouns using extraction pattern bootstrapping », *CoNLL-03*, p. 25-32, 2003.
- Rubin V. L., « Stating with Certainty or Stating with Doubt : Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements », *NAACL-HLT 2007*, Rochester, NY, p. 141-144, 2007.

- Rubin V. L., Liddy E. D., Kando N., « Certainty Identification in Texts : Categorization Model and Manual Tagging Results », *Computing Attitude and Affect in Text : Theory and Applications*, vol. 20, Springer Netherlands, p. 61-76, 2006.
- Saurí R., Pustejovsky J., « Determining Modality and Factuality for Text Entailment », *ICSC 2007*, Washington, DC, p. 509-516, 2007.
- Saurí R., A Factuality Profiler for Eventualities in Text, Thèse de doctorat, Brandeis University, Waltham, Boston, 2008.
- Saurí R., Verhagen M., Pustejovsky J., « Annotating and recognizing event modality in text », *FLAIRS 2006*, Melbourne Beach, Florida, p. 333-339, 2006.
- Sibanda T., Uzuner O., « Role of local context in automatic de-identification of ungrammatical, fragmented text », *NAACL-HLT 2006*, New York, p. 65-73, 2006.
- Teufel S., « Meta-discourse markers and problem-structuring in scientific articles », *ACL Workshop on Discourse Structure and Discourse Markers*, Montreal, p. 43-49, 1998.
- Thompson P., Venturi G., McNaught J., Montemagni S., Ananiadou S., « Categorising Modality in Biomedical Texts », *LREC Workshop Building and Evaluating resources for biomedical text mining*, Marrakech, p. 27-34, 2008.
- Tsuruoka Y., Tateishi Y., Kim J.-D., Ohta T., McNaught J., Ananiadou S., Tsujii J., « Developing a Robust Part-of-Speech Tagger for Biomedical Text », *Advances in Informatics : (PCI-2005)*, vol. LNCS 3746, Volos, Grece, p. 382-392, 2005.
- Vincze V., Szarvas G., Farkas R., Móra G., Csirik J., « The BioScope corpus : biomedical texts annotated for uncertainty, negation and their scopes », *BMC Bioinformatics*, vol. 9(Suppl. 11), p. S9, 2008.
- Vion R., « Modalités, modalisations et activités langagières », *Marges Linguistiques*, vol. 2, p. 209-231, 2001.
- Vold E. T., Modalité épistémique et discours scientifique, Thèse de doctorat, Universitetet i Bergen, Bergen, Norvège, 2008.
- Wilbur W., Rzhetsky A., Shatkay H., « New Directions in Biomedical Text Annotations : Definitions, Guidelines and Corpus Construction », *BMC Bioinformatics*, vol. 7, p. 356-365, 2006.

**A. Traduction des exemples**

- (1) a. *The patient denies suffering from abdominal pain.*  
(Le patient ne présente pas de douleur abdominale.)
- (2) a. *DLG4 : Cortical cytoskeleton protein found in a complex involved in maintaining cell polarity and epithelial integrity. Involved in the regulation of mitotic spindle orientation, proliferation, differentiation and tissue organization of neuroepithelial cells.*  
(DLG4 : Protéine du cytosquelette cortical appartenant à un complexe impliqué dans le maintien de la polarité des cellules et l'intégrité épithéliale. Impliquée également dans la régulation de l'orientation du fuseau mitotique, la prolifération, la différenciation et l'organisation tissulaire des cellules neuroépithéliales.)
- b. *DIP1 : May negatively regulate cell cycle progression. May act at least in part via inhibition of the cyclin-D1/CDK4 complex, thereby preventing phosphorylation of RB1 and blocking E2F-dependent transcription.*  
(DIP1 : Peut réguler négativement le cycle cellulaire. Peut agir au moins en partie via l'inhibition du complexe cycline D1/CDK4, prévenant ainsi la phosphorylation du gène RB1 et bloquant la transcription dépendante de la protéine E2F.)
- c. *HE6 : Could be involved in a signal transduction pathway controlling epididymal function and male fertility.*  
(HE6 : Pourrait être impliquée dans une voie de transduction du signal qui contrôle la fonction épидидymaire et la fertilité masculine.)
- (3) a. *The patient remained afebrile, and it was felt that it was probably not a contamination.*  
(Le patient est resté apyrétique. On en a déduit qu'il ne s'agissait probablement pas d'une contamination.)
- (4) a. *The patient had a stroke.*  
(Le patient a eu une attaque.)
- (5) a. *No aortic regurgitation present.*  
(Pas de régurgitation aortique présente.)
- b. *Review of systems shows that the patient denies any orthopnea.*  
(L'examen clinique montre que le patient ne présente aucune orthopnée.)
- c. *The patient has continued to feel well without any symptoms of shortness of breath.*  
(Le patient continue à se sentir bien, sans aucun symptôme d'essoufflement.)
- d. *Lower extremity dopplers negative for deep venous thrombosis.*  
(Échographie doppler des membres inférieurs négative pour la thrombose veineuse profonde.)
- e. *ABDOMEN : soft, nontender, nondistended.*  
(ABDOMEN : souple, dépressible et non distendu.)

- f. *She is afebrile with stable vital signs.*  
(Elle est apyrétique, avec une hémodynamique stable (avec signes vitaux stables).)
- (6) a. *No history of abnormal pap smears.*  
(Pas d'antécédents de frottis cervico-vaginal anormal.)  
b. *There were no findings suggestive of an accompanying hollow viscous perforation.*  
(Il n'y avait pas d'argument d'un organe perforé.)
- (7) a. *She [was] not [felt to be] acutely suicidal.*  
(Elle n'était pas à risque suicidaire.)  
b. *The baby did not [turn] blue at that time.*  
(Le bébé n'était pas cyanosé à ce moment-là.)  
c. *She did not have any further episodes of blurry vision or chest pain.*  
(Elle n'a pas eu d'autres épisodes de vision trouble ou de douleur à la poitrine.)
- (8) a. *Bradycardia is resolved after beta blockers and calcium channel blockers were stopped and norvasc was started.*  
(La bradycardie s'est résolue après l'arrêt des bêta-bloquants et des inhibiteurs calciques, et le norvasc a été introduit.)  
b. *By the following morning all mental status changes had cleared.*  
(Le lendemain matin, les troubles de la conscience avaient disparu.)  
c. *Lower extremity edema improved with diuresis.*  
(Les œdèmes des membres inférieurs ont diminué avec la diurèse.)
- (9) a. *He was to contact our office if there is any drainage.*  
(Il devait contacter notre service en cas d'écoulement.)  
b. *If any abdominal pain, fever or dizziness, go to the emergency room.*  
(En cas de douleur abdominale, de fièvre ou de vertiges, consultez au service des urgences.)  
c. *Return immediately for worsening signs of infection.*  
(Revenez immédiatement en cas d'aggravation des symptômes.)  
d. *He was advised to return to the emergency department, or call his physician for any new symptoms.*  
(On lui a conseillé de retourner au service des urgences, ou d'appeler son médecin en cas de nouveaux symptômes.)
- (10) a. *If his heart should stop or he should have a lethal arrhythmia, we should not do any chest compressions.*  
(Si son cœur devait s'arrêter ou s'il devait avoir une arythmie mortelle, nous ne devrions pas effectuer un massage cardiaque.)
- (11) a. *Watch for BZD withdrawal, agitation.*  
(Surveiller le syndrome de sevrage aux BZD et l'agitation.)



- b. *Wound care consult given anasarca to prevent decub.*  
(Consultation pour le soin des plaies étant donné anasarca pour prévenir un decub.)
- c. *Monitor wounds for infection - redness, drainage, or increased pain.*  
(Surveillance d'une surinfection de la plaie (rougeur, écoulement ou augmentation de la douleur.)
- (12) a. *Ativan 0.5 mg PO q4 hours prn for anxiety.*  
(Ativan 0,5 mg per os toutes les 4 heures si besoin pour l'anxiété.)
- b. *She was given Duoneb as needed for any shortness of breath.*  
(On lui a prescrit du Duoneb en cas de gêne respiratoire.)
- c. *Docusate Sodium 100 mg Capsule Sig : One (1) Capsule PO BID (2 times a day) : take as necessary for constipation.*  
(Docusate Sodium 100 mg Capsule Sig per os : une (1) capsule deux fois par jour en cas de constipation.)
- (13) a. *Patient reports shortness of breath upon climbing stairs.*  
(Le patient se plaint de gêne respiratoire quand il monte les escaliers.)
- b. *He was also short of breath with ambulation.*  
(Il se plaint également de gêne respiratoire quand il marche.)
- c. *He can climb approximately three flights of stairs before becoming short of breath.*  
(Il peut monter environ trois étages avant d'être essoufflé.)
- d. *He reports severe dyspnea on exertion.*  
(Il se plaint de dyspnée sévère à l'effort.)
- e. *Penicillin causes a rash.*  
(La Pénicilline a entraîné une éruption cutanée.)
- f. *The patient is allergic to Ciprofloxacin, morphine sulfate and Droperidol.*  
(Le patient est allergique à la Ciprofloxacine, au sulfate de morphine et au Dropéridol.)
- g. *Allergies included penicillin and iodine.*  
(Les allergies comprenaient la pénicilline et l'iode.)
- h. *ALLERGIES : TOMATO (causing anaphylaxis) and TETANUS SHOT.*  
(ALLERGIES : TOMATE (provoquant un choc anaphylactique) et VACCIN ANTI-TETANIQUE.)
- (14) a. *The patient gets very somnolent with Ativan.*  
(Le patient est très somnolent sous Ativan.)
- b. *She also has some shortness of breath and throat constriction with shrimp.*  
(Elle présente également une gêne respiratoire et une constriction de la gorge après l'ingestion de crevettes.)

- (15) a. *This is very likely to be an asthma exacerbation.*  
(Il est fort probable qu'il s'agisse d'une exacerbation d'asthme.)
- b. *Questionable pneumonia.*  
(Pneumonie possible.)
- (16) a. *A lucency and irregularity of the glenoid, possibly representing a fracture, was also seen.*  
(Une transparence et une irrégularité de la cavité glénoïde, possiblement représentant une fracture, ont aussi été observées.)
- b. *Florinef 0.1 mg po q day for presumed adrenal insufficiency.*  
(Florinef 0,1 mg per os une fois par jour pour insuffisance surrénalienne présumée.)
- c. *He was found on physical exam to have an asymmetric prostate with prominence on the left and a question of a nodule.*  
(L'examen clinique a montré une prostate asymétrique, plus importante à gauche et un probable nodule.)
- d. *He could have a baseline high and increased creatinine.*  
(Il pourrait avoir un taux de base haut et une créatinine élevée.)
- e. *Slight irregularity of the origin of the right vertebral artery may be atherosclerotic in nature.*  
(L'irrégularité au niveau de l'origine de l'artère vertébrale droite peut être de l'athérosclérose.)
- f. *The patient was already on an aspirin and plavix should this represent a small stroke.*  
(Le patient était déjà sous aspirine et sous plavix si en prévention d'un infarctus du myocarde.)
- (17) a. *We suspected aspiration or a chemical pneumonitis.*  
(Nous supposons une pneumopathie d'inhalation ou médicamenteuse.)
- b. *Still suspicion that she had a stroke while off her coumadin.*  
(Nous supposons encore qu'elle ait eu une attaque alors qu'elle n'était pas sous coumadine.)
- c. *The acute onset of chest pain was initially very concerning for a PE.*  
(Le début aigu de la douleur thoracique a initialement fait craindre une embolie pulmonaire.)
- d. *The patient was started on Florinef for concern of adrenal insufficiency.*  
(Le patient a commencé le Florinef pour l'insuffisance surrénalienne supposée.)
- (18) a. *We will make sure there is no associated B12 deficiency as well.*  
(Nous allons vérifier qu'il n'y a pas non plus de déficit en vitamine B12.)
- b. *We will monitor for signs of alcohol withdrawal.*  
(Nous allons surveiller les symptômes du sevrage alcoolique.)
- c. *She was admitted to rule out an acute cardiac event.*  
(Elle a été admise afin d'éliminer un épisode cardiaque aigu.)

- d. Please r/o acute process.  
(Veuillez éliminer un épisode aigu.)
- e. Troponins were checked for a chest pain.  
(Les troponines ont été vérifiées pour la douleur thoracique.)
- f. A magnetic resonance imaging study will be scheduled as an outpatient in three months to rule out a small vascular malformation if responsible for the hemorrhage.  
(Une imagerie par résonance magnétique sera réalisée dans trois mois en ambulatoire pour vérifier si une petite malformation vasculaire est responsable de l'hémorragie.)
- g. He underwent lower extremity study to rule out a DVT in both legs.  
(Il a eu un examen des membres inférieurs pour éliminer une thrombose veineuse profonde des deux membres.)
- h. He then underwent a CT scan to rule out a PE.  
(Il a ensuite eu un angioscanner afin d'éliminer une embolie pulmonaire.)
- (19) a. A repeat chest x-ray done for dyspnea showed persistent pleural effusion with atelectasis versus pneumonia.  
(Une nouvelle radiographie du thorax, réalisée pour la dyspnée, a montré un épanchement pleural persistant avec une atelectasie versus pneumonie.)
- b. Benign positional vertigo vs labyrinthitis.  
(Vertige paroxystique positionnel bénin vs labyrinthite.)
- c. Discussion was had at this time with the patient regarding whether or not this represented an infected abscess vs. post-surgical change.  
(Le patient a alors été informé sur l'éventualité d'un abcès infecté ou d'une complication postopératoire.)
- d. DM ? Arthritis.  
(Diabète ? Arthrite.)
- e. Bleed ?  
(Saignement ?)
- (20) a. The patient's father died of a myocardial infarction at age 40.  
(Le père du patient est décédé d'un infarctus du myocarde à l'âge de 40 ans.)
- b. Family history : not significant for coronary artery disease.  
(Antécédents familiaux : pas d'antécédents cardiovasculaires significatifs.)
- c. Avoid individuals who may be ill.  
(Éviter le contact avec des individus qui peuvent être malades.)
- d. The family and other caregivers should be considered for immunization against influenza.  
(La famille et les autres soignants devraient être immunisés contre la grippe.)
- (21) a. Without precipitating symptoms of Flolan overdose.  
(Sans provoquer les symptômes de surdosage du Flolan.)

- (22) a. *She denied any associated nausea, vomiting, diaphoresis, lightheadedness, or dizziness.*  
(*Elle n'a présenté ni nausées, ni vomissements ni sueurs, ni étourdissements, ni vertiges associés.*)
- (23) a. *Further identification revealed staphylococcus coag-negative, which was sensitive to gentamicin and clindamycin.*  
(*D'autres identifications ont révélé la présence de staphylocoques à coagulase négative, sensibles à la gentamicine et à la clindamycine.*)