

Portée de la négation : détection par apprentissage supervisé en français et portugais brésilien

Clément Dalloux¹ Natalia Grabar² Vincent Claveau¹ Claudia Moro³

(1) Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes

(2) UMR 8163 STL CNRS, Université de Lille 3 - France,

(3) Pontifícia Universidade Católica do Paraná (PUC-PR),

(1) prenom.nom@irisa.fr, (2) natalia.grabar@univ-lille3.fr,

(3) c.moro@pucpr.br

RÉSUMÉ

La détection automatique de la négation fait souvent partie des pré-requis dans les systèmes d'extraction d'information, notamment dans le domaine biomédical. Cet article présente nos contributions concernant la détection de la portée de la négation en français et portugais brésilien. Nous présentons d'une part deux corpus principalement constitués d'extraits de protocoles d'essais cliniques en français et portugais brésilien, dédiés aux critères d'inclusion de patients. Les marqueurs de négation et leurs portées y ont été annotés manuellement. Nous présentons d'autre part une approche par réseau de neurones récurrents pour extraire les portées.

ABSTRACT

Negation scope : sequence labeling by supervised learning in French and Brazilian-Portuguese.

Automatic detection of negated content is often a pre-requisite in information extraction systems, especially in the biomedical domain. This paper proposes our contributions on negation scope detection in French and Brazilian Portuguese. We introduce two corpora mainly built with excerpts from clinical trial protocols, describing the inclusion criteria of patients. The corpora have been manually annotated for marking up the negation cues and their scope. Then, we propose a recurrent neural networks approach to acquire the scopes.

MOTS-CLÉS : négation, français, portugais brésilien, apprentissage supervisé, réseau de neurones.

KEYWORDS: negation, French, Brazilian Portuguese, supervised learning, neural network.

1 Introduction

La détection de la négation dans les textes est un des pré-requis qui est utile dans de nombreuses tâches d'extraction d'informations. Dans le domaine biomédical, plusieurs informations existent sous forme non structurée et il y est donc nécessaire de pouvoir différencier, par exemple, l'absence et la présence d'une maladie, ou encore la prise ou non d'un médicament (Chapman *et al.*, 2001; Vincze *et al.*, 2008). Dans le cas d'essais cliniques, ces informations sont déterminantes dans le processus de recrutement des patients car elles indiquent et définissent les critères d'inclusion et d'exclusion de patients pouvant être inclus dans un essai. Il peut s'agir par exemple du statut fumeur ou non d'une personne, de la prise ou non d'un médicament donné, du fait qu'une patiente soit enceinte, etc.

L'article est structuré de la manière suivante. Dans la section qui suit, nous présentons les travaux existants qui proposent des ensembles de données annotées ainsi que des méthodes par systèmes expert ou par apprentissage supervisé pour la détection automatique de la négation et de sa portée. Ces travaux concernent majoritairement le traitement de textes en anglais. Nous présentons ensuite, dans la section 3, les ensembles de données que nous avons annotées, en français et portugais brésilien, dans le but d'automatiser la détection de la négation dans d'autres langues que l'anglais. Notre méthode de détection de la portée par apprentissage supervisé est présentée dans la section 4. Dans la section 5, nous présentons les résultats de notre méthode sur ces ensembles de données, ainsi qu'une analyse des erreurs les plus fréquentes. Enfin, nous présentons nos conclusions et perspectives dans la dernière section.

2 Travaux existants

Dans (Dalloux, 2017), nous avons présenté une grande partie des travaux pertinents pour cette tâche. Dans cette section, nous les décrivons brièvement et complétons notre état de l'art.

2.1 Les données

Ces dernières années, souvent dans le cadre de *shared tasks*, les énoncés/assertions négatifs ont été annotés dans plusieurs corpus de spécialité. On peut y distinguer deux catégories : (1) les corpus avec les marqueurs et leurs portées annotés et (2) les corpus se focalisant sur les concepts et entités nommées.

Dans la première catégorie, on retrouve Bioscope (Vincze *et al.*, 2008), qui est composé de comptes-rendus d'exams radiologiques, d'articles scientifiques ainsi que de résumés d'articles annotés avec l'incertitude et la négation. En tout, le corpus comprend 20 924 phrases dont environ 13 % sont négatives. Les ensembles de données de *SEM-2012¹ sont annotés de la même manière. Ce corpus est composé d'un roman et de trois nouvelles de Sherlock Holmes par Conan Doyle et comprend 5 520 phrases, dont 1 227 négatives. Les marqueurs de la négation sont constitués soit d'un préfixe soit d'un ou de plusieurs mots qui modifient, sauf exceptions, la polarité et donc le sens de la phrase. La portée est l'effet du marqueur, qui s'étend sur toute la phrase ou sur une partie de cette phrase. Dans cette catégorie de travaux, les marqueurs de la négation sont donc ciblés.

Les travaux de la seconde catégorie se focalisent sur le contexte entourant les entités nommées. Par exemple, la compétition i2b2/VA-2010 (Uzuner *et al.*, 2011) présentait trois tâches, dont la classification d'assertions se focalisant sur l'attribution des types d'assertions pour les concepts médicaux. En d'autres termes, à chaque concept était attribué l'une de ces 6 classes d'assertions : *present*, *absent*, *possible*, *conditional*, *hypothetical* ou *not associated with the patient*. Dans cette catégorie de travaux, ce sont les entités nommées qui sont les pivots des études.

Mipacq (Albright *et al.*, 2013) est un autre exemple de corpus, qui est constitué de données cliniques annotées avec plusieurs couches d'étiquettes syntaxiques et sémantiques. Dans ce corpus, chaque entité UMLS dispose de deux emplacements d'attributs : *Negation*, qui peut prendre deux valeurs : *true* ou *false*, et *status*, qui peut être *none*, *possible*, *HistoryOf* ou *FamilyHistoryOf*.

1. <https://www.clips.uantwerpen.be/sem2012-st-neg/>

2.2 Détection automatique

D'une manière générale, il existe deux familles d'approches pour aborder la détection automatique de la négation. La première famille couvre la conception de systèmes experts, tels que *NegEx* (Chapman *et al.*, 2001) et son adaptation française (Deléger & Grouin, 2012), *Negfinder* (Mutalik *et al.*, 2001), ou bien *ConText* (Harkema *et al.*, 2009) et son adaptation française (Abdaoui *et al.*, 2017). Notons que le système *ConText* a une portée plus large et permet de détecter la négation, la temporalité et le sujet dans les textes cliniques. Très récemment, Peng *et al.* (2018) ont proposé *NegBio* dont le code est disponible en ligne². Leur système repose sur des règles définies à partir des graphes de dépendances (*universal dependency graph* (UDG)).

La deuxième famille rassemble de nombreux travaux qui utilisent la classification par apprentissage supervisé à l'aide de méthodes telles que les champs aléatoires conditionnels (*Conditional Random Fields* ou CRFs), les machines à vecteurs support (SVM) ou les réseaux de neurones (Velldal *et al.*, 2012; Read *et al.*, 2012; Packard *et al.*, 2014; Fancellu *et al.*, 2016). Enfin, dans (Dalloux *et al.*, 2017), nous proposons déjà plusieurs approches par réseau de neurones et comparons leurs résultats pour le français et l'anglais.

Dans cet article, nous publions les résultats obtenus avec la version la plus récente de notre corpus, ainsi qu'avec un corpus en portugais brésilien, une langue rarement représentée dans les travaux de TAL.

3 Nos données

Nous présentons dans cette section deux corpus, l'un en français (FR), l'autre en portugais brésilien (PTBR), annotés manuellement pour marquer les informations liées à la négation : les marqueurs et leur portée. Ces corpus sont similaires car élaborés à partir de protocoles d'essais cliniques dans les deux langues. Le corpus français a été annoté par l'équipe de l'IRISA d'INRIA, tandis que le corpus brésilien a été annoté par trois étudiants de l'école de médecine de l'Université pontificale catholique du Paraná³.

Les protocoles en français sont issus du registre des essais cliniques de l'hôpital Gustave Roussy, ainsi que de l'Institut National du Cancer, auxquels nous avons ajouté quelques cas cliniques. Le corpus français est pré-traité avec *TreeTagger* (Schmid, 1994) pour effectuer l'étiquetage morpho-syntaxique et la lemmatisation. Les protocoles en portugais brésilien sont issus du site brésilien dédié aux essais cliniques⁴. L'étiquetage morpho-syntaxique (*Universal POS tags*) et les lemmes du corpus portugais brésilien sont obtenus à l'aide de *RDRPOSTagger* proposé par Nguyen *et al.* (2015).

Dans les exemples qui suivent, les marqueurs de négation sont en gras et leur portée est marquée entre crochets. Le rationale des annotations diffère légèrement dans les deux langues. Si l'annotation des marqueurs est identique, il existe des différences dans l'annotation de leur portée : dans le corpus brésilien (exemples 4, 5, 6) la portée se limite souvent aux groupes nominaux tandis qu'en français (exemples 1, 2, 3) toutes les séquences (groupes nominaux, propositions, subordinées...) impactées par le marqueur font partie de la portée. Ainsi, dans l'exemple 4, *apresentar* n'est pas annoté, ni

2. <https://github.com/ncbi-nlp/NegBio>

3. <https://www.pucpr.br/>

4. <http://ensaiosclinicos.gov.br/>

usuários de dans l'exemple 6.

1. **absence** [de ganglion métastatique]
2. En cas d'**in**[opérabilité] et/ou **im**[possibilité de réirradier],...
3. ...[des patients éligibles atteints d'une tumeur maligne et porteurs de la mutation BRAFV600] ayant précédemment été inclus et traités dans un protocole antérieur portant sur le Vemurafénib et **n'**[ayant] **pas** [satisfait aux critères du protocole sur la progression de la maladie]...
4. Grupo Controle : **Não** apresentar [DTM].
5. **Ausência de** [evidência clínica de imunossupressão].
6. **não** usuários de [drogas que causam dependência química].

Comme nous pouvons voir, la négation peut être exprimée de différentes manières : morphologique (*im-*), lexicale (*absence*, *ausência*) et grammaticale (*ne pas*, *não*).

Dans le tableau 1, nous présentons quelques statistiques sur nos corpus : le nombre de mots, la variété du vocabulaire, le nombre de phrases et le nombre de phrases négatives.

	FR	PTBR
Mots	134 386	48 204
Vocabulaire	8 133	6 453
Phrases	5 394	3 228
Phrases négatives	820 (15,20 %)	640 (19,83 %)

TABLEAU 1 – Statistique des corpus constitués pour notre travail

Dans le tableau 2, nous présentons des exemples de phrases en français et en brésilien annotées avec la négation et sa portée. Les champs renseignés sont : l'identifiant unique par phrase (*#Phrase*), la position de chaque token dans la phrase (*Position*), les différents descripteurs linguistiques (*Forme*, *Lemme*, *POS-tag*), ainsi le marqueur de la négation et sa portée. Le dernier exemple du tableau est issu du corpus brésilien. Comme d'autres phrases de ce corpus, il ne contient qu'un seul mot : *gestantes* signifiant que l'étude porte sur les femmes enceintes. Les phrases affirmatives sont indiqués par *** dans la sixième colonne.

4 Méthodes

Nous abordons la tâche de détection de la portée de la négation comme une tâche de classification binaire (*Inside-Outside*). Les informations recherchées doivent être prédites grâce à l'algorithme d'apprentissage supervisé, que nous décrivons dans la suite de cette section.

Inspirée de Fancellu *et al.* (2016), notre approche, dédiée à la détection de la portée de la négation, utilise un réseau de neurones Long Short-Term Memory (LSTM) bidirectionnel suivi, en sortie, d'une prédiction par champs aléatoires conditionnels (CRF). Un LSTM bidirectionnel est la combinaison de deux structures de réseaux de neurones récurrents (RNN) : le réseau de neurones récurrent bidirectionnel proposé par Schuster & Paliwal (1997), qui opère dans le sens de lecture et dans le sens contraire. La passe arrière est particulièrement importante dans le cas de la détection de la portée puisque les unités affectées peuvent également se trouver avant le marqueur. Le LSTM proposé par

#Phrase	Position	Forme	Lemme	POS-tag	Marqueur	Portée
5105	0	L'	le	DET :ART	—	—
5105	1	abdomen	abdomen	NOM	—	abdomen
5105	2	est	être	VER :pres	—	—
5105	3	souple	souple	ADJ	—	—
5105	4	et	et	KON	—	—
5105	5	sans	sans	PRP	sans	—
5105	6	défense	défense	NOM	—	défense
5105	7	.	.	SENT	—	—
2247	0	déficit	déficit	NOUN	—	—
2247	1	visual	visual	ADJ	—	—
2247	2	grave	grav	ADJ	—	—
2247	3	sem	sem	ADP	sem	—
2247	4	correção	correçã	NOUN	—	correção
2247	5	.	.	PUNCT	—	—
2916	0	Gestantes	gestant	ADJ	***	—
2916	1	.	.	PUNCT	***	—

TABLEAU 2 – Extraits des corpus annotés en français et brésilien

Hochreiter & Schmidhuber (1997) est également plus efficace que le RNN classique pour apprendre des dépendances de long-terme. Les CRFs (Lafferty *et al.*, 2001), modèles statistiques souvent utilisés pour des données séquentielles, semblent être particulièrement efficaces pour l'étiquetage de séquences dans les textes. Nous les préférons donc à la couche de prédiction la plus courante, la couche *softmax*.

Implémenté à l'aide de *Tensorflow* (Abadi *et al.*, 2016), notre système prend en entrée une instance $I(n, c, t)$, où chaque mot est représenté par un vecteur n (*word-embedding*), un vecteur c , qui détermine si le mot fait partie d'un marqueur (*cue-embedding*), ainsi que d'un vecteur t , qui est la représentation vectorielle de l'étiquetage morpho-syntaxique pour chaque mot (*postag-embedding*).

Pour chaque système, nous utilisons des paramètres d'entraînement définis de façon empirique. Nos embeddings sont de dimension $k = 50$. La couche cachée compte 200 unités (400 pour le BiLSTM, qui nécessite deux couches cachées concaténées). 50 périodes d'entraînement permettent d'atteindre le meilleur score F_1 sur l'ensemble de validation.

Les corpus sont segmentés en trois parties : 60% pour l'ensemble d'entraînement, 15% pour l'ensemble de validation et 25% pour l'ensemble de test. Les résultats obtenus sur le corpus de test sont évalués contre les données de référence avec les mesures d'évaluation classiques : la précision P , qui quantifie la pertinence de l'étiquetage, le rappel R , qui quantifie la sensibilité de l'étiquetage, ainsi que la moyenne harmonique de la précision et du rappel noté F_1 .

5 Résultats et discussion

Dans le tableau 3, nous présentons les résultats de notre approche sur nos deux corpus. Sont indiqués les résultats sur les ensembles de validation et de test. À titre de comparaison, nous indiquons

Dataset	Mots étiquetés			Portées exactes		
	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>P</i>	<i>R</i>	<i>F</i> ₁
français-valid	93,72	86,22	89,81	100	72,48	84,04
français-test	88,29	84,68	86,45	100	53,55	69,75
PTBR-valid	79,06	69,93	74,22	100	33,71	50,42
PTBR-test	77,61	64,58	70,50	100	34,29	51,06
SEM-2012	91,24	87,10	89,12	100	62,5	76,92
*SEM-2012**	92,62	85,13	88,72	99,40	63,87	77,7

TABLEAU 3 – Résultats de notre système sur nos corpus. Les résultats sont donnés en pourcentage. (Dalloux *et al.*, 2017)*, (Fancellu *et al.*, 2016)**

également les résultats que nous obtenons sur le corpus de *SEM-2012, ainsi que ceux de Fancellu *et al.* (2016) obtenus sur ce même corpus. Bien que l’ensemble d’entraînement français soit plus petit (d’environ 40 %), nous obtenons des résultats proches de ceux obtenus sur le corpus en anglais, avec un score *F*₁ de 86,45 en français contre 89,12 en anglais. Les résultats sont bien moins convaincants sur le corpus brésilien, qui est encore plus petit (d’environ 55 % pour l’ensemble d’entraînement). Il nous semble que les différences dans la façon d’annoter ont également une influence importante sur les résultats, l’annotation en français se rapprochant bien plus de celle de *SEM-2012 qu’en brésilien. Notons que la précision est privilégiée par le système, avec des valeurs proches ou égales à 100.

Un examen des résultats permet d’isoler des cas récurrents d’erreurs. Pour le corpus français, nous pouvons ainsi noter l’omission de groupes verbaux et l’inclusion d’incises :

- GOLD : Dans cet essai, **ni** [le patient,] **ni** [le médecin] **ne** [connaîtront quel traitement] (ProCervix ou placebo) [est administré].
- PRED : Dans cet essai, **ni** [le patient], **ni** [le médecin] **ne** [connaîtront quel traitement] (ProCervix ou placebo) **est administré**.
- GOLD : Par ailleurs [les patientes du second groupe ayant un risque de rechute potentiellement bas] (grade génomique bas) **ne** [recevront] **pas** [de chimiothérapie].
- PRED : Par ailleurs **les patientes du second groupe ayant** [un risque de rechute potentiellement bas] (**grade génomique bas**) **ne** [recevront] **pas** [de chimiothérapie].

Dans le corpus brésilien, les cas d’erreurs s’avèrent plus complexes. En effet, nous constatons qu’un grand nombre de prédictions combinent des erreurs de précision et de rappel au sein d’une même phrase. Les exemples ci-dessous illustrent cette situation :

- GOLD : **Ausencia** de diagnóstico de [doenças neuromusculares], [trauma], [tumores] ou [abscessos raquimedulares], [hemiplegia]/ [paresia], [lesão de plexo] ou [encefalopatia cerebral].
- PRED : **Ausencia** de diagnóstico [**de** doenças neuromusculares], [trauma], **tumores** ou [abscessos raquimedulares], **hemiplegia/paresia**, **lesão de plexo** ou [encefalopatia] **cerebral**.
- GOLD : que **não** apresentem outras [doenças neurológicas] ou [ortopédicas diagnosticadas].
- PRED : que **não** apresentem [**outras** doenças neurológicas] ou [ortopédicas] **diagnosticadas**.

Une des erreurs concerne l’inclusion ou non de déterminants (*de*, *outras*) dans la portée des marqueurs. Une autre erreur concerne la complétude de groupes nominaux (*encefalopatia cerebral* et *ortopédicas diagnosticadas*). Il s’agit sans doute des annotations non homogènes, qui peuvent être corrigées assez facilement.

6 Conclusions et perspectives

Dans cet article, nous proposons deux types de contributions.

Une première contribution concerne deux nouveaux corpus de données biomédicales, en français et en brésilien, annotées avec les informations sur la négation (les marqueurs et leur portée). Les corpus seront finalisés grâce à l'intégration de données nouvelles provenant d'autres sources, telles que des articles scientifiques ou bien des cas cliniques. L'accord inter-annotateur est en cours de calcul, s'en suivra une étape d'harmonisation des annotations.

La seconde contribution de notre travail consiste en l'exploitation d'un LSTM bidirectionnel pour la détection automatique de la portée de la négation. Les expériences sont effectuées et évaluées dans deux langues qui n'ont pas connu beaucoup de travaux de ce type : en portugais brésilien et en français. Sur le corpus français, notre système montre des performances qui s'approchent de celles de Packard *et al.* (2014) et Fancellu *et al.* (2016) sur les données de *SEM-2012, bien que le contenu de nos corpus soit très différent et qu'il nous semble y trouver plus de variations dans la structure de phrases négatives (énumérations, grandes portées discontinues, etc.).

Néanmoins, il nous reste plusieurs points à explorer. En effet, nous souhaitons exploiter différentes architectures de réseaux de neurones récurrents, ainsi que différents types de *word embeddings*. Nous prévoyons également de comparer nos résultats sur le français avec des systèmes experts tels que ceux proposés par Deléger & Grouin (2012) ou bien par (Abdaoui *et al.*, 2017).

Une première version de notre système est désormais accessible et utilisable en ligne à cette adresse : <https://allgo.inria.fr/webapps/173>

Remerciements

Ce travail a bénéficié d'une aide de l'état attribuée au labex COMIN LABS et gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'avenir » portant la référence ANR-10-LABX-07-01.

Nous remercions également les relecteurs pour les remarques constructives qui ont permis d'améliorer la présentation de notre travail.

Références

ABADI M., AGARWAL A., BARHAM P., BREVDO E., CHEN Z., CITRO C., CORRADO G. S., DAVIS A., DEAN J., DEVIN M. & ET AL. (2016). Tensorflow : Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv :1603.04467*.

ABDAOUI A., TCHECHMEDJIEV A., DIGAN W., BRINGAY S. & JONQUET C. (2017). *French ConText : Détecter la négation, la temporalité et le sujet dans les textes cliniques Français*, In 4e édition du Symposium sur l'Ingénierie de l'Information Médicale.

ALBRIGHT D., LANFRANCHI A., FREDRIKSEN A., STYLER IV W. F., WARNER C., HWANG J. D., CHOI J. D., DLIACH D., NIELSEN R. D., MARTIN J. *et al.* (2013). Towards comprehensive

syntactic and semantic annotations of the clinical narrative. *Journal of the American Medical Informatics Association*, **20**(5), 922–930.

CHAPMAN W. W., BRIDEWELL W., HANBURY P., COOPER G. F. & BUCHANAN B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, **34**(5).

DALLOUX C. (2017). Détection de l'incertitude et de la négation : un état de l'art. In *19es REcontres jeunes Chercheurs en Informatique pour le TAL (RECITAL 2017)*, p. 94–107.

DALLOUX C., CLAVEAU V. & GRABAR N. (2017). Détection de la négation : corpus français et apprentissage supervisé. In *SIIM 2017 - Symposium sur l'Ingénierie de l'Information Médicale*, p. 1–8, Toulouse, France.

DELÉGER L. & GROUIN C. (2012). Detecting negation of medical problems in french clinical notes. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*.

FANCELLU F., LOPEZ A. & WEBBER B. (2016). Neural networks for negation scope detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 1.

HARKEMA H., DOWLING J. N., THORNBLADE T. & CHAPMAN W. W. (2009). Context : an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of biomedical informatics*, **42**(5), 839–851.

HOCHREITER S. & SCHMIDHUBER J. (1997). Long short-term memory. *Neural Comput.*, **9**(8).

LAFFERTY J., MCCALLUM A., PEREIRA F. *et al.* (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning, ICML*, volume 1.

MUTALIK P. G., DESHPANDE A. & NADKARNI P. M. (2001). 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 : JAMIA*, **8**(6).

NGUYEN D. Q., NGUYEN D. Q., PHAM D. D. & PHAM S. B. (2015). A robust transformation-based learning approach using ripple down rules for part-of-speech tagging. *AI Communications*, **29**(3), 409–422.

PACKARD W., BENDER E. M., READ J., OEPEN S. & DRIDAN R. (2014). *Simple Negation Scope Resolution through Deep Parsing : A Semantic Solution to a Semantic Problem.*, In *ACL (1)*.

PENG Y., WANG X., LU L., BAGHERI M., SUMMERS R. & LU Z. (2018). Negbio : a high-performance tool for negation and uncertainty detection in radiology reports. *AMIA 2018 Informatics Summit*.

READ J., VELLDAL E., ØVRELID L. & OEPEN S. (2012). *Uio 1 : Constituent-based discriminative ranking for negation resolution*, In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1 : Proceedings of the main conference and the shared task, and Volume 2 : Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics.

SCHMID H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing, Manchester, UK*.

SCHUSTER M. & PALIWAL K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, **45**(11).

UZUNER Ö., SOUTH B. R., SHEN S. & DUVALL S. L. (2011). 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, **18**(5), 552–556.

VELLDAL E., ØVRELID L., READ J. & OEPEN S. (2012). Speculation and negation : Rules, rankers, and the role of syntax. *Computational Linguistics*, **38**(2).

VINCZE V., SZARVAS G., FARKAS R., MÓRA G. & CSIRIK J. (2008). The bioscope corpus : biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, **9**.

