



HAL
open science

Détection de la négation : corpus français et apprentissage supervisé

Clément Dalloux, Natalia Grabar, Vincent Claveau

► To cite this version:

Clément Dalloux, Natalia Grabar, Vincent Claveau. Détection de la négation : corpus français et apprentissage supervisé. *Revue des Sciences et Technologies de l'Information - Série TSI: Technique et Science Informatiques*, 2019, pp.1-21. hal-02402913

HAL Id: hal-02402913

<https://hal.science/hal-02402913>

Submitted on 10 Dec 2019

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

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

Détection de la négation : corpus français et apprentissage supervisé

Clément Dalloux¹, Natalia Grabar², Vincent Claveau¹

1. Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France

prenom.nom@irisa.fr

2. UMR 8163 STL CNRS, Université de Lille, France

natalia.grabar@univ-lille.fr

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 deux contributions liées à ce problème. Nous présentons d'une part un corpus constitué d'extraits de protocoles d'essais cliniques en français et dédié aux critères d'inclusion de patients. Ce corpus est annoté manuellement avec les marqueurs de négation et de leurs portées. Nous présentons d'autre part deux approches, basées sur l'apprentissage supervisé, pour extraire automatiquement les informations liées à la négation. De plus, une des approches est validée grâce à son application aux données de l'état de l'art en anglais : elle montre de très bons résultats en anglais, qui sont supérieurs à ceux présentés dans les travaux existants, et permet d'obtenir des résultats comparables en français. Enfin, nous proposons plusieurs expériences en variant les représentations vectorielles des mots.

ABSTRACT. Automatic detection of negated content is often a pre-requisite in information extraction systems, especially in the biomedical domain. This paper proposes two main contributions in this field. It first introduces a corpus built with excerpts from clinical trial protocols in French, describing the inclusion criteria for patient recruitment. The corpus is manually annotated for marking up the negation cues and their scope. Secondly, two supervised learning approaches are proposed for the automatic detection of negation. Besides, one of the approaches is validated on English data from the state of the art: the approach shows very good results and outperforms existing approaches, and it also yields comparable results on the French data. Finally, we propose several experiments with different vector representations of words.

MOTS-CLÉS : Négation, français, apprentissage supervisé, CRF, réseaux de neurones, plongement de mots, essai clinique

KEYWORDS: Negation, French, supervised learning, CRF, neural networks, word embeddings, clinical trials

1. Introduction

La détection de la négation dans les textes devient un des pré-requis incontournables et utiles dans de nombreuses tâches d'extraction d'informations. Dans le domaine biomédical en particulier, la négation est très fréquente et joue un rôle important (Elkin *et al.*, 2005 ; Denny, Peterson, 2007 ; Gindl *et al.*, 2008 ; Chapman *et al.*, 2001). Dans le cas d'essais cliniques, par exemple, elle peut fournir des critères déterminants pour recruter ou non un patient. Elle fournit des informations cruciales dans de nombreuses situations comme : détecter les pathologies et les co-morbidités d'un patient, déterminer le statut fumeur ou non d'une personne, détecter la prise ou non d'un médicament donné, définir si une patiente est enceinte ou non au moment du recrutement. Nous proposons donc de travailler sur la détection automatique de la négation. Cet article est la version étendue de nos travaux précédents (Dalloux *et al.*, 2017 ; Dalloux, 2017 ; Dalloux *et al.*, 2018). Nous proposons ainsi de faire la synthèse de ces travaux dédiés à la détection de la négation dans les documents de santé en français.

Dans ce qui suit, nous présentons d'abord une brève analyse du problème. Nous présentons ensuite les travaux qui existent. Nous décrivons les corpus constitués et exploités dans notre travail, de même que les méthodes que nous proposons. Nous rapportons ensuite les résultats obtenus que nous commentons. Nous terminons avec une conclusion et les pistes pour le travail futur.

2. Analyse du problème

Comme cela est souligné par plusieurs travaux existants (Elkin *et al.*, 2005 ; Denny, Peterson, 2007 ; Gindl *et al.*, 2008 ; Chapman *et al.*, 2001), la négation est très fréquente dans le domaine biomédical et y joue un rôle important. Cependant, son expression est très variable dans la langue, ce qui présente un premier défi pour sa détection automatique. Ainsi, les marqueurs de la négation sont d'habitude constitués d'un ou de plusieurs mots qui souvent modifient la polarité et donc le sens de la phrase. Notons que la négation peut également être marquée par les préfixes, comme les préfixes *an-*, *in-*, *im-*, *ir-*, *dis-*. Dans les exemples qui suivent les marqueurs sont soulignés :

1. Asthme : [*les hormones protègent*] les hommes, pas [*les femmes*].
2. L'étude vise également à vérifier que [*cette information*] est recevable, ne [*gène*] pas [*de stress*],[...].
3. Le traitement par tazemetostat continuera jusqu'à progression de la maladie ou l'apparition d'[un effet] in[*désirable*] inacceptable.
4. Le traitement par tazemetostat continuera jusqu'à progression de la maladie ou l'apparition d'[un effet *indésirable*] in[*acceptable*].
5. Une discipline pas **toujours** suivie par les personnes concernées.
6. il n'y a pas de traitement curateur de la maladie **en dehors de** l'allogreffe de moelle.

Même si la détection des marqueurs peut être une tâche assez complexe, à cause, entre autres, de leur variété et ambiguïté, elle n'est pas suffisante en soi. Ainsi, en plus de la détection des marqueurs, il est également nécessaire de calculer leur portée : de décider quel est l'effet du marqueur et si cet effet s'étend sur toute la phrase ou sur une partie de cette phrase. De plus, les marqueurs peuvent fonctionner sous différents régimes et leur portée peut aller à droite ou à gauche, s'étendre des deux côtés (exemple en (1)), être discontinue (exemple en (2)) ou se chevaucher (exemples en (3-4)). Il s'agit du deuxième défi lié à la détection et au traitement de la négation. Dans les exemples ci-dessus, la portée est marquée entre crochets. Finalement, la négation peut interagir avec d'autres éléments linguistiques de la phrase. Par exemple, elle peut être annulée par un adverbe de fréquence ou une locution prépositive (exemple en (5-6)). Il s'agit du troisième défi lié à la détection de la négation. Pour ces diverses raisons, la détection des informations liées à la négation présente de multiples difficultés que nous proposons de traiter.

3. Travaux existants

Nous présentons d'abord les corpus créés, dans lesquels la négation est annotée manuellement pour permettre sa détection automatique. Ensuite, nous décrivons assez brièvement les deux types d'approches existantes : les systèmes experts basés sur les règles et les approches par apprentissage automatique.

3.1. Corpus annotés avec la négation

Ces dernières années, avec la démocratisation des techniques d'apprentissage supervisé, plusieurs corpus de spécialité ont été annotés par les informations liées à la négation, ce qui a permis d'entraîner des modèles pour la détection automatique de la négation. Ces corpus décrivent le phénomène de la négation en langue anglaise. Les corpus peuvent être divisés en deux catégories : (1) les corpus annotés avec les marqueurs et les portées de la négation, comme Bioscope (Vincze *et al.*, 2008) ou *SEM-2012 par exemple, et (2) les corpus se focalisant également sur les concepts et les entités nommées, tels que SHARPN, I2B2 et Mipacq. Dans certains de ces corpus, le pivot d'annotation est la négation alors que dans d'autres ce sont les entités nommées. Nous décrivons brièvement ces corpus.

3.1.1. Corpus Bioscope

Le corpus Bioscope (Vincze *et al.*, 2008) est composé de comptes-rendus d'exams radiologiques, d'articles scientifiques ainsi que de résumés d'articles du domaine biomédical. Disponible au format XML, chaque phrase et chaque paire portée/marqueur (*scope* et *cue*) sont indexées par un identifiant unique. Dans ce corpus, le pivot d'annotation correspond aux marqueurs de négation et de l'incertitude. Le tableau 1 quantifie la composition du corpus Bioscope.

Tableau 1. Statistiques du corpus BioScope

	Compte-rendus	Articles	Résumés
Documents	1 954	9	1 273
Phrases	6 383	2 670	11 871
Phrases négatives	13,55 %	12,70 %	13,45 %
Marqueurs (négation)	877	389	1 848
Phrases incertaines	13,39 %	19,44 %	17,70 %
Marqueurs (incertitude)	1 189	714	2 769

3.1.2. Corpus *SEM-2012

Le corpus *SEM-2012 est composé d'un roman et de trois nouvelles de Sherlock Holmes par Conan Doyle et comprend 5 520 phrases, dont 1 227 négatives. Pour chaque occurrence de la négation, le marqueur et sa portée sont annotés, ainsi que l'objet de la négation s'il est présent dans la phrase. La particularité de ce corpus est que les marqueurs et les portées peuvent être discontinus, comme dans cet exemple :

[John had] never [*said* as much before].

Dans ce corpus, le pivot d'annotation correspond aussi aux marqueurs de négation. Le guide d'annotation utilisé dans ces travaux a été publié également (Morante *et al.*, 2011). Notons enfin qu'en plus des descripteurs lexicaux et syntaxiques, c'est-à-dire des lemmes, le corpus offre aussi des descripteurs syntaxiques : l'étiquetage morpho-syntaxique et l'analyse syntaxique de surface (*chunking*).

3.1.3. Corpus I2B2/VA-2010

La compétition I2B2/VA-2010 (Uzuner *et al.*, 2011) présentait trois tâches d'extraction d'informations à partir des dossiers cliniques américains. Une des tâches concernait la détection d'assertions et de leurs portées. Ainsi, à chaque concept médical devait être associée l'une de ces six classes d'assertions : *present*, *absent*, *possible*, *conditional*, *hypothetical* ou *not associated with the patient*. Dans ce corpus, ce sont donc les entités nommées qui sont les pivots des annotations et des études.

3.1.4. Corpus Mipacq

Le corpus Mipacq (Albright *et al.*, 2013) est un autre exemple de corpus, qui est constitué de données cliniques en anglais annotées par plusieurs couches d'étiquettes syntaxiques et sémantiques. Dans ce corpus, le pivot d'annotation correspond également aux entités nommées, et plus précisément aux entités UMLS. Chaque entité UMLS détectée dispose de deux emplacements d'attributs :

- *Negation*, qui peut prendre deux valeurs (*true* ou *false*),
- *Status*, qui peut prendre quatre valeurs (*none*, *possible*, *HistoryOf* ou *Family-HistoryOf*).

3.2. Détection automatique de la négation

D'une manière générale, il existe deux familles d'approches pour effectuer la détection automatique des marqueurs de la négation : l'utilisation de systèmes experts, raisonnant à partir de faits et règles connus pour répondre à des questions précises, et la classification par apprentissage automatique, 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. Nous décrivons ces deux familles d'approches en mentionnant quelques systèmes emblématiques.

3.2.1. Systèmes experts

Le fonctionnement des systèmes experts est basé sur un ensemble de faits connus (essentiellement des ressources et des règles) implémentés sous forme d'un algorithme ou d'un programme.

Parmi les systèmes experts dédiés à la détection de la négation, le plus populaire est sans doute *NegEx* (Chapman *et al.*, 2001), certainement à cause de son caractère précurseur et du fait que son algorithme et ses ressources soient librement disponible pour la recherche. Ce système utilise les expressions régulières pour détecter les marqueurs de négation et pour identifier les termes médicaux qui se trouvent sous leur portée. Ce système propose de prendre en compte l'orientation de la portée : droite (après le marqueur comme dans *absence de [maux de tête]*) ou gauche (avant le marqueur comme dans *[maux de tête] niés*). Ce système a été adapté à d'autres langues, comme le suédois (Velupillai *et al.*, 2011) ou le français (Deléger, Grouin, 2012).

D'autres systèmes basés sur l'exploitation de ressources et de règles sont apparus par la suite :

- *Negfinder* (Mutalik *et al.*, 2001) combine un analyseur lexical, qui utilise des expressions régulières afin de générer un automate à états finis, et un analyseur syntaxique, qui repose sur un sous-ensemble restreint de la grammaire non contextuelle LALR¹. Ainsi, *NegFinder* permet d'identifier les concepts impactés par la négation dans les textes médicaux lorsqu'ils sont proches de l'unité linguistique marquant la négation ;

- *ConText* (Harkema *et al.*, 2009) propose un algorithme dérivé de *NegEx* mais qui couvre des objectifs plus variés. Ainsi, il permet de détecter plusieurs types d'informations : la négation, la temporalité de même que le sujet concerné par ces informations dans les textes cliniques. Ce système a été adapté au français (Abdaoui *et al.*, 2017) ;

- *NegBio* (Peng *et al.*, 2018) repose sur des règles définies à partir des graphes de dépendances (*universal dependency graph* (UDG)). Le code de ce système est disponible en ligne².

1. Look-Ahead Left Recursive

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

3.2.2. Apprentissage supervisé

La deuxième famille rassemble de nombreux travaux qui utilisent la classification automatique 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. Les données nécessaires à la création de ces systèmes se trouvent dans les corpus annotés avec les informations requises (négation et sa portée).

Nous présentons ici quelques travaux typiques :

- Un système (Velldal *et al.*, 2012) considère l'ensemble de marqueurs de négation comme une classe fermée. Ce système utilise un SVM et de simples descripteurs de type n-grammes, calculés sur les mots et les lemmes, à droite et à gauche des marqueurs candidats. Ce système propose une détection hybride de la portée de la négation. Il combine ainsi des règles expertes, opérant sur des arbres syntaxiques de dépendances, avec un SVM d'ordonnement (*ranking*), qui opère sur des noeuds d'arbres syntaxiques des constituants syntagmatiques ;

- Un autre système (Read *et al.*, 2012) est basé sur le travail précédent (Velldal *et al.*, 2012) auquel les auteurs ont ajouté la prise en charge de l'identification des marqueurs de négation affixaux ;

- Dans un autre travail, les auteurs exploitent deux approches (Packard *et al.*, 2014). La première est basée sur un indexeur par *minimal recursion semantics* (MRS), alors que la seconde approche reprend le système de (Read *et al.*, 2012) ;

- Finalement, un système récent (Fancellu *et al.*, 2016) propose deux approches par réseau de neurones pour résoudre le problème de détection de la portée de la négation. Une première approche utilise un réseau de neurones *Feed-forward*, c'est-à-dire un réseau de neurones artificiel où les connexions entre les unités ne forment pas de boucle. La seconde approche, qui est la plus efficace pour la tâche ciblée, utilise un réseau de neurones Long Short-Term Memory (LSTM) bidirectionnel.

Comme c'est le cas pour d'autres travaux présentés, les systèmes par apprentissage automatique sont développés pour traiter les textes en anglais. Leur adaptation à d'autres langues dépend de la disponibilité de corpus annotés.

3.3. Extension sur les travaux existants

Dans cette section, nous décrivons les corpus et méthodes qui seront utilisés dans l'extension de l'article.

Le corpus de français médical du CRTT³ (CRTT-MED) est constitué d'articles de recherche dans le domaine médical tirés des revues disponibles sur la base de données Science Direct. Ce corpus est actuellement disponible sous sa forme brute et sous sa forme étiquetée en parties du discours et lemmatisée.

3. http://perso.univ-lyon2.fr/~maniezf/Corpus/Corpus_medical_FR_CRTT.htm

Le corpus QUAERO Médical du français⁴ est une ressource développée dans le cadre de la recherche en reconnaissance d'entité et normalisation (Névéol *et al.*, 2014). Ce corpus est composé de titres MEDLINE et de documents EMEA qui ont été annotés manuellement. Les annotations s'appuient sur les concepts de l'UMLS (Unified Medical Language System).

4. Généralités

Dans cette section, nous présentons les généralités sur les méthodes de représentations vectorielles des mots et de classification utilisées dans nos expérimentations.

4.1. Représentations vectorielles des mots

Les algorithmes et modèles d'apprentissage artificiel emploient des entrées numériques. Il est donc nécessaire de donner à chaque mot un identifiant unique qui le représente. Dans cette section, nous présentons plusieurs méthodes couramment utilisées dans ce but.

4.1.1. Encodage one-hot

Dans notre cas, un encodage one-hot consiste à représenter chaque mot par un vecteur binaire composée d'une suite de 0 et d'un 1, dont la taille dépend de la taille du vocabulaire. Par exemple :

- ['Cancer', 'du', 'rein'] devient [001, 010, 100]
- ['Cancer', 'du', 'rein', 'métastatique'] devient [0001, 0010, 0100, 1000]

4.1.2. Plongement de mots

Word embedding, ou plongement de mots (Vukotic *et al.*, 2015), est l'appellation regroupant un ensemble de méthodes de représentation des mots selon des principes distributionnels, où les mots ou syntagmes du vocabulaire sont représentés par des vecteurs de nombres réels. Les méthodes exploitées pour générer cette représentation incluent par exemple les réseaux de neurones et les modèles probabilistes. Dans cette section, nous présentons deux méthodes couramment utilisées : Word2vec et GloVe.

Word2vec (Mikolov *et al.*, 2013) est un modèle prédictif particulièrement efficace pour l'apprentissage des plongements de mots à partir du texte brut. Les plongements de mots peuvent être calculés selon deux modèles : le modèle utilisant un sac de mots continu (CBOW) et le modèle dit *Skip-Gram* (SG). Algorithmiquement, ces modèles sont similaires, sauf que CBOW prédit les mots cibles à partir des mots du contexte source, tandis que le modèle *skip-gram* fait l'inverse et prédit les mots du contexte source à partir des mots cibles. Statistiquement, cette différence a pour effet que CBOW lisse une grande partie de l'information distributionnelle en traitant un

4. <https://quaerofrenchmed.limsi.fr/>

contexte entier comme une observation. Cela s'avère être une chose utile pour les jeux de données plus petits. Cependant, le modèle *skip-gram* traite chaque paire contextuelle comme une nouvelle observation, ce qui a tendance à donner de meilleurs résultats lorsque les ensembles de données sont plus volumineux.

GloVe (*Global Vectors for Word Representation*) (Pennington *et al.*, 2014) est un algorithme d'apprentissage non supervisé également permettant d'obtenir des représentations vectorielles des mots. D'après les auteurs, *Word2vec* est sous-optimal car il n'exploite pas complètement les informations statistiques des co-occurrences des mots. Les auteurs proposent donc de dépasser cette limite. Ils développent ainsi le modèle *GloVe*, qui intègre les avantages du modèle *Skip-Gram* de *Word2vec* pour les tâches d'analogie, et les avantages des méthodes de factorisation de matrices, qui peuvent exploiter des informations statistiques globales. *GloVe* récupère des statistiques de cooccurrence de mots sous la forme d'une matrice de co-occurrences des mots X . Chaque élément X_{ij} de cette matrice représente la fréquence avec laquelle un mot i apparaît dans le contexte du mot j . Pour chaque terme, l'algorithme cherche les termes du contexte dans une certaine zone définie par une taille de fenêtre avant le terme et une taille de fenêtre après le terme. Les mots éloignés du mot pivot reçoivent moins de poids que les mots qui lui sont plus proches. Enfin, pour chaque paire de mots, l'algorithme applique des contraintes douces ainsi qu'une fonction objective.

4.2. Champs aléatoires conditionnels

Les champs aléatoires conditionnels (*conditional random fields* ou CRF) (Lafferty *et al.*, 2001) sont des méthodes statistiques utilisées en traitement automatique du langage naturel (TALN) pour étiqueter des séquences de mots. En entraînant un modèle sur les descripteurs appropriés (mots, lemmes, l'étiquetage morpho-syntaxique, etc.) et les étiquettes à prédire, les CRF obtiennent généralement de bons résultats avec un temps d'entraînement bien inférieur aux réseaux de neurones.

4.3. Réseau de neurones récurrents

Le fonctionnement d'un réseau de neurones récurrents (RNN, selon l'acronyme anglais de *Recurrent Neural Network*) est motivé par le fait qu'un être humain raisonne en s'appuyant sur les connaissances qu'il a acquises et qui restent dans sa mémoire. Si les réseaux de neurones « classiques » ne sont pas capables de reproduire ce processus, les RNN répondent en partie à ce problème. En effet, ce sont des réseaux qui prennent en compte les données vues précédemment, en plus de la donnée vue actuellement, pour adapter leur décision. Ce fonctionnement est mis en place grâce aux boucles dans l'architecture du réseau, ce qui permet aux informations de persister en mémoire, comme le montre la figure 1.

Parmi les RNN proposés dans l'état de l'art, les réseaux dits *Long short-term memory* (LSTM) (Hochreiter, Schmidhuber, 1997) sont les plus efficaces pour apprendre des dépendances de long terme et sont donc plus à même de résoudre le problème

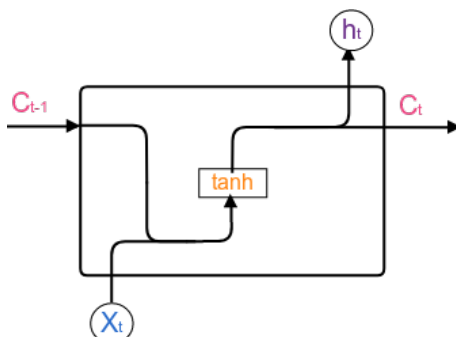


Figure 1. Cellule RNN classique.

de la portée discontinue, par exemple. Par ailleurs, les cellules LSTM sont plus efficaces pour retenir les informations utiles lors de la rétropropagation du gradient, ce qui permet de corriger les différences entre les prédictions sortantes et les catégories de référence en calculant le gradient de l'erreur pour chaque neurone, en allant de la dernière couche vers la première. La figure 2 illustre les quatre couches interactives (sigmoïde et tanh), les trois portes (oubli f , entrée i et sortie o) et les opérations *pointwise* qui traitent le vecteur x à l'intérieur d'une cellule LSTM à un temps t .

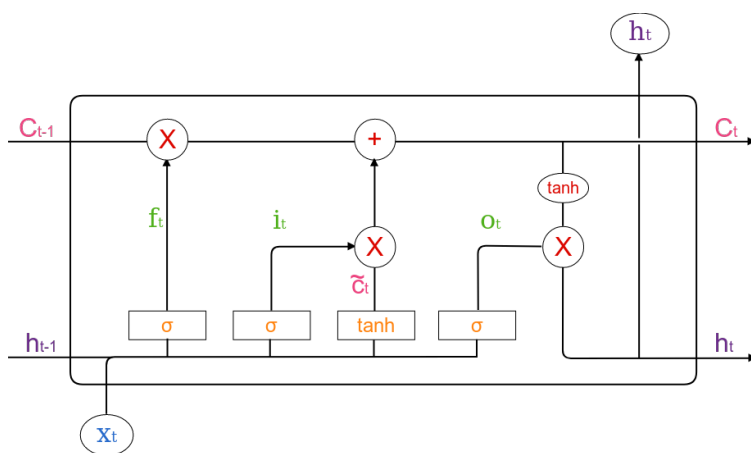


Figure 2. Cellule LSTM.

Le réseau Gated Recurrent Unit (GRU) (Cho *et al.*, 2014) est une variante du LSTM où les portes d'oubli et d'entrée sont fusionnées en une unique porte de mise à jour. L'état de la cellule C ainsi que l'état caché h sont aussi fusionnés, comme présenté à la figure 3. Le modèle produit est donc plus simple que le modèle obtenu avec un LSTM standard. En pratique, cette approche permet de réduire le temps de calcul d'un modèle tout en conservant des résultats équivalents à ceux d'un réseau LSTM.

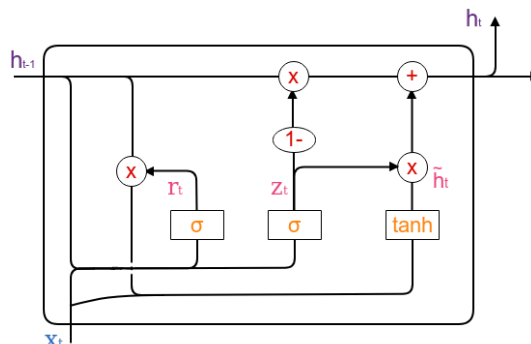


Figure 3. Cellule GRU.

5. Description et préparation du corpus

Le premier objectif que nous traitons consiste à créer un corpus annoté manuellement avec les marqueurs de la négation et leur portée. Les données cliniques étant difficile à obtenir et à partager, nous proposons un corpus de langue française de données médicales publiques partageables⁵. Notre corpus contient des protocoles d'essais cliniques en français ainsi que plusieurs cas cliniques. Les protocoles d'essais cliniques proviennent du registre des essais cliniques de l'hôpital Gustave Roussy⁶ et du registre de l'Institut National du Cancer⁷. Au sein des textes décrivant les protocoles des essais cliniques, nous avons décidé de nous concentrer sur les critères d'inclusion des patients et sur la description du déroulement des essais. Dans les deux cas, la négation est fréquente et très importante pour la description de la cohorte visée et le recrutement des patients. Les cas cliniques proviennent des articles scientifiques publiés en français. Il s'agit de textes qui ont des contenus proches des portions textuelles des dossiers médicaux. La négation est également fréquente dans ce type de textes.

Dans le tableau 2, nous présentons quelques statistiques sur ce corpus : le nombre de mots, la variété du vocabulaire, le nombre de phrases et le nombre de phrases négatives (avec leur pourcentage). Nous indiquons également l'évolution de ce corpus par rapport à nos travaux précédents (Dalloux *et al.*, 2017).

Le corpus est pré-traité avec TreeTagger (Schmid, 1994) pour effectuer l'étiquetage morpho-syntaxique et la lemmatisation.

Pour disposer des données de référence nécessaires à l'apprentissage automatique, le corpus a été annoté manuellement pour marquer les informations liées à la négation : les marqueurs et leur portée. Le tableau 3 fait le bilan de cette annotation et indique les marqueurs présents dans notre corpus et leur nombre d'occurrences. Nous voyons

5. Une version du corpus est disponible à cette adresse : <https://bit.ly/2NhrAaq>

6. <https://www.gustaveroussy.fr/fr/essais-cliniques>

7. <http://bit.ly/2ypwAoA>

Tableau 2. Statistique du corpus constitué et annoté pour notre travail

	Actuel	(Dalloux <i>et al.</i> , 2017)
Mots	134 386	104 432
Vocabulaire	8 133	6 643
Phrases	5 392	4 198
Phrases négatives	820 (15, 20 %)	650 (15, 48 %)

par exemple que certains marqueurs sont sous-représentés dans notre corpus, ce qui présente un problème connu pour les systèmes par apprentissage supervisé.

Tableau 3. Distribution des marqueurs

(ne/n') pas	253	in-	222	non(-)	171
Sans	125	ni	100	Absence de/du/d'	94
aucun	24	(ne/n') plus	9	(ne/n') jamais	8
excepté	4	sauf	3	à la place de	3
négatif	3	à l'exception de/du	2	absent	1
exclu	1	a-	1	im-	1
nul	1	dis-	1		

Dans le tableau 4, nous présentons l'exemple d'une phrase annotée avec différents niveaux d'informations linguistiques (l'ordre dans une phrase donnée, la forme, le lemme, l'étiquette morpho-syntaxique ou le POS-tag) et les catégories de l'annotation manuelle de référence (le marqueur de la négation et sa portée), qui se trouvent dans les deux dernières colonnes.

Tableau 4. Extrait du corpus, où les colonnes contiennent les informations linguistiques et de référence, à l'instar du format proposé par *SEM-2012

#P	Position	Forme	Lemme	POS Tag	Marqueur	Portée
0	0	Dans	dans	PRP	-	-
0	1	les	le	DET:ART	-	-
0	2	formes	forme	NOM	-	-
0	3	peu	peu	ADV	-	-
0	4	sévères	sévère	ADJ	-	-
0	5	,	,	PUN	-	-
0	6	une	un	DET:ART	-	une
0	7	antibiothérapie	antibiothérapie	NOM	-	antibiothérapie
0	8	spécifique	spécifique	ADJ	-	spécifique
0	9	n'	ne	ADV	n'	-
0	10	est	être	VER:pres	-	est
0	11	habituellement	habituellement	ADV	-	habituellement
0	12	pas	pas	ADV	pas	-
0	13	nécessaire	nécessaire	ADJ	-	nécessaire
0	14	.	.	SENT	-	-

Enfin, il est nécessaire de préciser qu'à ce jour, l'annotation du corpus a été réalisée par un seul annotateur. La méthode d'annotation est grandement inspirée de celles de

BioScope et de *SEM-2012. Le guide d’annotation est en cours de rédaction et un accord inter-annotateur sera calculé dès qu’au moins deux annotateurs supplémentaires auront annotés une portion suffisante du corpus.

6. Méthodes

Notre deuxième objectif consiste à proposer une approche pour la détection automatique et efficace de la négation et de sa portée. Nous abordons cette tâche comme une tâche de classification : les informations recherchées doivent être prédites grâce aux algorithmes d’apprentissage supervisé. Dans cette section, nous présentons notre approche pour la détection des marqueurs et notre approche pour la détection de la portée des marqueurs. Dans les deux cas, une évaluation est effectuée. Nous décrivons également l’extension de notre approche par rapport à nos travaux précédents.

6.1. Détection des marqueurs

Notre approche pour la détection des marqueurs est basée sur l’utilisation des CRF présentés plus haut. Nous effectuons la descente de gradient en utilisant la méthode *L-BFGS* (*Limited-memory BFGS*), un algorithme d’optimisation de la famille des méthodes de quasi-Newton, qui calcule approximativement l’algorithme de Broyden-Fletcher-Goldfarb-Shanno (BFGS) en utilisant une quantité limitée de mémoire. C’est un algorithme couramment utilisé en apprentissage artificiel pour l’estimation de paramètres. Notre implémentation utilise `sklearn-crfsuite`⁸.

En entrée de l’algorithme, nous exploitons les decrypteurs suivants : les mots, les lemmes ainsi que l’étiquetage morpho-syntaxique (POS), comme présenté dans le tableau 4. Nous testons différentes tailles de fenêtres : le CRF est entraîné sur une taille de fenêtre allant de deux à cinq mots/lemmes/POS avant et après un mot donné.

Pour évaluer notre système, nous utilisons la validation croisée à 10 plis (*10-fold Cross-validation*) et les mesures d’évaluation classiques : la précision P , qui quantifie la pertinence de la catégorisation automatique, le rappel R , qui quantifie la sensibilité de la catégorisation automatique, ainsi que la moyenne harmonique de la précision et du rappel notée F_1 .

6.2. Détection de la portée

Notre approche initiale (Dalloux *et al.*, 2017) est basée sur l’utilisation des réseaux de neurones récurrents. Les RNN sont bidirectionnels, c’est-à-dire qu’ils opèrent 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 mots affectés peuvent se trouver également avant le marqueur.

8. <https://sklearn-crfsuite.readthedocs.io/en/latest/>

Implémenté à l'aide de *Tensorflow* (Abadi *et al.*, 2016), notre système comprend une version adaptée au français du LSTM bidirectionnel (Fancellu *et al.*, 2016), ainsi que son adaptation en GRU bidirectionnel. En sortie, la prédiction est assurée soit par une couche *softmax*, qui est la méthode la plus courante, soit par une couche CRF, une méthode qui semble être particulièrement efficace pour l'étiquetage de séquences.

Les réseaux de neurones sont entraînés uniquement sur les phrases porteuses d'une négation. Le système de base 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*),
- un vecteur t qui est la représentation vectorielle de l'étiquetage morpho-syntaxique pour chaque mot (*postag-embedding*).

Pour chaque système, nous utilisons les mêmes paramètres d'entraînement définis de façon empirique :

- Les embeddings sont de dimension $k = 50$;
- La couche cachée compte 200 unités (400 pour le BiLSTM qui nécessite deux couches cachées concaténées);
- 30 périodes d'entraînement permettent d'atteindre le meilleur score F_1 possible sur l'ensemble de validation.

Comme l'illustre la figure 4, dans l'état actuel du corpus, il apparaît que l'augmentation de la quantité d'exemples de l'ensemble d'entraînement n'apporte plus d'amélioration de la F-mesure au-delà de 350 phrases.

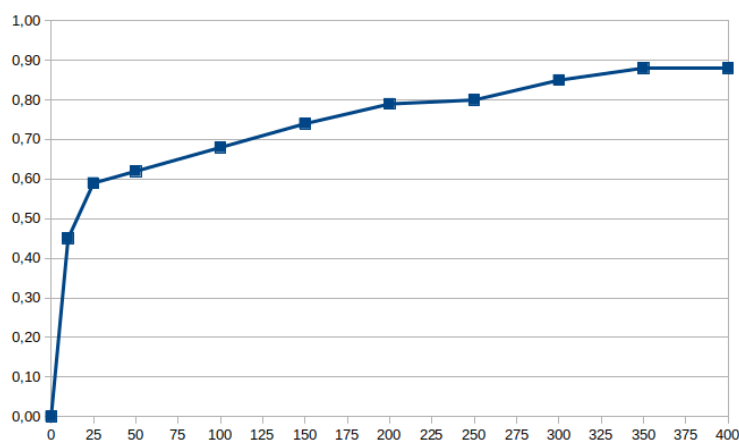


Figure 4. Progression de la F-mesure selon le nombre d'exemples utilisés pour l'entraînement

6.3. Extension de la méthode pour la détection de la portée

Notre méthode initiale a été étendue. Dans cette nouvelle expérimentation, nous comparons l'efficacité de deux types de représentations vectorielles des mots avec le *BiLSTM-CRF*. L'encodage *one-hot*, utilisé dans les expérimentations précédentes, et Word2vec, que nous proposons d'exploiter en plus. Deux modèles Word2vec sont entraînés.

Le premier modèle Word2vec est entraîné sur notre corpus ainsi que corpus additionnels, CRTT-MED et QUAERO. Malgré l'intérêt certain des annotations proposées par ces deux corpus, nous n'avons besoin que du texte brut pour entraîner nos plongement de mots. Par ailleurs, puisque cet ensemble de données reste relativement restreint, nous utilisons l'algorithme *COW* et la dimensionnalité des vecteurs est fixée à 50.

Le second modèle est entraîné avec l'algorithme *Skip-Gram* (SG) sur les mêmes données que le premier auxquelles est ajoutée l'intégralité des articles du Wikipedia français. La dimensionnalité des vecteurs est alors fixée à 300.

Ces modèles sont censés donner une représentation plus pertinente aux vecteurs de mot n que l'encodage *one-hot* (simple vecteur binaire).

À chaque entraînement, le corpus est segmenté en trois parties :

- 60 % pour l'ensemble d'entraînement,
- 15 % pour l'ensemble de validation,
- 25 % pour l'ensemble de test.

Les résultats obtenus sur le corpus de test sont évalués par comparaison avec les données de référence avec les mêmes mesures que dans le cas des marqueurs (précision P , rappel R et F-mesure F_1). Chaque système est entraîné et testé plusieurs fois sur des ensembles de données définis aléatoirement. Nous indiquons donc l'écart type des moyennes des F-mesures de différents modèles produits.

De plus, nous considérons deux applications de ces mesures. L'une sur les mots étiquetés, c'est-à-dire que la mesure est calculée sur chaque mot. L'autre est calculée sur la portée exacte, c'est-à-dire que toutes les étiquettes de l'énoncé doivent être correctes pour compter comme un vrai positif.

7. Résultats et discussion

Nous présentons les résultats obtenus sur le corpus français à deux étapes : la détection des marqueurs et la détection de la portée, en version initiale et étendue. Nous comparons également notre approche aux travaux de l'état de l'art en l'appliquant aux données en anglais.

7.1. Détection des marqueurs

Le tableau 5 présente les résultats obtenus avec notre approche par CRF, en fonction de la taille de la fenêtre. Bien que les résultats de ces expériences soient très proches, il apparaît qu'une fenêtre de taille 4 donne un meilleur équilibre entre précision et rappel. En outre, cette expérimentation a permis de systématiser les annotations et de compléter quelques occurrences oubliées de la négation.

Tableau 5. Résultats de l'approche par CRF sur notre corpus. Les résultats sont donnés en pourcentage en indiquant la taille de la fenêtre entre parenthèses

Système	<i>P</i>	<i>R</i>	<i>F</i> ₁
CRF (2)	98,01	96,91	97,46
CRF (3)	98,50	96,91	97,70
CRF (4)	98,44	97,12	97,77
CRF (5)	98,57	96,70	97,63

7.2. Détection de la portée avec l'approche initiale

Tableau 6. Résultats de différentes approches appliquées au corpus français (3 entraînements). Écart type des moyennes de *F*-mesures *s*_{*F*₁}. Les résultats sont donnés en pourcentage et les meilleurs scores sont indiqués en gras.

Système	Mots étiquetés				Portées exactes			
	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>s</i> _{<i>F</i>₁}	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>s</i> _{<i>F</i>₁}
BiLSTM	89,07	86,54	87,68	1,56	99,48	60,10	74,93	5,37
BiLSTM+CRF	87,69	87,20	87,40	1,24	99,91	61,95	76,48	2,37
BiGRU	89,69	82,68	85,95	1,26	99,95	58,42	73,74	1,30
BiGRU+CRF	90,93	82,83	86,67	1,08	99,95	59,76	74,77	2,79

Tableau 7. Comparaison des systèmes les plus performants sur les données de test en anglais de *SEM-2012. Les résultats sont donnés en pourcentage et les meilleurs scores sont indiqués en gras.

Système	Mots étiquetés			Portées exactes		
	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>P</i>	<i>R</i>	<i>F</i> ₁
(Read <i>et al.</i> , 2012)	81,99	88,81	85,26	87,43	61,45	72,17
(Lapponi <i>et al.</i> , 2012)	86,03	81,55	83,73	85,71	62,65	72,39
(Packard <i>et al.</i> , 2014)	86,1	90,4	88,2	98,8	65,5	78,7
(Fancellu <i>et al.</i> , 2016)	92,62	85,13	88,72	99,40	63,87	77,7
BiLSTM+CRF (nous)	92.70	85.36	88.87	99.43	66.16	79.45

Le tableau 6 présente les résultats obtenus avec notre approche initiale (Dalloux *et al.*, 2017) sur les données en français. Le tableau 7 présente les résultats obtenus sur

les données en anglais de la compétition *SEM-2012 par les travaux existants et par notre approche initiale.

Premièrement, comme nous nous y sommes attendus, les LSTM donnent de meilleurs résultats que les GRU. En effet, la revue des performances des RNN sur de multiples tâches (Jozefowicz *et al.*, 2015) indique que les GRU montrent toujours de meilleurs résultats que les LSTM, excepté pour la tâche de modélisation du langage. Bien que toutes les portes aient un impact positif sur les résultats, les expérimentations proposées dans ce travail montrent que c'est la porte d'oubli qui donne l'avantage au LSTM. Par ailleurs, la prédiction par CRF donne, pour le LSTM comme le GRU, de meilleurs résultats que la couche *softmax* classique dans la détection de portées exactes. Les CRF sont particulièrement efficaces pour l'étiquetage de séquences, ce qui pourrait expliquer le léger gain (1/1,5 points) obtenu.

Concernant l'ensemble de données de *SEM-2012, nous obtenons un score F_1 légèrement supérieur à celui du BiLSTM de (Fancellu *et al.*, 2016) en terme de mots correctement étiquetés. Cela s'explique par un meilleur équilibre entre la précision et le rappel. Notre score de F-mesure est légèrement supérieur lors de la détection de la portée exacte.

À titre de comparaison, les résultats des méthodes NegEx et ConText présentés dans le tableau 8 (Abdaoui *et al.*, 2017) sont sensiblement inférieurs aux nôtres. Cela dit, la comparaison peut difficilement être faite sans expérimenter ces différentes méthodes sur les mêmes ensembles de données.

Tableau 8. Résultats de NegEx en français et de French ConText sur deux types de textes cliniques.

			Précision	Rappel	F1
Dossiers patient	Ancienne adaptation de NegEx	Nié	0,989	0,648	0,783
		Affirmé	0,901	0,999	0,947
	French ConText	Nié	0,904	0,799	0,837
		Affirmé	0,963	0,986	0,975
Certificats de décès	Ancienne adaptation de NegEx	Nié	1	0,659	0,794
		Affirmé	0,986	1	0,993
	French ConText	Nié	0,970	0,780	0,865
		Affirmé	0,991	0,998	0,994

Un examen des résultats permet d'isoler les erreurs récurrentes. L'exemple ci-dessous, dans lequel le sujet est séparé par plusieurs mots de la négation le concernant, impacte le rappel :

– GOLD : *L'objectif de cet essai est de comparer l'efficacité et la tolérance de la gemcitabine administrée seule ou en association avec de l'oxaliplatine chez [des patients] ayant un cancer de la vessie à un stade avancé et ne [pouvant être traité par du cisplatine].*

– PRED : *L'objectif de cet essai est de comparer l'efficacité et la tolérance de la gemcitabine administrée seule ou en association avec de l'oxaliplatine chez **des patients** ayant un cancer de la vessie à un stade avancé et **ne** [pouvant être traité par] **du** [cisplatine].*

Un deuxième exemple illustre un autre problème d'erreur impactant la précision. Ici, le modèle prédit que « *et* » fait porter la négation sur l'énoncé qui le suit car il a sans doute rencontré ce type de phrase lors de l'entraînement :

– GOLD : *[La prise en charge thérapeutique] **ne** [sera] **pas** [modifiée par l'étude] et sera conforme aux référentiels.*

– PRED : *[La prise en charge thérapeutique] **ne** [sera] **pas** [modifiée par l'étude **et**] sera [conforme aux référentiels].*

7.3. Détection de la portée avec l'approche étendue

Le tableau 9 présente les résultats obtenus par l'approche étendue (*BiLSTM-CRF* avec différentes représentations de mots et entraînement sur un corpus français plus gros) sur le corpus actuel.

L'utilisation de nos modèles pré-entraînés sur des corpus français plus gros occasionne une baisse globale des performances. Cela dit, l'écart type s'améliore pour le modèle *Skip-Gram*. Ce dernier ne perd pas significativement en précision non plus. Il apparaît donc que l'utilisation d'un ensemble de données plus gros et en plus n'appartenant pas au domaine de spécialité traité, provoque une perte significative du rappel. En outre, le modèle *CBOW* occasionne une perte de plus d'un point en précision comme du rappel. Par ailleurs, ces résultats indiquent également que l'utilisation d'une représentation plus complexe des mots ne constitue pas forcément une solution dans la tâche que nous explorons.

Tableau 9. Résultats des différentes approches sur le corpus français (10 entraînements). Écart type des moyennes de F -mesures s_{F_1} . Les résultats sont donnés en pourcentage et les meilleurs scores sont indiqués en gras.

Système	Mots étiquetés				Portées exactes			
	P	R	F_1	s_{F_1}	P	R	F_1	s_{F_1}
Encodage one-hot	86,97	84,81	85,87	1,26	99,71	56,98	72,52	2,66
Word2vec (CBOW)	85,52	83,41	84,45	1,42	99,61	53,44	69,56	2,61
Word2vec (SG)	86,83	81,92	84,30	0,86	99,83	53,27	69,47	1,73

8. Conclusions et perspectives

Les travaux sur la détection automatique de la négation en langue anglaise par apprentissage supervisé se sont multipliés ces dernières années. Dans cet article, après avoir présenté les difficultés liées à cette tâche et après un bref rappel des travaux existants, nous avons présenté un nouveau corpus de données biomédicales, de langue

française, annotées avec les informations sur la négation (les marqueurs et leur portée). Avant sa diffusion auprès de la communauté de la recherche, le corpus sera finalisé grâce à l'intégration de données nouvelles provenant de protocoles d'essais cliniques, de cas cliniques et possiblement d'autres sources, suivie d'une étape d'harmonisation des annotations.

Une autre contribution de notre travail consiste à exploiter différents types de plongements de mots et de réseaux de neurones récurrents pour la reconnaissance automatique de la portée de la négation. Les expériences sont effectuées et évaluées dans deux langues : en anglais, qui dispose de données de référence indépendantes, et en français, qui n'a pas connu beaucoup de travaux de ce type. Notre système montre des performances très proches du meilleur système de l'état de l'art en anglais, et des performances très proches en français. Nous proposons donc une méthode simple et efficace pour la détection des marqueurs et de leur portée sur les données en français et en anglais.

Nos travaux indiquent également que les architectures neuronales fondées sur les LSTM sont plus efficaces que les GRU dans la tâche de détection de la portée. En outre, la couche de sortie CRF semble apporter de légers gains en terme de rappel par rapport à la couche *Softmax* grâce à une meilleure prise en compte des dépendances entre les catégories.

Dans l'avenir, nous prévoyons d'étendre le système sur la détection des incertitudes et de leurs portées. Notre système sera comparé avec un système expert existant, comme l'adaptation française de *NegEx* (Deléger, Grouin, 2012), sur les mêmes données.

Une première version de notre système est désormais utilisable en ligne via le service *ALLGO* : <https://algo.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.

Bibliographie

- Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C. *et al.* (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*. Consulté sur <https://arxiv.org/abs/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. *et al.* (2013). Towards comprehensive syntactic and semantic annotations of the clinical narrative. *Journal of the American Medical Informatics Association*, vol. 20, n° 5, p. 922–930.

- Chapman W. W., Bridewell W., Hanbury P., Cooper G. F., Buchanan B. G. (2001, Oct). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, vol. 34, n° 5.
- Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H. *et al.* (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. Consulté sur <https://arxiv.org/abs/1406.1078>
- Dalloux C. (2017). Détection de l'incertitude et de la négation : un état de l'art. In *19es rencontres jeunes chercheurs en informatique pour le tal (recital 2017)*, p. 94-107.
- Dalloux C., Claveau V., Grabar N. (2017, novembre). 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. Consulté sur <https://hal.archives-ouvertes.fr/hal-01659637>
- Dalloux C., Grabar N., Claveau V., Moro C. (2018). Portée de la négation : détection par apprentissage supervisé en français et portugais brésilien. *25e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*.
- Deléger L., Grouin C. (2012). Detecting negation of medical problems in french clinical notes. In *Proceedings of the 2nd acm sighth international health informatics symposium*.
- Denny J., Peterson J. (2007). Identifying qt prolongation from ecg impressions using natural language processing and negation detection. In *Medinfo*, p. 1283-8.
- Elkin P., Brown S., Bauer B., Husser C., Carruth W., Bergstrom L. *et al.* (2005). A controlled trial of automated classification of negation from clinical notes. *BMC Med Inform Decis Mak.*, vol. 5, n° 13.
- 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*, vol. 1. Consulté sur http://www.research.ed.ac.uk/portal/files/25518265/neural_networks_negation_1.pdf
- Gindl S., Kaiser K., Miksch S. (2008). Syntactical negation detection in clinical practice guidelines. In *Stud health technol inform*, p. 187-92.
- Harkema H., Dowling J. N., Thornblade T., Chapman W. W. (2009). Context: an algorithm for determining negation, experienter, and temporal status from clinical reports. *Journal of biomedical informatics*, vol. 42, n° 5, p. 839-851.
- Hochreiter S., Schmidhuber J. (1997, novembre). Long short-term memory. *Neural Comput.*, vol. 9, n° 8. Consulté sur <http://dx.doi.org/10.1162/neco.1997.9.8.1735>
- Jozefowicz R., Zaremba W., Sutskever I. (2015). An empirical exploration of recurrent network architectures. In *Proceedings of the 32nd international conference on machine learning (icml-15)*, p. 2342-2350.
- 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*, vol. 1.
- Lapponi E., Velldal E., Øvreliid L., Read J. (2012). Uio 2: sequence-labeling negation using dependency features. 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. Consulté sur <http://dl.acm.org/citation.cfm?id=2387687>
- Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, p. 3111–3119.
- Morante R., Schrauwen S., Daelemans W. (2011). Annotation of negation cues and their scope. guidelines v1.0.
- Mutalik P. G., Deshpande A., Nadkarni P. M. (2001, Dec). 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*, vol. 8, n° 6.
- Névéal A., Grouin C., Leixa J., Rosset S., Zweigenbaum P. (2014). The quæro french medical corpus: A resource for medical entity recognition and normalization. In *In proc biotextm, reykjavik*.
- Packard W., Bender E. M., Read J., Oepen S., Drīdan R. (2014). Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem. In *Acl (1)*. Consulté sur <https://www.aclweb.org/anthology/P/P14/P14-1007.pdf>
- 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*. Consulté sur <http://arxiv.org/abs/1712.05898>
- Pennington J., Socher R., Manning C. D. (2014). Glove: Global vectors for word representation. In *Empirical methods in natural language processing (emnlp)*, p. 1532–1543. Consulté sur <http://www.aclweb.org/anthology/D14-1162>
- Read J., Velldal E., Øvreid 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. Consulté sur <http://dl.acm.org/citation.cfm?id=2387686>
- Schmid H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of international conference on new methods in language processing, manchester, uk*.
- 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*, vol. 18, n° 5, p. 552–556.
- Velldal E., Øvreid L., Read J., Oepen S. (2012, Jun). Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics*, vol. 38, n° 2.
- Velupillai S., Dalianis H., Kvist M. (2011). Factuality levels of diagnoses in swedish clinical text. In *Stud health technol inform*, p. 559-63.
- 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*, vol. 9.
- Vukotic V., Claveau V., Raymond C. (2015, juin). IRISA at DeFT 2015: Supervised and Unsupervised Methods in Sentiment Analysis. In *DeFT, Défi Fouille de Texte, joint à la conférence TALN 2015*. Caen, France. Consulté sur <https://hal.archives-ouvertes.fr/hal-01226528>

Biographies des auteurs

Clément Dalloux est doctorant en informatique à l'université de Rennes 1 depuis décembre 2016. Sa thèse, intitulée « Fouille de texte et extraction d'informations dans les données cliniques », s'inscrit dans le cadre du projet LABEX CominLabs Big-Clin et ses travaux sont principalement axés sur le traitement automatique du langage naturel appliqué aux données médicales.

Natalia Grabar est chercheur CNRS depuis 2010. Elle est affiliée à l'UMR 8163 STL à Lille. Ses thématiques de recherche sont liées à la construction et utilisation de terminologies ; la recherche et extraction d'informations ; la qualité et compréhension d'informations médicales et de santé.

Vincent Claveau est chargé de recherche au CNRS depuis 2005. Au sein du laboratoire IRISA à Rennes, cet informaticien s'intéresse au traitement automatique des langues, à la recherche d'information et à l'analyse des documents multimédia. Depuis quelques années, il développe des recherches sur l'extraction d'information que ce soit dans le domaine médical ou dans le contexte des réseaux sociaux.