Speculation and negation detection in french biomedical corpora

Clément Dalloux, Vincent Claveau, Natalia Grabar

HAL Id: hal-02284444
https://hal.archives-ouvertes.fr/hal-02284444

SPECULATION and NEGATION DETECTION in FRENCH BIOMEDICAL CORPORA

Clément Dalloux and **Vincent Claveau**

Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France
`name.surname@irisa.fr`

Natalia Grabar

UMR 8163 STL CNRS, Université de Lille, France
`natalia.grabar@univ-lille.fr`

Abstract

In this work, we propose to address the detection of negation and speculation, and of their scope, in French biomedical documents. It has been indeed observed that they play an important role and provide crucial clues for other NLP applications. Our methods are based on CRFs and BiLSTM. We reach up to 97.21 % and 91.30 % F-measure for the detection of negation and speculation cues, respectively, using CRFs. For the computing of scope, we reach up to 90.81 % and 86.73 % F-measure on negation and speculation, respectively, using BiLSTM-CRF fed with word embeddings.

1   INTRODUCTION

The detection of speculation and negation in texts has become one of the unavoidable pre-requisites in many information extraction tasks. Both are common in language and provide information on factuality and polarity of facts, which is particularly important for the biomedical field (Elkin et al., 2005b; Denny and Peterson, 2007; Gindl et al., 2008; Chapman et al., 2001). Indeed, negation and speculation provide there crucial information for detecting patient's present, speculated or absent pathologies and co-morbidities, detecting whether a particular medication has been, may have been, or has not been prescribed or taken, defining the certainty of diagnosis, etc. In order to efficiently identify speculation and negation instances, it is first necessary to identify their cues, i.e., words (or morphological units) that express speculation and negation, and then their scopes, i.e., tokens within the sentence which are affected by the negation or speculation. In this paper, we present two French datasets annotated with negation and speculation cues and their scope. We also propose machine learning and deep learning based systems to tackle the automatic detection of cues and scopes that achieve high performance.

2   EXPRESSION of NEGATION and SPECULATION

In French, the expression of negation and speculation have some specifics that are described below.

2.1   NEGATION

The negation cues may consist of one word or prefix, or of multiple words. Moreover, negation can be expressed via a large panel of cues which can be morphological (*an, in, im, ir, dis*), lexical (*absence de(absence of)*, *à l'exception de(excepting)*), and grammatical (*non, ne...pas, ni...ni*). In the following examples, we present sentences with instances of negation (the cues are underlined).

1. *En alternative des traitements locaux (chirurgie, radiothérapie, radiofréquence, cryoablation) peuvent être indiqués mais ils <u>ne</u> sont <u>pas</u> **toujours** faisables.* (Alternatively, local treatments (surgery, radiotherapy, radiofrequency, cryoablation) may be indicated but are <u>not</u> **always** feasible.)

2. *Il <u>n'</u>existe **toujours** <u>pas</u> aujourd'hui de consensus quant à une définition précise de ce phénomène hétérogène ou des modalités de sa prise en charge.* (There is **still** <u>no</u> consensus today on a precise definition of this heterogeneous phenomenon or the modalities of its management.)

3. *il <u>n'</u>y a <u>pas</u> de traitement curateur de la maladie **en dehors de** l'allogreffe de moelle.* (there is <u>no</u> curative treatment for this disease **apart from** bone-marrow homograft)

4. *Lymphome <u>non</u> hodgkinien à cellules B matures récidivant/réfractaire.* (Relapsed/refractory mature B-cell <u>non</u>-Hodgkin's lymphoma.)

5. *Elle n'est soulagée que par la marche et doit donc écouter la télévision en faisant les cent* **<u>pas</u>** *dans son salon.* (She is only relieved by walking and must therefore listen to the TV pacing in her living room.)

Examples (1-2) show the possible effect of frequency adverbs, here *toujours (always)*. In the first sentence *traitements locaux (chirurgie, radiothérapie, radiofréquence, cryoablation)*, the content would be negated without *toujours(always)*. In the second sentence, with or without *toujours*, the meaning of the sentence does not change, therefore, the scope of the negation remains the same.

Example (3) shows how the preposition *en dehors de(apart from)*, can stop the scope of negation. Many other prepositions such as *à part, à l'exception de* or *excepté*, with more or less the same meaning than *en dehors de (apart from)*, would have the same effect on the negation scope. However, these prepositions can also play the role of negation by themselves.

Examples (4) show that cues can also be included in medical concepts such as **non hodgkinien** *(non-Hodgkin's)*. In our work, we chose to label *hodgkinien* as part of the negation scope.

Finally, example (5) shows the context in which the ambiguous word *pas*, meaning both *no/not* and *footstep*, has the non-negation meaning. In this example, *pas* is part of the idiomatic expression **faire les cent pas** *(pacing, walking around)*. Another ambiguity is related to the adverb **plus** meaning either **more** or, in conjunction with **ne**, **no more**.

2.2 SPECULATION

The expression of speculation can be even more complex than negation. Indeed, speculation can be triggered by many specific sequences of words. We describe several of them below.

1. *En effet, l'arrêt du traitement antituberculeux en soi <u>pourrait</u>* **permettre un rétablissement de la fonction normale des héptocytes** *en éliminant la source de l'atteinte hépatique.* (Indeed, stopping TB treatment per se could restore normal hepatocyte function by eliminating the source of liver damage.)

2. **Le bénéfice de l'association lénalomide + R-CHOP au rapport par rapport au R-CHOP** <u>*reste à démontrer*</u>. (The benefit of the combination of lenalomide + R-CHOP compared with R-CHOP remains to be demonstrated.)

3. *Elle <u>aurait</u>* **eu la pose à 2 reprises d'un dispositif intrautérin** *(DIU).* (She would have had the pose of an intrauterine device (IUD) twice.)

Example (1) shows a typical occurrence of *pourrait (could)*. As in English, where *can, could, may, etc.* express speculation, in French, *pouvoir* can occur in many forms.

Example (2) shows the effect of *reste à (remains to)* combined with an infinitive verb, here *démontrer*. Other infinitive verbs, such as *expliquer (to explain)* or *vérifier (to verify)*, associated with *reste à* trigger speculation.

Example (3) shows how the conditional tense triggers speculation. In English, *would* triggers speculation in this case, which is simpler to detect. Indeed, in French, the conditional tense is expressed via suffixes (*-ais, -ais, -ait, -ions, -iez, -aient*), which makes the detection harder, especially for supervised learning techniques.

3 RELATED WORK

We present several corpora and methods that have been proposed in the existing work to tackle the tasks of speculation and negation detection.

3.1 DATA

In the recent years, several specialized corpora in English have been annotated with speculation and negation, which has resulted in models for their automatic detection. These corpora can be divided into two categories: (1) corpora annotated with cues and scopes, such as Bioscope (Vincze et al., 2008) or **SEM-2012*, and (2) corpora focusing on concepts and named entities, such as I2B2 and Mipacq. We briefly describe these corpora. The Bioscope corpus (Vincze et al., 2008) contains reports of radiological examinations, scientific articles, and abstracts from biomedical articles. Each sentence and each negation and speculation cue/scope pair receives unique identifier. The *SEM-2012 corpus (Morante and Blanco, 2012) consists of a Sherlock Holmes novel and three other short stories written by Sir Arthur Conan Doyle. It contains 5,520 sentences, among which 1,227 sentences are negated. Each occurrence of the negation, the cue and its scope are annotated, as well as the focus of the negation if relevant. In this corpus, cues and scopes can be discontinuous. The I2B2/VA-2010 challenge (Uzuner et al., 2011) featured several tasks using US clinical records. One task aimed the detection of statements and of their

scope. Medical concepts had to be associated with the corresponding statement: *present*, *absent*, *possible*, *conditional*, *hypothetical* or *not associated with the patient*. Mipacq (Albright et al., 2013) is another corpus with clinical data in English annotated with syntactic and semantic labels. Each detected UMLS entity has two attribute locations: *negation* (*true* or *false*) and *status* (*none*, *possible*, *HistoryOf* or *FamilyHistoryOf*).

3.2 EXPERT SYSTEMS

Among the rule based systems dedicated to the negation detection, *NegEx* (Chapman et al., 2001) pioneered the area. It uses regular expressions to detect the cues and to identify medical terms in their scope. It was later adapted to various languages including French (Deléger and Grouin, 2012). *ConText* (Harkema et al., 2009), derived from *NegEx*, covers more objectives: negation, temporality, and the subject concerned by this information in the clinical texts. It has been adapted to French (Abdaoui et al., 2017). In another work, medical concepts may receive additional labels (positive, negative or uncertain) Elkin et al. (2005a). Özgür and Radev (2009); Øvrelid et al. (2010); Kilicoglu and Bergler (2010) exploit lexical, grammatical and syntactic information to detect speculation and its scope. ScopeFinder (Apostolova et al., 2011) detects the scope of negation and speculation with rules built automatically from BioScope (lexico-syntactic patterns extraction). *NegBio* (Peng et al., 2018) detects both negation and speculation in radiology reports with rules based on universal dependency graphs.

3.3 SUPERVISED LEARNING

To our knowledge, Light et al. (2004) is the first work to include supervised learning for speculation detection. It relies on SVM to select speculative sentences in MEDLINE abstracts. Tang et al. (2010) proposes a cascade method based on CRF and SVM classifiers to detect speculation cues and another CRF classifier to identify their scopes. Velldal et al. (2012) proposes a SVM-based cue detection system, trained on simple n-grams features computed on the local lexical context (words and lemmas). This system offers a hybrid detection of the scope, which combines expert rules, operating on syntactic dependency trees, with a ranking SVM that learns a discriminative ranking function over nodes in constituent trees. It was further improved by Read et al. (2012) and is used

	French clin. trials	French clin. cases
Documents	–	200
Sentences	6,547	3,811
Tokens	150,084	87,487
Vocabulary (types)	7,880	10,500
Negative sentences	1,025	804
Speculative sentences	630	226

Table 1: Statistics on the two French corpora

as a fall-back by Packard et al. (2014) when the main MRS (minimal recursion semantics) Crawler cannot parse the sentence. Qian et al. (2016) addresses the scope detection with an approach based on a convolutional neural network which extracts features from various syntactic paths between the cues and the candidate tokens in constituency and dependency parsed trees. Fancellu et al. (2016) uses neural networks to solve the problem of negation scope detection. One approach uses Feed-forward neural network, while the other, which appears to be more efficient for the task, uses a bidirectional Long Short-Term Memory (BiLSTM) neural network. Given the results from the latter approach, it inspired our work.

4 FRENCH MEDICAL CORPORA

We manually annotated two corpora from the biomedical field. Table 1 presents some statistics on these corpora: the number of words, the variety of the vocabulary, the number of sentences, the number of negative sentences with one or more negations. The Inter Annotator Agreement (IAA) on negation annotation is high (Cohen's κ=0.8461).

4.1 ESSAI: FRENCH CORPUS with CLINICAL TRIALS

One corpus contains clinical trial protocols in French. They were mainly obtained from the National Cancer Institute registry[1]. The typical protocol consists of two parts: the summary of the trial, which indicates the purpose of the trial and the methods applied; and a detailed description of the trial with the inclusion and exclusion criteria.

---

[1] https://www.e-cancer.fr

Form	Lemma	POS	Cue	scope
Pas	pas	ADV	pas	_
de	de	PRP	_	de
dyspnée	dyspnée	NOM	_	dyspnée
.	.	SENT	_	_

Table 2: Excerpt from the CAS corpus. The columns contain linguistic (lemmas, POS-tag) and reference (cue, scope) information.

4.2 CAS: FRENCH CORPUS with CLINICAL CASES

This corpus contains clinical cases published in scientific literature and training material. They are published in different journals from French-speaking countries (France, Belgium, Switzerland, Canada, African countries, tropical countries) and are related to various medical specialties (cardiology, urology, oncology...). The purpose of clinical cases is to describe clinical situations of patients. Hence, their content is close to the content of clinical narratives (description of diagnoses, treatments or procedures, evolution, family history, expected audience, etc.). In clinical cases, the negation is frequently used for describing the patient signs, symptoms, and diagnosis. Speculation is present as well but less frequently.

4.3 ANNOTATION LAYERS

These two corpora are Part-of-Speech tagged and lemmatized with TreeTagger (Schmid, 1994). For the creation of the reference data necessary for machine learning, both corpora were annotated manually to mark up negation and speculation cues and their scope. However, the 200 annotated clinical cases did not include enough examples of speculation for a machine learning models to train properly. Therefore, speculation detection is either trained and tested with ESSAI alone or with ESSAI and CAS. Table 2 presents an annotated sentence from CAS: *No dyspnea.*. The corpora also includes two additional columns (sentence number and token position).

5 METHODOLOGY

As indicated on Figure 1, our methods rely on specifically trained word vectors and supervised learning techniques (BiLSTM and CRF). The objective is to classify each word as being part or not of the negation/speculation cue and/or scope.
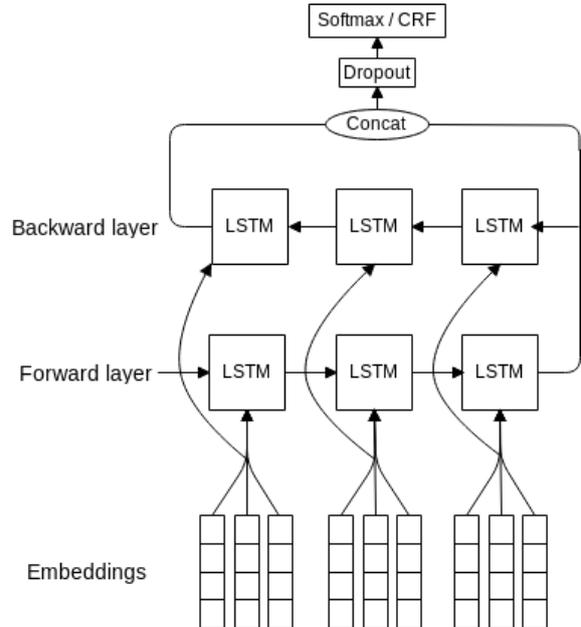

Figure 1: Our bidirectional RNN uses LSTM cells with either a softmax or CRF output layer. Features are either words/lemmas/PoS for cue detection or words/PoS/Cue information for scope detection.

5.1 WORD VECTOR REPRESENTATIONS

In the recent years, several models have been introduced to generate vector representations of words helping machine learning approaches to better capture their semantics. The models used in the negation/speculation detection task are the following ones.

word2vec (Mikolov et al., 2013) is a predictive model to learn word embeddings from plain text. The embeddings can be calculated using two model architectures: the continuous bag-of-words (CBOW) and *Skip-Gram* (SG) models. In this work, we use use the SG model; it treats each context-target pair as new observation, which is suitable for large datasets.

fastText (Bojanowski et al., 2017) addresses the Word2vec's main issue: the words, which do not occur in the vocabulary, cannot be represented. Hence, this algorithm uses subword information: each word is represented as a bag of all possible character n-grams it contains. The word is padded using a set of unique symbols which helps singling out prefixes and suffixes. The full sequence is added to the bag of n-grams as well. The vector now denotes every char n-gram and the word vector is the sum of its char n-gram vectors. Since the char n-gram representations across words are

often shared, rare words can also get reliable representations.

These two word embedding models are trained using the *Skip-Gram* algorithm, 100 dimensions, a window of 5 words before and after each word, a minimum count of five occurrences for each word and negative sampling. The training data are composed of the French Wikipedia articles and biomedical data. The latter includes the ESSAI and CAS corpora, the French Medical Corpus from CRTT[2] and the Corpus QUAERO Médical du français[3] (Névéol et al., 2014). These models are trained using the Gensim[4] (Rehurek and Sojka, 2010) python library.

5.2 RECURRENT NEURAL NETWORK

Recurrent neural network takes into account the previously seen data in addition to the currently seen data. This is implemented with loops in the architecture of the network, which allows the information to persist in memory. Among the RNNs, long short-term memory networks (LSTM) (Hochreiter and Schmidhuber, 1997) are the most efficient for the learning of long-term dependencies and are therefore more suitable to solve the problem of discontinuous scope, which is typical for the negation. LSTM cells are also more efficient at retaining useful information during back-propagation.

We use a bidirectional LSTM, which operates forward and backward on the sentence, to detect cues and scopes. The backward pass is relevant for the scope detection because the scope may be before or after the cue. Prediction is computed by either a *softmax* or a CRF (suitable for the sequence labeling) output layer. We use embeddings of dimension $k = 100$ and a dimensionality of the output space of 400 units per layer (backward/forward) with 0.5 dropout. 50 epochs achieve the highest F_1 score on the validation sets.

5.3 CONDITIONAL RANDOM FIELDS

Conditional random fields (CRFs) (Lafferty et al., 2001) are statistical methods used to label word sequences. By training a model on appropriate features and labels to be predicted, the CRFs generally obtain good results with much lower training time than neural networks.

[2] https://bit.ly/2LOJfEW
[3] https://quaerofrenchmed.limsi.fr/
[4] https://radimrehurek.com/gensim/

We performed the gradient descent using the L-BFGS (Limited-memory BFGS) method with 0.1 L1 penalty and 0.01 L2 penalty. We only experiment with CRFs for the cue detection task, in comparison with BiLSTM-CRF.

5.4 EVALUATING LABELING SYSTEMS

We use standard evaluation measures: precision P, recall R, and F_1 score. The scope detection is evaluated in two ways: (1) on individual scope tokens which is the standard evaluation, and (2) on exact scopes to assess more strictly how efficient our models are. For the latter, we use the available evaluation script[5]. Each corpus is randomly segmented into the training set (80%, 20% for validation), and the test set (20%).

6 CUE DETECTION

The speculation and negation cue detection is the first step of the task. To tackle this problem, we experiment with two supervised learning approaches. First, we train a CRF using several features (words, lemmas and POS-tags) with empirically defined window over features. Our second approach uses a BiLSTM with a CRF output layer, which is trained on the same features. We did not use any pre-trained embeddings for this task.

Table 3 presents the results obtained with our approaches on the ESSAI and CAS corpora. We can see that cue detection shows high evaluation values: 93.92 to 97.21 F-measure for negation cues, and 86.88 to 91.30 F-measure for speculation cues. Although there is little room for improvement on negation cue detection, indeed 10k-fold cross-validation with our CRF reaches more than 95 F-measure on both corpora, speculation cue detection would benefit from more training examples. Indeed, the potential number of cues and the numerous contexts in which they appear and do or do not express speculation makes them harder to detect and will require more annotated examples. For both negation and speculation cue detection, our CRF is slightly more efficient than the BiLSTM-CRF when the CAS corpus is involved, which indicates that the CAS corpus contains less complex examples than ESSAI.

7 SCOPE DETECTION

In all the scope detection experiments proposed, we only train the neural networks on nega-

[5] https://github.com/ffancellu/NegNN

	System	Corpus	window size	P	R	F_1
Negation	CRF	ESSAI	(4)	96.05	91.89	93.92
	BiLSTM-CRF		None	95.10	94.58	**94.84**
	CRF	CAS	(4)	97.05	97.37	**97.21**
	BiLSTM-CRF		None	97.02	97.02	97.02
Speculation	CRF	ESSAI	(4)	91.43	82.76	86.88
	BiLSTM-CRF		None	91	83.84	**87.27**
	CRF	ESSAI+CAS	(4)	93.93	88.82	**91.30**
	BiLSTM-CRF		None	91.22	87.92	89.54

Table 3: Precision, Recall and F_1-score for the cue detection task on the two corpora (bold: best scores).

	Corpus	System	WE	Scope tokens			Exact scope match		
				P	R	F_1	P	R	F_1
Negation	ESSAI	BiLSTM-S	WI	86.21	82.85	**84.50**	100	55.61	71.47
			W2V	83.54	83.68	83.61	100	56.59	72.27
			FT	80.79	86.41	83.51	100	56.59	72.27
		BiLSTM-CRF	WI	84.65	84.09	84.37	100	59.51	74.62
			W2V	83.86	83.10	83.48	100	61.95	**76.51**
			FT	82.38	84.84	83.59	100	59.51	74.61
	CAS	BiLSTM-S	WI	93.72	87.30	90.40	100	73.21	84.54
			W2V	93.03	88.69	**90.81**	100	75.59	86.10
			FT	91.50	88.69	90.08	100	72.02	83.74
		BiLSTM-CRF	WI	91.87	88.59	90.20	100	68.45	81.27
			W2V	91.47	88.29	89.85	100	76.19	86.49
			FT	94.82	87.10	90.80	100	78.57	**88.00**
Speculation	ESSAI	BiLSTM-S	WI	89.27	82.14	85.56	100	52.76	69.07
			W2V	88.61	84.42	**86.47**	100	56.69	72.36
			FT	85.84	84.58	85.21	100	58.27	**73.63**
		BiLSTM-CRF	WI	89.77	83.03	86.27	100	51.97	68.39
			W2V	87.85	83.77	85.76	100	55.91	71.72
			FT	91.04	79.61	84.94	100	57.48	73.00
	ESSAI+CAS	BiLSTM-S	WI	88.90	83.94	86.35	100	57.56	73.06
			W2V	86.20	85.19	85.69	100	58.14	73.53
			FT	85.15	87.46	86.29	100	59.30	74.45
		BiLSTM-CRF	WI	89.49	81.16	85.12	100	56.98	72.59
			W2V	88.48	85.04	**86.73**	100	65.12	**78.87**
			FT	89.15	83.14	86.04	100	62.79	77.14

Table 4: Precision, Recall and F_1-score for the scope detection task (bold: best scores).

System		Train	Test	Scope tokens			Exact scope match		
				P	R	F_1	P	R	F_1
BiLSTM-CRF	Negation	ESSAI	CAS	76.73	76.36	**76.54**	100	36.08	**53.03**
		CAS	ESSAI	82.36	55.23	66.12	100	28.20	43.99
	Speculation	ESSAI	CAS	76.46	68.03	**72.00**	100	45.37	**62.42**
		CAS	ESSAI	72.43	65.05	68.54	100	30.79	47.09

Table 5: Cross-corpora Precision, Recall and F_1-score for the scope detection task (bold: best scores).

tive/speculative sentences. The base system takes as input an instance $I(w, c, t)$, where each word is represented by: w vector (*word-embedding*), c vector (*cue-embedding*), indicating if the word is a cue or not, t vector (*POS-tag-embedding*). Preliminary tests showed that adding lemmas as features only decreases the F-measure. For each system, we use the same empirically defined hyperparameters given before. During training, embeddings weights are updated.

Table 4 indicate the results obtained for the scope detection task. One can see that it is easier to predict the scope of negation cues (up to 90.81 F-measure) than of speculation cues (up to 86.73 F-measure). Results show that using pre-trained embeddings improves F-measures for exact scope detection by up to 6 points for both negation and speculation. Moreover, the CRF output layer either outperforms the *softmax* layer or reaches an equivalent F-measure for exact scope detection. In another experiment, we trained the models on one corpus and tested them on the other (Table 5): the models trained on ESSAI are more efficient, and the negation and speculation structures are more stable in the CAS corpus. However, even though CAS (speculation) was only trained on 226 examples, the model still shows decent results in scope tokens detection.

7.1 ERROR ANALYSIS

An analysis of the results makes it possible to isolate frequent types of errors. In the following examples, the speculation and negation cues are underlined, the scope is between brackets, while the segments in bold correspond to predictions errors.

7.1.1 NEGATION

In the first example, the prediction fails at labeling **rénale** *(renal)*. In the majority of cases in the reference data, the scopes associated to the cue **sans** often only include one token, which may be causing this error that impacts recall:

- GOLD: *Le patient sortira du service de réanimation guéri et <u>sans</u> [insuffisance rénale] après huit jours de prise en charge et cinq séances d'hémodialyse.*

- PRED: *Le patient sortira du service de réanimation guéri et <u>sans</u> [insuffisance] **rénale** après huit jours de prise en charge et cinq séances d'hémodialyse.*

  (The patient will be discharged from the intensive care unit without renal failure after eight days of management and five hemodialysis sessions.)

The second example illustrates the error that impacts precision. Here, the model wrongly predicts that all tokens in the sentence are within the scope. In the reference data, the cue **aucun** *(any, no)* often occurs at the beginning of sentences, and in sentences with many instances of negation. The model, mostly trained on this kind of examples, may try to reproduce these structures which causes bad prediction in some cases.

- GOLD: *Les colorations spéciales (PAS, coloration de Ziehl-Neelsen, coloration de Grocott) <u>ne</u> [mettaient en évidence] <u>aucun</u> [agent pathogène].*

- PRED: ***[Les colorations spéciales (PAS, coloration de Ziehl-Neelsen, coloration de Grocott]) <u>ne</u> [mettaient en évidence] <u>aucun</u> [agent pathogène].***

  (Special stains (PAS, Ziehl-Neelsen stain, Grocott stain) showed no pathogens.)

In the third example, the error impacts both precision and recall. In this example, we have two instances of negation with the same cues: **n...pas**. Usually, its scope follows, however, in the first instance it precedes. As we do not have many examples of this kind to train on, the model fails to correctly label the sequence. In the second negation instance, the scope may be shorter than usual, which impacts precision.

- GOLD: *[Le retrait du matériel d'ostéosynthèse incriminé] <u>n'</u>[est] <u>pas</u> [systématique], ce qui explique qu'il <u>n'</u>[ait] <u>pas</u> [été proposé] à notre patient asymptomatique.*

- PRED: ***Le retrait du matériel d'ostéosynthèse incriminé** <u>n'</u>[est] <u>pas</u> [systématique], ce qui explique qu'il <u>n'</u>[ait] <u>pas</u> [été proposé **à notre patient asymptomatique**].*

  (The removal of the implicated osteosynthesis material is not systematic, which explains why it has not been proposed to our asymptomatic patient.)

7.1.2 SPECULATION

In the first example, the scope of the speculation has been predicted up to the end of the preposition while in the reference data, the scope covers the verbal group. This typically impacts precision.

- GOLD: *Ce médicament n'a pas la toxicité de la chimiothérapie, mais entraine une disparition des lymphocytes B normaux pendant plusieurs mois, ce qui* pourrait *[favoriser la survenue d'infections graves] car ces lymphocytes participent à la défense immunitaire.*

- PRED: *Ce médicament n'a pas la toxicité de la chimiothérapie, mais entraine une disparition des lymphocytes B normaux pendant plusieurs mois, ce qui* pourrait *[favoriser la survenue d'infections graves **car ces lymphocytes participent à la défense immunitaire**]*

  (*This drug does not have the toxicity of chemotherapy, but causes the disappearance of normal B lymphocytes for several months, which could increase the occurrence of serious infections because these lymphocytes participate in the immune defense.*)

However, most of our errors impact recall, like in the following example.

- GOLD: *Elles possèdent les caractéristiques de la tumeur et* pourraient *[permettre à l'avenir de faire le diagnostic de tumeur sans biopsie ainsi que de suivre l'évolution de la tumeur traitée], les CTC disparaissant quand le traitement fonctionne.*

- PRED: *Elles possèdent les caractéristiques de la tumeur et* pourraient *[permettre à l'avenir de faire le diagnostic de tumeur] **sans biopsie ainsi que de suivre l'évolution de la tumeur traitée**, les CTC disparaissant quand le traitement fonctionne.*

  (*It has the characteristics of the tumor and could in the future be used to diagnose the tumor without biopsy and to follow-up the evolution of the treated tumor, the CTC disappearing when the treatment is efficient.*)

Several examples show errors where both precision and recall are impacted. Usually, the reason is that they are rare cases in our corpora where the scope is reversed. For instance, in the example below, most of the scope of *devrait* precedes it, while in the reference data, its scope follows this cue.

- GOLD: *L'objectif de cet essai est d'évaluer [la diminution des complications postopératoires de l'oesophagectomie qui]* devrait *[être obtenue] en réalisant une partie de l'intervention sous coelioscopie, chez des patients ayant un cancer de l'oesophage résécable.*

- PRED: *L'objectif de cet essai est d'évaluer la diminution des complications postopératoires de l'oesophagectomie qui* devrait *[être obtenue **en réalisant une partie de l'intervention sous coelioscopie**], chez des patients ayant un cancer de l'oesophage résécable.*

  (*The objective of this trial is to evaluate the decrease in postoperative complications of esophagectomy that should be achieved by performing part of the procedure under laparoscopy, for patients with resectable esophageal cancer.*)

8 CONCLUSION and FUTURE WORK

The interest for the automatic detection of speculation and negation in English with supervised machine learning has increased in the recent years. Yet, the lack of data for other languages and for specialized domains hampers the further development of such approaches. In our work, we presented new bodies of biomedical data in French annotated with negation and speculation (cues and their scope). Prior to the dissemination to the research community, the French clinical trial protocols corpus will be finalized through the integration of new data and the computation of the inter-annotator agreement. The French CAS corpus will be distributed as more clinical cases are manually annotated, as we need more speculative sentences to train supervised learning models on this dataset. Another contribution of our work is the study of different types of word vector representations and recurrent neural networks for the detection of negation and speculation. There has not been much work of this type on French corpora, especially for the biomedical domain which contains specific negation and speculation phenomena. We showed that a *CRF* layer yields better performance than *softmax* on exact scope match. Finally, the models have been applied in a cross-corpus context. Besides, we plan to improve our neural network performance by providing richer feature. In particular, recent embedding techniques, such as BERT or ELMO (Devlin et al., 2018; Peters et al., 2018) may provide more accurate representation of the sentences. Moreover, in order to provide more accurate features, we plan to move from TreeTagger, which makes a substantial number of mistakes on our datasets, to a POS tagger/lemmatizer dedicated to French biomedical texts. Syntactic parsing of sentences may also provide useful features for the detection of scope.

References

Amine Abdaoui, Andon Tchechmedjiev, William Digan, Sandra Bringay, and Clement Jonquet. 2017. French context: Détecter la négation, la temporalité et le sujet dans les textes cliniques français. In *4ème Symposium sur l'Ingénierie de l'Information Médicale, SIIM'17*.

Daniel Albright, Arrick Lanfranchi, Anwen Fredriksen, William F Styler IV, Colin Warner, Jena D Hwang, Jinho D Choi, Dmitriy Dligach, Rodney D Nielsen, James Martin, et al. 2013. Towards comprehensive syntactic and semantic annotations of the clinical narrative. *Journal of the American Medical Informatics Association* 20(5):922–930.

Emilia Apostolova, Noriko Tomuro, and Dina Demner-Fushman. 2011. Automatic extraction of lexico-syntactic patterns for detection of negation and speculation scopes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, pages 283–287.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5:135–146.

W Chapman, W Bridewell, P Hanbury, GF Cooper, and BG Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* 34(5):301–10.

Louise Deléger and Cyril Grouin. 2012. Detecting negation of medical problems in french clinical notes. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*.

JC Denny and JF Peterson. 2007. Identifying qt prolongation from ecg impressions using natural language processing and negation detection. In *Medinfo*. pages 1283–8.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Peter L. Elkin, Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R. Bergstrom, and Dietlind L. Wahner-Roedler. 2005a. A controlled trial of automated classification of negation from clinical notes. *BMC medical informatics and decision making* 5. https://doi.org/10.1186/1472-6947-5-13.

PL Elkin, SH Brown, BA Bauer, CS Husser, W Carruth, LR Bergstrom, and DL Wahner-Roedler. 2005b. A controlled trial of automated classification of negation from clinical notes. *BMC Med Inform Decis Mak.* 5(13).

Federico Fancellu, Adam Lopez, and Bonnie Webber. 2016. Neural networks for negation scope detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. volume 1.

S Gindl, K Kaiser, and S Miksch. 2008. Syntactical negation detection in clinical practice guidelines. In *Stud Health Technol Inform*. pages 187–92.

Henk Harkema, John N Dowling, Tyler Thornblade, and Wendy W Chapman. 2009. Context: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of biomedical informatics* 42(5):839–851.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9(8). https://doi.org/10.1162/neco.1997.9.8.1735.

Halil Kilicoglu and Sabine Bergler. 2010. A high-precision approach to detecting hedges and their scopes. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*. Association for Computational Linguistics, pages 70–77.

John Lafferty, Andrew McCallum, Fernando Pereira, 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.

Marc Light, Xin Ying Qiu, and Padmini Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *HLT-NAACL 2004 Workshop: Linking Biological Literature, Ontologies and Databases*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*.

Roser Morante and Eduardo Blanco. 2012. * sem 2012 shared task: Resolving the scope and focus of negation. In * *SEM 2012: 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 (SemEval 2012)*. volume 1, pages 265–274.

Aurélie Névéol, Cyril Grouin, Jeremy Leixa, Sophie Rosset, and Pierre Zweigenbaum. 2014. The quaero french medical corpus: A ressource for medical entity recognition and normalization. In *In Proc BioTextM, Reykjavik*. Citeseer.

Lilja Øvrelid, Erik Velldal, and Stephan Oepen. 2010. Syntactic scope resolution in uncertainty analysis. In *Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics, pages 1379–1387.

Arzucan Özgür and Dragomir R Radev. 2009. Detecting speculations and their scopes in scientific text. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*. Association for Computational Linguistics, pages 1398–1407.

Woodley Packard, Emily M Bender, Jonathon Read, Stephan Oepen, and Rebecca Dridan. 2014. Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 69–78.

Yifan Peng, Xiaosong Wang, Le Lu, Mohammadhadi Bagheri, Ronald Summers, and Zhiyong Lu. 2018. Negbio: a high-performance tool for negation and uncertainty detection in radiology reports. *AMIA 2018 Informatics Summit* .

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365* .

Zhong Qian, Peifeng Li, Qiaoming Zhu, Guodong Zhou, Zhunchen Luo, and Wei Luo. 2016. Speculation and negation scope detection via convolutional neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. pages 815–825.

Jonathon Read, Erik Velldal, Lilja Øvrelid, and Stephan Oepen. 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, pages 310–318.

Radim Rehurek and Petr Sojka. 2010. Software framework for topic modelling with large corpora. In *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Citeseer.

Helmut Schmid. 1994. Probabilistic part-ofispeech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.*.

Buzhou Tang, Xiaolong Wang, Xuan Wang, Bo Yuan, and Shixi Fan. 2010. A cascade method for detecting hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning— Shared Task*. Association for Computational Linguistics, pages 13–17.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association* 18(5):552–556.

Erik Velldal, Lilja Øvrelid, Jonathon Read, and Stephan Oepen. 2012. Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics* 38(2).

Veronika Vincze, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* 9.