Procesamiento del Lenguaje Natural, Revista nº 62, marzo de 2019, pp. 37-44 recibido 22-11-2018 revisado 07-01-2019 aceptado 07-02-2019

Deep learning approach for negation trigger and scope recognition

Experimentación basada en deep learning para el reconocimiento del alcance y disparadores de la negación

Hermenegildo Fabregat[1], Lourdes Araujo[1,2], Juan Martinez-Romo[1,2]
[1]NLP & IR Group, Dpto. Lenguajes y Sistemas Informáticos,
Universidad Nacional de Educación a Distancia (UNED),
Juan del Rosal 16, Madrid 28040, Spain
[2]IMIENS: Instituto Mixto de Investigación, Escuela Nacional de Sanidad,
Monforte de Lemos 5, Madrid 28019, Spain
gildo.fabregat@lsi.uned.es, lurdes@lsi.uned.es, juaner@lsi.uned.es

Abstract: The automatic detection of negation elements is an active area of study due to its high impact on several natural language processing tasks. This article presents a system based on deep learning and a non-language dependent architecture for the automatic detection of both, triggers and scopes of negation for English and Spanish. The presented system obtains for English comparable results with those obtained in recent works by more complex systems. For Spanish, the results obtained in the detection of negation triggers are remarkable. The results for the scope recognition are similar to those obtained for English.
Keywords: Negation scope, negation triggers detection, deep learning

Resumen: La detección automática de los distintos elementos de la negación es un frecuente tema de estudio debido a su alto impacto en diversas tareas de procesamiento de lenguaje natural. Este articulo presenta un sistema basado en *deep learning* y de arquitectura no dependiente del idioma para la detección automática tanto de disparadores como del alcance de la negación para inglés y español. El sistema presentado obtiene para ingles resultados comparables a los obtenidos en recientes trabajos por sistemas más complejos. Para español destacan los resultados obtenidos en la detección de claves de negación. Por último, los resultados para el reconocimiento del alcance de la negación, son similares a los obtenidos en inglés.
Palabras clave: Detección de negación, disparadores de la negación, *deep learning*

1 Introduction

The study of negation is an active research topic due to its effects and importance within the different challenges of the natural language processing (NLP) research area. Although many tasks and domains are affected by negation, its study in the biomedical domain is of particular relevance. Chapman et al. (2001a) shows the importance of consider possible negated phrases during the analysis of electronic health records (EHR), documents in which much of the information contained is expressed in a negated way. The contributions of negation in tasks such as sentiment analysis and relationship extraction stand out due to the performance improvements obtained after considering it. Councill, McDonald, and Velikovich (2010) examine the achieved improvements after including the study of negation in the task of sentiment analysis in online product reviews. The authors take into account the negation during the evaluation of the score of each term in a sentence, modifying the score sign if the term is part of a negation. On the other hand, Chowdhury and Lavelli (2013) highlight the significant performance improvements obtained in the detection of drug-drug relationships after considering negation.

There are many possible elements of study concerning negation. This paper deals with both, identification of negation triggers and the delimitation of negation scope. The detection of negation triggers, can be considered a basic task in the study of negation.

ISSN 1135-5948. DOI 10.26342/2019-62-4

It refers to the identification of expressions that work as markers of negation. The identification of negation scope refers to finding segments of a sentence that are part of one or more negations. This study presents the experimentation carried out using a deep learning approach for the study of the negation scope for both English and Spanish. The following sections present a study of the state of the art for the proposed work (Section 2), a summarization of the proposed approach including a brief description of the explored datasets (Section 3) and finally, a review of the obtained results (Section 4) and the conclusions reached after the experimentation (Section 5).

2 Background

Chapman et al. (2001b) presented an algorithm called NegEx based on the use of regular expressions for the detection of negation in clinical documents. Tools such as cTAKES (Savova et al., 2010), designed for processing medical documents in free text format, use NegEx for the treatment of negation. Nowadays, NegEx is considered a baseline in many of the works dealing with the automatic study of negation. Although this algorithm shows a high *performance*, an important issue to take into account, is the low precision obtained evaluating sentences where the term "no" appears. Goldin and Chapman (2003) extend the study of this case developing a set of experiments in order to compare the results obtained by NegEx with those obtained by a set of different machine learning algorithms. Among them Naive Bayes (NB) and Decision Tree (DT), achieved better results than NegEx. Although NegEx has been designed for English, recent works such as Chapman et al. (2013) and Skeppstedt (2011) have studied its use for other languages such as French, German, Swedish and Spanish. Cotik et al. (2016) show the results obtained by an adaptation of NegEx to Spanish. Their results are better than the use of dictionaries and comparable with those obtained by a system of rules based on patterns of PoS Tagging.

There are many systems evaluated using the Bioscope corpus (Vincze et al., 2008) which is a linguistic resource containing annotations about negation and speculation in the biomedical domain. Fancellu, Lopez, and Webber (2016) shows the good performance of Bidirectional Long Short-Term Memory (bi-LSTM) based models for the identification of multi-term expressions such as "by no means of" and "no longer". Fancellu et al. (2017) extend the study to other domains and languages (Chinese), presenting, among others, results for the Bioscope corpus and for the SFU corpus (Konstantinova et al., 2012). The study shows a comparison of the results obtained by a bidirectional long short-term memory (bi-LSTM) based model with some state of the art systems. For different domains, Li and Lu (2018) deal with both, the detection of negation triggers and the recognition of the scope of negation, using different kinds of conditional random fields (CRF), *linear CRF* (Lafferty, McCallum, and Pereira, 2001), *semi-CRF* (Sarawagi and Cohen, 2005) and *latent variable CRF*. Taking into account the obtained results, one of the conclusions the authors reached was the good performance of this kind of algorithms in sequence labeling tasks, having obtained remarkable results even after extending the evaluation to languages such as Chinese.

3 Materials and methods

The aim of this work has been to look for a simple deep learning architecture for negation detection valid for different languages. We consider both, the detection of negation triggers and the recognition of the scopes. We have built a simple architecture in which input data have proven to be useful. In the following sections, we describe the proposed model and details of the corpora we use for the evaluation. We also explain the details of both, text pre-processing and the system output post-processing processes.

3.1 Features

The proposed model uses the following features:

Words. For Spanish, we have used the embedding vectors presented by Cardellino (2016). They are vectors of 300 dimensions and collect a total of 1000653 unique tokens. They were generated using the word2vec algorithm by means of multiple repositories of information in Spanish for training. For English, the word embedding presented by Pyysalo et al. (2013) were used. They are 200-dimensional vectors and collect about twenty-four million unique tokens. This resource was generated using word2vec and taking as a source of information several wikipedia dumps and some biomedical repositories such as PubMed and PMC.

PoS-Tagging. We have used FreeLing PoS-tagger (Padró and Stanilovsky, 2012) for Spanish and the maximum entropy PoS tagger implemented in NLTK (Bird and Loper, 2004) for English.

Casing. Another feature used is a matrix for the representation of word casing information. Each token has been represented with the corresponding index of the matrix embedding. The casing embedding matrix is a hot-one encoding matrix of size 14. This feature provides additional support to the model by representing each token in a summary category.

Chars. We use character embeddings in order to collect expressions not included in the pre-trained word embeddings vocabulary, taking into account that the vocabulary of health records is not standard and that the vocabulary of product reviews may contain spelling errors. This vectorial representation allows to represent the information contained in both prefixes and suffixes.

Both PoS-tagging, casing and character embedding models have been implemented using three Keras Embedding Layers initialized using a random uniform distribution. In Section 4 the performance improvement obtained after considering each of the features is shown.

3.2 Proposed model

Figure 1 shows the architecture of the proposed model. This architecture consists of a character-level processing module (Santos and Zadrozny, 2014) and a word-level processing module (Fabregat, Araujo, and Martinez-Romo, 2018). Character-level processing is essentially a transformation of the characters of each word into character embeddings and on the concatenation of the most important features obtained by the application of a convolutional layer. The result of this process is concatenated to the input of the other part of the model.

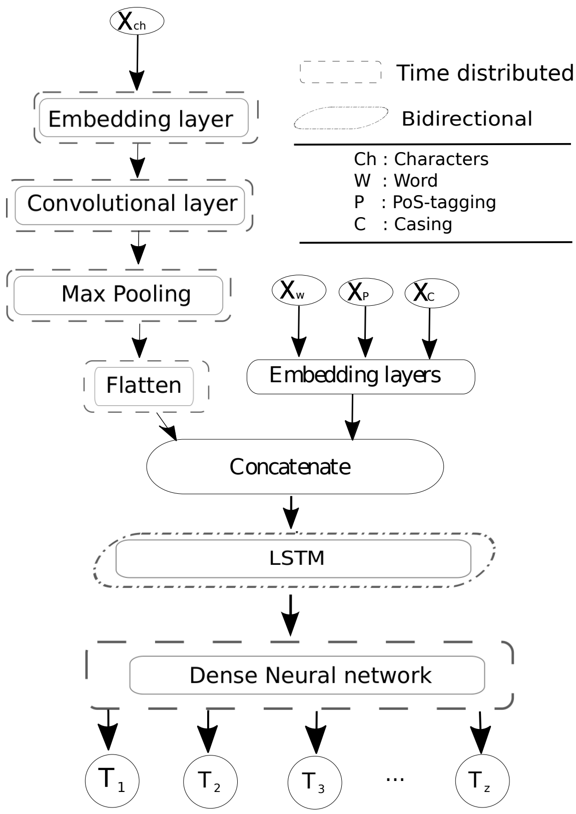

Figure 1: Architecture of the proposed model, where $X_C h$ and X_W (Ch: Characters, W: Raw word) are the encoded word inputs and X_P and X_C are the encoded inputs representing the PoS-tagging and casing information. Bi-LSTM inputs (Y_x) are the concatenated embedded features for each word. In the output layer, T_x represents the assigned tag.

The second part of the model consists of a Bidirectional Long Short-Term Memory (Bi-LSTM) network connected to a neural network. On the one hand, the Bi-LSTM is responsible of processing each part of the concatenated features in order to obtain positional/semantic relationships between terms of a sentence. On the other hand, the neural network is responsible of generating the correct classification sequence. This neural network has softmax as an activation function and processes the output of each part of the sequence returned by the previous LSTM.

Taking into account the obtained results in Fabregat, Araujo, and Martinez-Romo (2018), which deals with a problem of recognition of negation triggers by means of an approach based on deep learning, the adjustment of parameters has been made according to the configuration of the model presented in that paper. For the configuration of the additional layers that form the character-level processing, we have used of a configuration focused on extracting syntactic/semantic features from immediately adjacent characters. The final configuration is as follows:

- Convolutional layer (kernel_size / filter): 3 / 30
- Embeddings dimension (Casing / PoS-tagging / Char): 12 / 50 / 50
- LSTM output dimension: 250
- Dropout: 0.5
- Batch size / Model optimizer: 32 / Adam
- Hidden Dense units (output dimension / activation function): 17 / softmax

The number of neurons in the output layer of the neural network corresponds to the total of classes to be considered in the annotation.

3.3 Pre-processing

During the pre-processing phase, the different datasets are transformed into the BILOU labelling scheme (Ratinov and Roth, 2009). In this annotation scheme the information is represented applying the following map: {I:In - For tokens part of the annotation. O: Out - For tokens outside the annotation. B:Begin - For the first token of each annotation. L:Last - For the last token of each annotation. U:Unique - Those annotations that have a single token.}. This annotation scheme, used in entity recognition tasks, allows the partial overlapping and nesting of one entity within another, a characteristic necessary to represent cases such as two or more negations starting in the same term or a negation included within another. We have carried out this encoding in order to be able to deal with this problem from the perspective of a classification problem. We have used the labelling code to represent both, the scope and the negation triggers separately. These two codifications are combined into a single one by concatenating the labels. For example, if an expression is both, the beginning of a negation and the beginning of a negation triggers, this will be re-labeled with the label "BB". Table 1 shows an example of BILOU annotation format. The first column contains the word and the second column contains the label after joining the scope label and the trigger associated label. The example shows the annotation of both scopes. While the first one spans from the first term "no" up to the term "dinero", the other is nested and spans from the second term "no" up to the term "gusta".

Word	Label	Word	Label
no	BU	no	BU
tendré	IO	me	IO
jamás	IU	gusta	LO
que	IO	por	IO
aceptar	IO	el	IO
un	IO	dinero	LO
trabajo	IO	.	OO
que	IO		

Table 1: SFU Review SP-NEG fragment with tag assignment.

Considering that a negation must has associated a scope, i.e. there are combinations of labels that cannot occur, a total of 17 labels are generated.

3.4 Post-processing

The post-processing phase aims to ensure that the format generated by the model is correct. This format must satisfy the following requirements: Each scope must have at least one associated negation trigger and each annotation must have both, a start and an end label, except in the case of a single token annotation. This phase applies the following rules:

- If a scope does not have at least one negation trigger associated to it, it is not a scope.
  - **- Sentence** Don't you think it's late?
  - **- Proposed labels** BO BO BO BO BO
  - **- Processed labels** OO OO OO OO OO
- If a negation trigger does not have one negation scope associated to it, it is not a negation trigger.
  - **- Sentence** Don't you think it's late?
  - **- Proposed labels** IS OO OO OO OO
  - **- Processed labels** OO OO OO OO OO
- If an annotation starts but does not closes, then it finishes in the last term considered by the system as part of the annotation.
  - **- Sentence** don't buy it.
  - **- Proposed labels** BS IO IO

- **Processed labels** BS IO LO

- If an annotation closes a scope but it is not open, it starts with the first trigger of the phrase detected by the system.
  - **Sentence** don't buy it.
  - **Proposed labels** IS IO LO
  - **Processed labels** BS IO LO

3.5 Corpora

In order to extract conclusions for both languages, English and Spanish, we have selected two corpora with similar annotation guidelines. They are Bioscope corpus and SFU Review SP-NEG corpus (Jiménez-Zafra et al., 2018). Bioscope corpus consists of three parts, electronic health records (EHRs) presented in free text format, full biological articles and abstracts of both scientific and biological articles. The domains included in this corpus present a complex structure, being the most different the domain of EHRs for the use of a free writing style. The subset of abstracts stands out because it contains more negations than the rest and it is the largest subset. On the other hand, SFU Review SP-NEG corpus consists of a collection of 400 comments on cars, hotels, washing machines, books, mobile phones, music, computers and films from the *Ciao.es* website. This corpus presents a mixture of free-writing and formal writing styles.

In both collections, each document has been annotated at token and sentence level with labels related to negation triggers and their linguistic scope. In addition, both collections have used an annotation style without gaps. In summary, using both datasets and taking into account the different writing scenarios of their documents, this work studies the performance of the proposed architecture for free text in Spanish and for both free text and well-structured text in English.

4 Evaluation

This model has been evaluated using 10 fold cross-validation and we have made a study of the improvement obtained considering each attribute of the model. We have used two separate workflows for the model evaluation: one for the evaluation of negation scope detection and a second for the evaluation of negation triggers recognition. For the evaluation of negation scope detection task, we used the percentage of correctly identified scopes (PCS), F1 measure at scope level (F1s) and F1 measure at token level (F1t). F1s measure only considers as false positives those scopes that have been identified but were not found in the gold standard. For the evaluation of triggers, we used precision, recall and F1-measure. In both cases, the evaluation metrics have been used in previous works. The results obtained during this preliminary evaluation suggest that the proposed model is appropriate to deal with the different domains proposed. The experiments carried out show an improvement in performance af-

	Bioscope									SFU SP-NEG		
	Abstracts			Clinical records			Full papers			All categories		
Training Features	PCS	F1s	F1t	PCS	F1s	F1t	PCS	F1s	F1t	PCS	F1s	F1t
W	75.6	84.8	75.52	89.11	94.07	91.28	46.77	59.55	46.17	69.76	82.04	67.32
W+P	81.83	89.9	80.3	94.57	97.12	94.78	49.05	64.79	50.41	72.40	84.05	71.34
W+P+C+Ch	80.52	88.54	80.05	90.03	94.63	91.85	58.99	70.67	58.50	74.29	85.25	72.00

Table 2: Evaluation of each elements considered in the proposed model for the identification of the scope, {W:Words - P:PoS - C: Casing - Ch: Chars}. PCS (percentage of correctly identified scopes), F1 at scope level (F1s) and F1 at token level (F1t) are the metrics analyzed

	Bioscope									SFU SP-NEG		
	Abstracts			Clinical records			Full papers			All categories		
Training Features	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
W	96.03	94.19	95.0	99.11	96.14	97.59	90.99	80.6	84.81	99.04	91.18	94.95
W+P	99.59	95.67	97.59	99.76	97.42	98.57	99.42	82.58	90.02	97.10	92.43	94.70
W+P+C+Ch	97.4	94.28	95.75	98.82	96.24	97.46	93.25	83.21	87.45	99.69	91.85	95.60

Table 3: Evaluation of each elements considered in the proposed model for the exact identification of negation triggers, {W:Words - P:PoS - C: Casing - Ch: Chars} Prec: Precision, Rec: Recall and F1 measure are the metrics analyzed

ter the addition of each of the features studied. The decrease of performance for the corpus Bioscope in the categories of abstracts and clinical records after adding casing and chars features is remarkable. This may be because the model is over adjusted to certain patterns discovered with these features. According to the difference in negation occurrences between the different subsets in the Bioscope corpus, evaluating the subset of articles we obtain the highest standard deviation (±11%) and the worst performance. In this study, some errors were detected in the treatment of double negation and in the handling of multi-term expressions. The Spanish results show improvements as new features are added to the study. Many of the errors detected in the identification of the scope correspond to sentences with large negations. In order to study the robustness of the studied model, Table 4 shows the results obtained by evaluating the behaviour of the model detecting the scope, training it with data from a subset and validating it with data from the other sets. We have only been able to carry out this experiment with the corpus in English because it is the only one that shows a structure divided into categories with strong differences in the writing style. As can be observed, the best inter-domain performance is obtained when training with the abstracts subset. It is because the set of abstracts is the group that contains more negations and it uses a language very similar to the rest of subsets. The main problem detected is the mean length difference of the negations contained in the different subsets. The system trained with the set of abstracts tends to lose performance evaluating long sentences contained in the set of articles.

Finally, results in both datasets have been compared with the results of the state of the art systems results (Table 5 and Table 6). Competitive results have been obtained evaluating with Bioscope, although with a remarkable lower performance in terms of precision. This is mainly due to the fact that our system tends to generate a shorter length range than that collected in the gold standard. Some detected errors are those in which the first negation trigger appears far from the beginning of the scope. In these

	Testing with					
	Abstracts		Clinical records		Full papers	
Training with	PCS	F1	PCS	F1	PCS	F1
Abstracts	–.–	–.–	**84.90**	**91.83**	**55.85**	**71.67**
Clinical records	40.22	53.68	–.–	–.–	40.05	53.94
Full papers	**76.81**	**64.00**	81.68	89.91	–.–	–.–

Table 4: Bioscope corpus (English) - Evaluation of interdomain scope recognition. Results obtained by training with one of the Bioscope subsets (abstracts, clinical records and full papers) and testing with other

	Abstracts		Clinical records		Full papers	
System	PCS	F1	PCS	F1	PCS	F1
Proposed model	80.52	88.54	90.03	94.63	58.99	70.67
Li and Lu (2018)	**84.1**	91.3	94.4	95.59	**60.1**	69.23
Fancellu et al. (2017)	81.38	**92.11**	94.21	**97.94**	54.54	77.73
Fancellu, Lopez, and Webber (2016)	73.72	91.35	**95.78**	97.66	51.24	**77.85**

Table 5: Bioscope corpus (English) - Evaluation of negation scope recognition: Comparison with other state-of-the-art approaches

	Triggers			Scope		
System	P	R	F1	PCS	F1s	F1t
Proposed Model	**99.69**	**91.85**	**95.60**	**74.29**	**85.25**	**72.00**
Fabregat, Araujo, and Martinez-Romo (2018)	79.45	59.58	67.97	–	–	–
Loharja, Padró, and Turmo Borras (2018)	91.48	82.18	86.45	–	–	–

Table 6: SFU Review SP-NEG corpus - Evaluation of recognition of both negation scope (PCS, F1s, F1t) and negation triggers (P: Precision, R: Recall, F1): Comparison of obtained results by the proposed model with results from other state-of-the-art approaches

cases, the system makes mistakes such as ignoring the presence of multi-term expressions. The post-processing process also generates certain errors, especially in cases of double negation. Regarding the results obtained evaluating with the corpus SFU SP-NEG, as far as we know, there are only results for negation triggers recognition and only using training and test evaluation which makes it difficult to reach conclusions about the state of the art improvements. However, the results obtained for both the detection of negation triggers and for the recognition of the negation scope, are comparable to those obtained for English using the Bioscope corpus. Some of the errors reported in works about negation triggers detection that use the SFU SP-NEG corpus have been corrected incorporating the BILOU format and using of character embeddings.

5 Conclusions and future work

This research has focused on the generation of a common model to deal with negation in both English and Spanish languages. In order to generalize its application to different languages, the model has been trained mainly using non-language dependent writing features. Results show that it is a robust architecture based on a single supervised learning model for both detection of negation triggers and recognition of their scope. Performance obtained for English is comparable to state of the art and results obtained for Spanish are only slightly lower than for English. Possible future work lines are the study of more non-language dependent features and the improvement of the extraction of relationships between terms introducing n-gram embeddings as a feature. Some detected errors related to multi-term expressions suggest that working with n-gram embeddings can improve current precision results.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Science and Innovation within the projects PROSA-MED (TIN2016-77820-C3-2-R) and EX-TRAE (IMIENS 2017).

References

Bird, S. and E. Loper. 2004. Nltk: The natural language toolkit. In *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions*, ACLdemo '04, Stroudsburg, PA, USA. ACL.

Cardellino, C. 2016. Spanish Billion Words Corpus and Embeddings, March.

Chapman, W. W., W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan. 2001a. Evaluation of negation phrases in narrative clinical reports. In *Proceedings of the AMIA Symposium*, pages 105–109. AMIA.

Chapman, W. W., W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan. 2001b. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301–310.

Chapman, W. W., D. Hilert, S. Velupillai, M. Kvist, M. Skeppstedt, B. E. Chapman, M. Conway, M. Tharp, D. L. Mowery, and L. Deleger. 2013. Extending the negex lexicon for multiple languages. *Studies in health technology and informatics*, 192:677.

Chowdhury, M. F. M. and A. Lavelli. 2013. Exploiting the scope of negations and heterogeneous features for relation extraction: A case study for drug-drug interaction extraction. In *Proceedings of the 2013 Conference of the North American Chapter of the ACL: Human Language Technologies*, pages 765–771. ACL.

Cotik, V., V. Stricker, J. Vivaldi, and H. Rodríguez Hontoria. 2016. Syntactic methods for negation detection in radiology reports in spanish. In *Proceedings of the 15th Workshop on BioNLP 2016: Berlin, Germany, August 12, 2016*, pages 156–165. ACL.

Councill, I. G., R. McDonald, and L. Velikovich. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the workshop on negation and speculation in NLP*, pages 51–59. ACL.

Fabregat, H., L. Araujo, and J. Martinez-Romo. 2018. Deep learning approach for negation cues detection in spanish. In *NEGES 2018: Workshop on Negation in Spanish: Seville, Spain: September 19-21, 2018: proceedings book*, pages 43–48.

Fancellu, F., A. Lopez, and B. Webber. 2016. Neural networks for negation scope detection. In *Proceedings of ACL (Vol. 1: Long Papers)*, volume 1, pages 495–504. ACL.

Fancellu, F., A. Lopez, B. Webber, and H. He. 2017. Detecting negation scope is easy, except when it isn't. In *Proceedings of European Chapter of the ACL: Vol. 2, Short Papers*, volume 2, pages 58–63. ACL.

Goldin, I. and W. W. Chapman. 2003. Learning to detect negation with 'not'in medical texts. In *Proceedings of the Workshop on Text Analysis and Search for Bioinformatics, ACM SIGIR*. ACM SIGIR.

Jiménez-Zafra, S. M., R. Morante, M. Martin, and L. A. U. Lopez. 2018. A review of spanish corpora annotated with negation. In *Proceedings of the Conference on Computational Linguistics*, pages 915–924.

Konstantinova, N., S. C. de Sousa, N. P. Cruz, M. J. Maña, M. Taboada, and R. Mitkov. 2012. A review corpus annotated for negation, speculation and their scope. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. ELRA.

Lafferty, J. D., A. McCallum, and F. C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning*, pages 282–289, San Francisco, CA, USA. ICML.

Li, H. and W. Lu. 2018. Learning with structured representations for negation scope extraction. In *Proceedings of ACL (Vol. 2: Short Papers)*, volume 2, pages 533–539. ACL.

Loharja, H., L. Padró, and J. Turmo Borras. 2018. Negation cues detection using crf on spanish product review texts. In *NEGES 2018: Workshop on Negation in Spanish: Seville, Spain: September 19-21, 2018: proceedings book*, pages 49–54.

Padró, L. and E. Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*. LREC.

Pyysalo, S., F. Ginter, H. Moen, T. Salakoski, and S. Ananiadou. 2013. Distributional semantics resources for biomedical text processing. In *Proceedings of the 5th International Symposium on Languages in Biology and Medicine*, pages 39–44. LBM.

Ratinov, L. and D. Roth. 2009. Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on CoNLL*, pages 147–155. ACL.

Santos, C. D. and B. Zadrozny. 2014. Learning character-level representations for part-of-speech tagging. In *Proceedings of the International Conference on Machine Learning*, pages 1818–1826. ICML.

Sarawagi, S. and W. W. Cohen. 2005. Semi-markov conditional random fields for information extraction. In *Advances in Neural Information Processing Systems 17*, pages 1185–1192. NIPS.

Savova, G. K., J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute. 2010. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. *Journal of the AMIA*, 17(5):507–513.

Skeppstedt, M. 2011. Negation detection in swedish clinical text: An adaption of negex to swedish. *Journal of Biomedical Semantics*, 2(3):S3, Jul.

Vincze, V., G. Szarvas, R. Farkas, G. Móra, and J. Csirik. 2008. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(S11):S9.