

Cross-lingual transfer-learning approach to negation scope resolution

Anastassia Shaitarova¹, Lenz Furrer¹, Fabio Rinaldi²

¹Department of Computational Linguistics, University of Zurich

²Dalle Molle Institute for Artificial Intelligence Research, Switzerland

{*anastassia.shaitarova, lenz.furrer*}@uzh.ch, *fabio.rinaldi@idsia.ch*

Abstract

Detecting instances of negation in text is crucially important for several applications, yet it is often neglected. Several decades of research in automated negation detection have not yet provided a reliable solution, especially in a multilingual context. Negation scope resolution poses particular challenges since identifying the scope of influence of a negation cue in a sentence requires a deeper level of natural language understanding. Little work has been done on negation scope resolution in languages other than English. Meanwhile, transfer learning is in wide use and large multilingual models are available to the public. This paper explores the feasibility of a cross-lingual transfer-learning approach to negation scope resolution. Preliminary experiments with the Multilingual BERT model and data in English, French, and Spanish show solid results with the highest F1-score 84.73 on zero-shot transfer between English and French.

1 Introduction

Negation detection has been a kind of stumbling block in NLP research for many years. Linguists have not yet presented a proper generalization that can explain this highly complex linguistic phenomenon. The ability to negate is what makes us human, claims Horn (2001), which is not a comforting message to the NLP community, where *human* is practically synonymous with *ambiguous*.

Indeed, negation expresses itself through high syntactic and morphological variability which

complicates automated detection significantly. However, there is a certain degree of logical and semantic uniformity in negation, which is exhibited cross-linguistically. This uniformity can be potentially exploited by the new state-of-the-art NLP models. The work on negation detection is further complicated by the sparsity of annotated data, particularly in languages other than English. Therefore the search for annotation-independent approaches must continue.

The task of negation detection classically consists of two steps: 1) negation trigger or cue detection, and 2) negation scope resolution. Negation cues are words (*no*, *not*) or parts of a word (*un-* in *unhealthy*) that signal negation, while negation scope includes the part of a sentence that is semantically influenced by this signal. Identifying scope is more computationally challenging since the sphere of influence of each negation cue depends on a number of factors.

In this research we explore the ability of a multilingual BERT model (here: mBERT) released by Devlin (2018) to solve a fine-grained linguistic task of negation scope resolution across languages. We focus on surface form pertinent negations. The scope tokens are selected based on a binary classification negated/affirmed. Our research is guided by two main objectives:

- explore the feasibility of a zero-shot model transferring approach. In other words, can a model that is fine-tuned on labeled data in one language resolve negation scope in another language?
- test the possibility of using very few labeled examples as training data.

In Section 2 we highlight studies and approaches relevant for the question of cross-lingual negation scope resolution. Section 3 describes the datasets involved in this study. The preliminary experiments and their results are described in Section 4.

2 Related Work

Computational endeavours in negation detection began in the medical domain since it has a direct impact on the reliability of diagnosis. NegFinder (Mutalik et al., 2001) and NegEx (Chapman et al., 2001) were the first two algorithms specifically designed to handle negation. These and other early systems regarded negation detection from a medical rather than a linguistic perspective. Evaluation accounted only for correctly identified negated medical terms. The growing demand for reliable negation detection in other computational fields such as, *inter alia*, sentiment analysis (Councill et al., 2010), textual entailment (de Marneffe et al., 2006), and machine translation (Baker et al., 2010) promoted the development of negation detection as an NLP task in its own right.

Early lexical rule-based approaches like NegEx and its expanded versions (e.g. ConText, Harkema et al., 2009) used a fixed scope length that depends on a predefined number of tokens and the presence of delimiter words such as *but* and *however*. This is often sufficient for clinical texts that commonly contain short and/or ungrammatical sentences. Researchers also observed that negation scope is truly syntactic in nature (Szarvas et al., 2008; Morante and Blanco, 2012). Thus, most other rule-based, hybrid, and machine learning systems used syntax and dependency parsing to identify negation scope patterns. Context-free grammars were often used to employ these patterns as rules.

The entire first decade of research, however, was dedicated to solving negations in English. Only later, and gradually, the work on negation detection in other languages started taking place.

2.1 NegEx and its cross-lingual adaptations

The NegEx algorithm has been adopted for several European languages including Swedish (Skeppstedt, 2010), French (Deléger and Grouin, 2012), German (Chapman et al., 2013), and Spanish (Costumero et al., 2014). A comparative study conducted by Chapman et al. revealed several significant challenges associated with the adaptation process, the main one being the collection of triggers.

Negation trigger lists are inherently incomplete despite the fact that, following Zipf's law, a handful of triggers is responsible for most of the negations in the text. A set of negation triggers depends

greatly on the type of text in which it is found. Even if texts belong to the same domain but are of different types (e.g. radiology reports, discharge summaries, and surgical notes), they contain different negation triggers.

NegEx rules can be considered language agnostic but the compilation of negation cues has to be language specific. The same triggers have various levels of ambiguity and usage frequency depending on the language. Whether a cue negates to its left or right varies from language to language. Translating cues from English into languages that have a richer morphosyntactic variability (like French) or exhibit different or more flexible word order (like Swedish) introduces additional problems.

The restrictions described above make it advisable that a native speaker, and preferably a domain specialist, is involved in the compilation of a NegEx trigger list in the target language. Abdaoui et al. (2017) involved a bilingual text mining specialist to validate French cues automatically translated from English. A specialist, however, is not always available.

Despite all these difficulties NegEx is still in wide use because it is simple, reliable and needs no labelled data. Some researchers show that NegEx is enough for biomedical texts (Cotik et al., 2016; Elazhary, 2017) and assert that no other complex approach is necessary. Others disagree, referring to the inherent inability of rule-based systems to generalize (Wu et al., 2014; Sergeeva et al., 2019).

2.2 Machine learning approaches

The first study focusing on negation scope resolution was spearheaded by Morante et al. (2008). They framed it as a chunking problem and proved that it can be handled as a classification task at a token level. They used a k-nearest neighbor algorithm to assess each token in relation to the negation trigger.

The work was conducted on the BioScope corpus (Szarvas et al., 2008; Vincze et al., 2008). BioScope was the first publicly available sizeable corpus annotated for negation cues and scopes. It contains biomedical texts in English. The results showed an F1-score of 88.40 for negation scope resolution with the use of gold-standard cues.

Morante and Blanco (2012) went further and organized the *SEM2012 Shared Task that was ded-

icated entirely to resolving the scope and focus of negation. Additionally, in order to compensate for the lack of negation training data in the general domain, they annotated several Sherlock Holmes stories by Sir Arthur Conan Doyle. The best F1-score for scope detection on the Conan Doyle corpus (here: Sherlock) was 85.26, which was later outperformed with a score of 88.2 by Packard et al. (2014).

Fancellu et al. (2016) noted that previous systems for negation scope resolution were highly engineered, parser-dependent, and specific to English. They outperformed all previous results on the Sherlock corpus with an F1-score of 88.72 using BiLSTMs, pre-trained embeddings, and universal POS tags. The team then worked with a parallel English-Chinese corpus, NegPar (Liu et al., 2018) and asked the question that inspired this paper’s research:

Can we learn a model that detects negation scope in English and use it in a language where annotations are not available? (Fancellu et al., 2018, p. 1)

They used universal dependencies to abstract away from word order that differs between languages. The best performing cross-lingual model reached an F1-score of 72.46 and set the precedent for zero-shot cross-lingual negation scope detection.

When Google AI open-sourced the Bidirectional Encoder Representation for Transformers (BERT: Devlin et al., 2019), it “marked the beginning of a new era in NLP” (Alammar, 2018). Negation detection research followed suit, exploiting the transformer’s architectural features as well as linguistic knowledge gained from pre-training on massive amounts of data (Khandelwal and Sawant, 2019; Britto and Khandelwal, 2020). The results showed solid improvement on scope resolution across domains in English. The developed architecture, NegBERT, was customized for the experiments in this paper.

3 Data

There are several negation annotated corpora available to the public, but most of them are in English. Moreover, they all suffer from a lack of standardization in terms of annotation guidelines (Jiménez-Zafra et al., 2020). For example, Sherlock (Morante and Daelemans, 2012) is the

only corpus in English that annotates morphological negation cues such as affixes. It also allows discontinuous scopes which is not the case in other corpora. Sherlock and BioScope differ on whether they include the negation cue and the clause’s subject into the scope. Both corpora are freely available for download.

The SFU Review-NEG corpus (Konstantinova et al., 2012), a large multi-domain corpus of product reviews, mostly follows BioScope’s guidelines and does not include cues into the scope of negation. It is available on request from Simon Fraser University. We use all three English corpora in our experiments. We also combine them together into one data set that provides 7044 negation sentences.

The French data that we use here is described in Dalloux et al. (2017) and is publicly available on request¹. It has 3790 sentences total and is loosely modeled on the Sherlock corpus. The data in Spanish comes from the SFU ReviewSP-NEG corpus (Jiménez-Zafra et al., 2018) that can be requested via the same link as The SFU Review-NEG corpus above. It has 9445 sentences and its annotations follow the guidelines of the three aforementioned English corpora.

4 Experiments

We used NegBERT for our experiments but instead of the `BERT-base-uncased` model we imported `bert-base-multilingual-cased`. mBERT has been pre-trained on non-parallel data in 104 languages with the same training objectives as BERT: masked language modeling and next sentence prediction. The switch between the models is simple since they share a tokenizer and, for our task, can be initialized with a built-in class call `BertForTokenClassification` from HuggingFace².

The preprocessing of the three aforementioned English corpora is completed by NegBERT which duplicates sentences with multiple negations into multiple copies containing a single negation. The French and Spanish corpora are stored in formats that differ from the English corpora. We pre-processed these corpora ourselves and, for the sake of simplicity and consistency in the experiments, considered only sentences with one negation scope. Here we refer to these subcorpora

¹<http://people.irisa.fr/Clement.Dalloux/>

²https://huggingface.co/transformers/_modules/transformers/modeling_bert.html

| models | frXX | spXX | models | frXX | spXX | models | frXX | spXX |
|----------|-------|-------|----------------|-------|-------|----------------|-------|-------|
| SHERLOCK | 77.31 | 77.39 | SHERLOCK_FR100 | 84.64 | 77.51 | SHERLOCK.SP100 | 80.14 | 83.43 |
| BIOSCOPE | 82.17 | 77.40 | BIOSCOPE_FR100 | 87.15 | 78.12 | BIOSCOPE.SP100 | 84.20 | 83.24 |
| SFU | 75.89 | 78.36 | SFU_FR100 | 85.40 | 78.93 | SFU.SP100 | 81.62 | 82.80 |
| EN | 84.50 | 78.29 | EN_FR100 | 87.82 | 79.37 | EN.SP100 | 83.09 | 83.30 |
| FR | — | 78.44 | FR100 | 83.27 | 72.96 | FR.SP100 | — | 83.87 |
| SP | 81.55 | — | SP_FR100 | 86.68 | — | SP100 | 75.55 | 82.72 |

Table 1: F1-scores for negation scope resolution across languages and corpora. Each model is an mBERT model that has been fine-tuned on a subcorpus indicated in its name (eg. model **FR100** was fine-tuned on a minimal data set of 100 sentences in French).

as OneScopeFR for French and OneScopeSP for Spanish. OneScopeFR consists of 717 sentences while OneScopeSP has 2197.

Following our objectives, we trained models that can be used for zero-shot transfer and minimal training data experiments. Thus we fine-tuned mBERT on the English corpora, OneScopeSP, OneScopeFR, mini subsets of OneScopeFR and OneScopeSp as well as various combinations of corpora and subsets. All models were fine-tuned with MAX_LEN 250, batch size 8, and learning rate 1e-5. The reported F-scores are calculated on the per-token basis for label 1 (token in scope).

4.1 Zero-shot model transfer vs. rule-based approach for biomedical French texts.

Since the French corpus is a collection of medical reports, we decided to assess the ease and efficacy of NegEx and use the results as a baseline. We used the publicly available NegEx adaptation for Python³ which comes with cues in English. In order to compile a list of triggers in French we initially contacted [Deléger and Grouin](#). They readily provided the list they had produced for their own adaptation of NegEx. Processing OneScopeFR with these triggers resulted in an F1-score **45.55%**.

In order to improve the result and to have a fair comparison, we collected gold negation cues from OneScopeFR. Consecutive multiword cues such as *absence de* (en: lack of) were collected as multiword units. Non-consecutive multiword cues such as *ne ... pas ... ni ... ni, sans ... ne ... aucun* were collected as single token cues *ne, pas, ni, sans, aucun*. The final list consisted of 27 triggers.

Manual examination of the data helped decide on the directionality of each trigger. For example, our observations revealed that *absent* (en: absent)

³<https://code.google.com/archive/p/negex/downloads?page=2>

and *négatif* (en: negative) and their morphological variations generally affect context to the left while all the other cues negate to the right. As a result, 22 right- and five left-negating cues were added. Scope delimiters such as *mais* (en: but) and pseudo-triggers such as *ne cause pas* (en: does not cause) were copied from the original French list. Additionally, NegEx was modified not to include the full stop in the scope.

These adjustments boosted the results up to an F1-score of **63.89%**. OneScopeFR was also tested with mBERT fine-tuned on the combined English data which produced the best zero-shot result across the board with an F1-score of **84.73%**.

4.2 Zero-shot and minimal training data for French and Spanish texts.

In this set of experiments we test zero-shot approach using both OneScopeFR and OneScopeSP. Additionally, we create minimal training data sets for both languages by randomly selecting 125 sentences from each subcorpus. In the process of fine-tuning mBERT on these miniature sets, 20% of the data (25 sentences) is used for validation and prevention of overfitting through early stopping. We call the resulting models **FR100** and **SP100**.

The motivation behind this setup is a potential situation where annotated material for a particular language might not exist. It is, however, conceivable for a hypothetical researcher or a developer to be able to annotate about a hundred sentences without unreasonable effort.

Table 1 shows the results of our experiments. All the models are tested on sentences that are left in OneScopeFR and OneScopeSP after the extraction of the mini-sets. Thus, **frXX** contains 592 sentences in French and **spXX** has 2072 sentences in Spanish.

The results suggest that fine-tuning mBERT even on one hundred sentences produces results

that are worth considering. **FR100** → **frXX** shows an F1-score of 83.68, **SP100** → **spXX** is at 82.72.

These results are further improved by adding available training material in other languages to the mini-set. The best score for French (87.82) is achieved by the addition of all English data **EN_FR100** → **frXX**, while Spanish benefits most from the addition of French **FR_SP100** → **spXX**: 83.87. The highest score for zero-shot transfer is achieved by **EN** → **frXX** which is consistent with the previous experiment.

4.3 Discussion

The results produced by our experiments are positive. With low effort and little to no training data we can obtain results comparable to former state-of-the-art models. It is difficult to say, however, how much this outcome was to be expected.

Pires et al. (2019) showed that mBERT performs best for languages that share the same word order features. English, French, and Spanish are all SVO languages, meaning that their sentences mostly follow the subject-verb-object pattern. Thus, it was reasonable to expect good results from our experiments. When it comes to negation, however, the languages differ typologically. Dahl (1979) examined negation patterns in 240 different languages in 40 different language families and concluded that English, French, and Spanish belong to three different categories. In English the negation marker immediately follows the verb, but in Spanish the marker precedes it, while French shows both patterns.

A look into numerous studies dedicated to understanding BERT did not provide clear ideas of what to expect from our experiments. On one hand, BERT’s contextualized representations seem to possess robust knowledge over syntactic and dependency parse trees (Hewitt and Manning, 2019). On the other hand, Warstadt et al. (2019) claim that the syntactic knowledge in BERT appears to be partial and inconclusive. Rogers et al. (2020) analyzed over forty studies on BERT and concluded that BERT does not understand negation.

An examination of the output provides our own insights. An example of a French sentence with its translation into English and gold standard annotation is shown below. The negation cue is marked in **bold**, while scope is enclosed between the [square brackets].

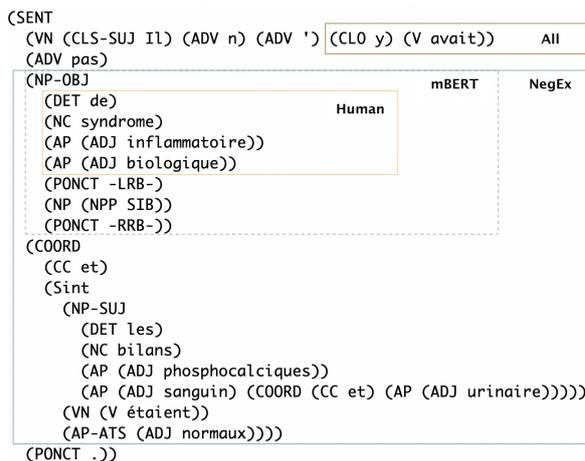


Figure 1: A syntax tree of a French sentence with negation scope identified by human annotators, mBERT fine-tuned on English data, and NegEx. The negation trigger is an embracing multiword cue **n’ ... pas**.

French: *Il n’ [y avait] **pas** [de syndrome inflammatoire biologique] (SIB) et les bilans phosphocalciques sanguin et urinaire étaient normaux .*

English: *[There was] **no** [biological inflammatory syndrome] (BIS) and blood and urine calcium phosphate levels were normal.*

Fig. 1 shows the same sentence as a constituent tree⁴ which provides syntactic visualization of scope resolution by different agents. The scope of negation marked by BERT in French is generally very consistent. It usually ends before a punctuation mark, another cue, a finite verb, or the conjunction *et*. It tends to trace the constituent boundaries. In fact, it is unclear why, in this example, human annotation did not include the abbreviation of the syndrome inside the scope.

5 Conclusion

Considering the linguistic complexity of negation scope resolution, the F1-score of 84.73 on a zero-shot transfer test is substantial. Fine-tuning on minimal training data also provided decent results. A deeper examination of the output could provide us with further improvements. Potentially, a standardized annotation scheme for the English corpora could improve all outcomes. Aside from negation, this study could be turned into a behavioral probing task that further explores BERT’s linguistic and cross-linguistic abilities. It would be interesting to test typologically different languages as well.

⁴Created with Berkeley Neural Parser

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](#).
- Jay Alamar. 2018. [The Illustrated BERT, ELMo, and co. \(How NLP Cracked Transfer Learning\)](#).
- Kathryn Baker, Michael Bloodgood, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin, and Christine Pitko. 2010. [A Modality Lexicon and its use in Automatic Tagging](#). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Benita Kathleen Britto and Aditya Khandelwal. 2020. Resolving the Scope of Speculation and Negation using Transformer-Based Architectures. *ArXiv*, abs/2001.02885.
- Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. [A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries](#). *Journal of Biomedical Informatics*, 34(5):301–310.
- Wendy W. Chapman, Dieter Hillert, Sumithra Velupillai, Maria Kvist, Maria Skeppstedt, Brian E. Chapman, Mike Conway, Melissa Tharp, Danielle L. Mowery, and Louise Deléger. 2013. Extending the NegEx Lexicon for Multiple Languages. *Studies in health technology and informatics*, 192:677–81.
- Roberto Costumero, Federico Lopez, Consuelo Gonzalo-Martín, Marta Millan, and Ernestina Menasalvas. 2014. [An Approach to Detect Negation on Medical Documents in Spanish](#). In *Brain Informatics and Health*, Lecture Notes in Computer Science, pages 366–375, Cham. Springer International Publishing.
- Viviana Cotik, Roland Roller, Feiyu Xu, Hans Uszkoreit, Klemens Budde, and Danilo Schmidt. 2016. [Negation Detection in Clinical Reports Written in German](#). In *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*, pages 115–124, Osaka, Japan. The COLING 2016 Organizing Committee.
- Isaac Councill, Ryan McDonald, and Leonid 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 Natural Language Processing*, pages 51–59, Uppsala, Sweden. University of Antwerp.
- Östen Dahl. 1979. [Typology of sentence negation](#). *Linguistics*, 17(1-2):79–106. Publisher: De Gruyter Mouton Section: Linguistics.
- Clément Dalloux, Vincent Claveau, and Natalia Grabar. 2017. [Détection de la négation : corpus français et apprentissage supervisé](#). In *SIIM 2017 - Symposium sur l'Ingénierie de l'Information Médicale*, pages 1–8, Toulouse, France.
- Louise Deléger and Cyril Grouin. 2012. [Detecting Negation of Medical Problems in French Clinical Notes](#). In *Proc of Int Health Inform*, Miami Beach, FL.
- Jacob Devlin. 2018. [Multilingual BERT Readme document](#). Library Catalog: github.com.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Hanan Elazhary. 2017. [NegMiner: An Automated Tool for Mining Negations from Electronic Narrative Medical Documents](#).
- 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: Long Papers)*, pages 495–504, Berlin, Germany. Association for Computational Linguistics.
- Federico Fancellu, Adam Lopez, and Bonnie L. Webber. 2018. Neural Networks for Cross-lingual Negation Scope Detection. *ArXiv*, abs/1810.02156.
- 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.
- John Hewitt and Christopher D. Manning. 2019. [A Structural Probe for Finding Syntax in Word Representations](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4129–4138, Minneapolis, Minnesota. Association for Computational Linguistics.
- Laurence R. Horn. 2001. *A natural history of negation*. The David Hume series. CSLI, Stanford, Calif.
- Salud María Jiménez-Zafra, M. Teresa Martín-Valdivia, M. Dolores Molina-González, and L. Alfonso Ureña-López. 2018. [Relevance of the SFU ReviewSP-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish](#). *Information Processing & Management*, 54(2):240–251.

- Salud María Jiménez-Zafra, Roser Morante, María Teresa Martín-Valdivia, and L. Alfonso Ureña-López. 2020. [Corpora Annotated with Negation: An Overview](#). In *Computational Linguistics*, volume 0, pages 1–87.
- Aditya Khandelwal and Suraj Sawant. 2019. NegBERT: A Transfer Learning Approach for Negation Detection and Scope Resolution. *ArXiv*, abs/1911.04211.
- Natalia Konstantinova, Sheila C.M. de Sousa, Noa P. Cruz, Manuel J. Maña, Maite Taboada, and Ruslan Mitkov. 2012. [A review corpus annotated for negation, speculation and their scope](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3190–3195, Istanbul, Turkey. European Language Resources Association (ELRA).
- Qianchu Liu, Federico Fancellu, and Bonnie Webber. 2018. [NegPar: A parallel corpus annotated for negation](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Marie-Catherine de Marneffe, Bill MacCartney, Trond Grenager, Daniel Cer, Anna Rafferty, and Christopher D Manning. 2006. Learning to distinguish valid textual entailments. page 6.
- 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)*, pages 265–274, Montréal, Canada. Association for Computational Linguistics.
- Roser Morante and Walter Daelemans. 2012. ConanDoyle-neg: Annotation of negation in Conan Doyle stories. page 6.
- Roser Morante, Anthony Liekens, and Walter Daelemans. 2008. [Learning the Scope of Negation in Biomedical Texts](#). In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 715–724, Honolulu, Hawaii. Association for Computational Linguistics.
- Pradeep Mutalik, Aniruddha M. Deshpande, and Prakash M. Nadkarni. 2001. [Research Paper: Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: A Quantitative Study Using the UMLS](#). *JAMIA*.
- 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)*, pages 69–78, Baltimore, Maryland. Association for Computational Linguistics.
- Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is Multilingual BERT? In *ACL*.
- Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. [A Primer in BERTology: What we know about how BERT works](#). *arXiv:2002.12327 [cs]*. ArXiv: 2002.12327.
- Elena Sergeeva, Henghui Zhu, Peter Prinsen, and Amir Tahmasebi. 2019. [Negation Scope Detection in Clinical Notes and Scientific Abstracts: A Feature-enriched LSTM-based Approach](#). *AMIA Summits on Translational Science Proceedings*, 2019:212–221.
- Maria Skeppstedt. 2010. Negation Detection in Swedish Clinical Text. In *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents*, Louhi '10, pages 15–21, USA. Association for Computational Linguistics. Event-place: Los Angeles, California.
- György Szarvas, Veronika Vincze, Richárd Farkas, and János Csirik. 2008. [The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts](#). In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 38–45, Columbus, Ohio. Association for Computational Linguistics.
- 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(S11):S9.
- Alex Warstadt, Yu Cao, Ioana Grosu, Wei Peng, Hagen Blix, Yining Nie, Anna Alsop, Shikha Bordia, Haokun Liu, Alicia Parrish, Sheng-Fu Wang, Jason Phang, Anhad Mohananey, Phu Mon Htut, Paloma Jeretic, and Samuel R. Bowman. 2019. [Investigating BERT’s Knowledge of Language: Five Analysis Methods with NPIs](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2877–2887, Hong Kong, China. Association for Computational Linguistics.
- Stephen Wu, Timothy Miller, James Masanz, Matt Coarr, Scott Halgrim, David Carrell, and Cheryl Clark. 2014. [Negation’s Not Solved: Generalizability Versus Optimizability in Clinical Natural Language Processing](#). *PLOS ONE*, 9(11):1–11.