Annotating Negation in Spanish Clinical Texts

Noa P. Cruz Díaz[1], Roser Morante Vallejo[2], Manuel J. Maña López[3], Jacinto Mata Vázquez[3] and Carlos L. Parra Calderón[1]

[1]IBiS/Virgen del Rocío University Hospital/CSIC/University of Seville
[2]CLTL Lab, VU Amsterdam
[3]Departamento de Tecnologías de la Información, University of Huelva
{noapatricia.cruz.exts,carlos.parra.sspa}@juntadeandalucia.es
{r.morantevallejo}@vu.nl
{manuel.mana,jacinto.mata}@dti.uhu.es

Abstract

In this paper we present on-going work on annotating negation in Spanish clinical documents. A corpus of anamnesis and radiology reports has been annotated by two domain expert annotators with negation markers and negated events. The Dice coefficient for inter-annotator agreement is higher than 0.94 for negation markers and higher than 0.72 for negated events. The corpus will be publicly released when the annotation process is finished, constituting the first corpus annotated with negation for Spanish clinical reports available for the NLP community.

1 Introduction

In this paper we present the UHU-HUVR corpus of Spanish clinical reports annotated with negation information. Negation processing (Morante and Blanco, 2012) is an emergent task in Natural Language Processing (NLP). Initially, negation processing systems were developed to detect negation in clinical texts in order to extract accurate information about the patients. The systems were rule-based (Chapman et al., 2001) because of the lack of annotated corpora to train machine learning systems. Currently, annotated corpora are still scarce due to the legal and ethical requirements that have to be fulfilled. In addition, most of the corpora only contain documents written in English. An example is the BioScope corpus (Szarvas et al., 2008) which contains 1,954 radiology reports from the Computational Medicine Center in Cincinnati, annotated for negations and uncertainty along with the scopes of each phenomenon.

For Spanish clinical reports, to the best of our knowledge, there are only a few attempts to annotate negation, mostly as a secondary task within broader projects, and none of the annotated resources are publicly available. For example, Costumero et al. (2014) adapted the NegEx algorithm (Chapman et al., 2001) to detect negated clinical conditions in Spanish written medical documents. Although the authors mention that 500 medical texts have been manually annotated, they do not provide more information about this gold-standard. Oronoz et al. (2015), annotated negated adverse drug reactions, but they did not annotate the words expressing negation. The negation was annotated as a modifier of the disorder or drug. Stricker et al. (2015) used a dataset composed of about 85,600 reports of ultrasonography studies performed in a Spanish public hospital. Clinical findings were annotated in a subset of the reports as being affirmed (if it was possible to infer that the finding was present in the patient) or negated (if the finding was absent). The annotations were recently expanded by Cotik et al. (2016), including the values *probable* (if it was not certain that the finding was present, but it was probable) and *doubt* (if the finding corresponded to the past or if it was not clear for the annotator if the finding was present or not). None of the studies provides information about the availability of the corpora.

Being aware of the limitations of the current approaches and the necessity of filling this gap, we present the first version of the UHU-HUVR Corpus, a set of clinical documents annotated with negation markers and their linguistic scope. The annotated corpus and its guidelines will be made publicly available. The paper is structured as follows: Section 2 describes the annotation process and provides information about the corpus. Section 3 presents the results of the agreement analysis and discusses difficult or interesting cases. Finally, conclusions and future work are presented in Section 4.

2 Annotating Negation in Spanish Clinical Reports

The corpus that we present consists of 604 clinical reports from the Virgen del Rocío Hospital in Seville (Spain). Specifically, the set of documents consists of diagnosis information of 276 radiology reports and the personal history of 328 anamnesis reports written in free text, as shown in Table 1. The reports were randomly collected among reports from the first semester of 2013, and were fully anonymized in order to satisfy legal and ethical issues. The radiology documents are longer than the anamnesis documents (339.05 and 157.66 average words per document respectively), but the average words per sentence is similar (17.50 versus 16.87).

	Anamnesis	Radiology
#Docs	328	276
#Sentences	3,065	5,347
#Words	51,712	93,579
Avg. length docs (sentences)	9.34	19.37
Avg. length docs (words)	157.66	339.05
Avg. length sentences (words)	16.87	17.50

Table 1: Statistics about the UHU-HUVR corpus.

The entire corpus was annotated by two domain experts who followed the guidelines that we developed. The annotation process started with two pilot experiments. In the first one, the authors annotated 7 reports in order to gain insight to produce the first version of the guidelines, which was written by one of the authors, who is a linguistics expert. In the second one 100 anamnesis documents were annotated by the two annotators in order to train them, and test and improve the guidelines. These documents are not included in the final corpus. During all the annotation process, the annotators were not allowed to communicate with each other. The problematic cases that they encountered were discussed with the expert linguist and author of the guidelines. As work in progress, another author who is expert annotator is currently solving the disagreement cases, acting as adjudicator, in order to generate the gold-standard corpus. The annotation tool used was CAT (Lenzi et al., 2012). Two markables were defined (negation marker and negated event) plus a negation relation between the marker and the event.

The annotation task consisted on annotating the events that are affected by contextual negation, as well as the words that express negation. We will refer to these words as *markers*. In the examples that follow, negated events are marked within brackets and negation markers in bold. In example 1, the word *no* is the negation marker and *alteraciones en el luminograma aéreo de tráquea* is the event affected by this negation. Examples of events with negative polarity are the following ones:[1]

1. **No** visualizamos [alteraciones en el luminograma aéreo de tráquea]. (En. We did not visualize alterations in the aerial luminogram of trachea.)

2. **No** [aumentos ganglionares hilio mediastínicos] **ni** [hallazgos de interés a lo largo del esófago]. (En. No mediastinal hilum nodal enlargement nor findings of interest along the esophagus.)

3. **Sin** [lesiones óseas focales que sugieran metástasis en este barrido]. (En. No focal bone lesions suggestive of metastasis in this scan.)

4. Desde hace 20 años **abandono** radical del [hábito tabáquico]. (En. During the last 20 years radical stopping of smoking habit.)

The annotation guidelines follow closely the Thyme corpus guidelines (Styler IV et al., 2014) with some adaptations. We defined as clinical event any event that is relevant to elaborate the clinical chronology of a patient such as a diagnosis, tumors, habits, medical tests, or events related to the functional evaluation of the patient. Events that do not contribute clinical information are not annotated. The difficulty of annotating events lies not just in identifying what is an event, but in determining which is the chain of words that express the event. We decided to annotate all the words that express the event, excluding punctuation marks. In Example (1) the full NP is annotated as event: *alteraciones en el luminograma aéreo de tráquea*. For negation markers, the minimum number of words is annotated.

We do not always annotate the mentions of negative results of a test as negation markers because the fact that a test is negative does not necessarily imply that a clinical event is not happening. In Example (5) no negation is annotated because, although the results of the Z-N stain are negative,

---

[1]We provide between parentheses English translations for all examples.

the test has taken place. However, when the name of the test is the same as the name of the clinical condition, such as a disease, then the negative results are annotated as negation because the negative result indicates that the patient does not have the disease (see Example (6)).

5. Técnicas de Z-N (normal y largo) negativo. (En. Negative Z-N stain (normal and long).)

6. Serología materna: [Toxoplasma]: **Negativo**. [VHB]: **Negativo**. [Rubeola]: **Negativo**. [Lues]: **Negativo**. (En. Maternal serology: Toxoplasma: Negative. VHB: Negative. Rubella: Negative. Lues: negative.)

In addition, negative results are sometimes expressed with the "-" sign, which was annotated as negation marker in the same cases as indicated above.

As for affixal negation, we take a pragmatic approach and annotate the full word where the affix occurs as a negation marker and as a negated event, when an event is negated. It is the case of *afebril* (En. without fever), where *a-* is the negation marker. Due to limitations of the tools we could not split the affixes in order to mark them independently.

Finally, some coordination structures involve negation markers. In this cases, each negation marker in the structure has its own scope, as shown in Example (7), where the first coordination marker *no* scopes over *alteraciones a nivel de los distintos ligamentos y estructuras músculo-tendinosas* and Example (8), where the second coordination marker *así como* scopes over *alteraciones ... de las restantes partes blandas*. In clinical reports, some expressions like *así como*, which are usually not negation markers, are used as such, and thus, we decided to mark them as negation markers, even if what brings negation to the event is the coordination structure and not only the second coordinating element. We have also marked "/" as negation marker when it acts as a coordinating particle, as in Examples (9) and (10).

7. **No** hemos observado [alteraciones a nivel de los distintos ligamentos y estructuras músculo-tendinosas], así como de las restantes partes blandas. (En. We have not observed alterations at the level of the different ligaments and musculo-tendinous structures, as well as of the other soft parts.)

8. No hemos observado [alteraciones] a nivel de los distintos ligamentos y estructuras músculo-tendinosas, **así como** [de las restantes partes blandas]. (En. We have not observed alterations at the level of the different ligaments and muscle-tendon structures, as well as of the other soft parts.)

9. Actualmente **no** hay evidencia de [nódulos pulmonares] / adenomegalias mediastínicas. (En. No evidence of pulmonary nodules / mediastinal adenomegalies)

10. Actualmente no hay evidencia de nódulos pulmonares **/** [adenomegalias mediastínicas].

2.1 Negation in Spanish Clinical Texts

The annotations show that negation is a frequent phenomenon in Spanish clinical texts. As it is shown in Table 2, more than 22% of the sentences in radiology reports contain negation markers, whereas in anamnesis reports this percentage is even higher, 35.20%. This fact is related to the nature of the two types of documents. Whereas a radiology report reflects the radiologists observations, the anamnesis report describes the history of the patient, including clinical conditions that the patient has not gone through.

In Spanish clinical domain, (Oronoz et al., 2015) reported that 27.58% of diseases presented in a corpus of electronic health records are negated. Although this percentage is not directly comparable with the percentage of negated sentences shown in Table 2, the numbers seem similar. The frequency of negation in Spanish for the reviews domain is sensibly higher. Martí et al. (2016) analysed the 75% of SFU Review SP-NEG corpus and claimed that 46.64% of sentences contain at least one negation. However, unpublished analysis about the entire SFU Review SP-NEG corpus shown that the amount of sentences that contain at least one negation in this corpus is the 31.90%, which represents a similar value to negation frequency in anamnesis reports.

	#Negation sentences	% Negation sentences	# Negation markers
Anamnesis	1,079	35.20	1,572
Radiology	1,219	22.80	1,985

Table 2: Negation statistics in the UHU-HUVR corpus.

Negation markers in the UHU-HUVR corpus amounted to 69 in the anamnesis subcollection and 52 in the radiology subcollection, with the top 10 most frequent markers shown in Tables 3 and 4. It is interesting to note that, in Spanish, the first two markers, *no* ('no') and *sin* ('without'), constitute more than 70% of the total frequency of all the markers found in the corpus, while the remaining markers cover only about 30%. 28 of the markers are common for radiology and anamnesis reports. It is also interesting to see that signs as "/" or "-" are considered to be negation markers. A specific negation marker of the radiology reports is *namc* which stands for *no alergias medicamentosas conocidas* (En. no known drug allergies).

Marker	Frequency	%	Acc. %
no	936	59.54	59.54
sin	197	12.53	72.07
ni	143	9.10	81.17
niega	66	4.20	85.37
namc	31	1.97	87.34
negativo	26	1.65	88.99
nunca	20	1.27	90.26
o	19	1.21	91.47
-	14	0.89	92.36
independiente	11	0.70	93.06

Table 3: The most frequent negation markers in the anamnesis subcollection of the UHU-HUVR corpus.

Marker	Frequency	%	Acc. %
sin	877	44.18	44.18
no	662	33.35	77.53
ni	258	13.00	90.53
así como	28	1.41	91.94
ausencia	27	1.36	93.30
o	27	1.36	94.66
tampoco	14	0.71	95.37
desaparecido	7	0.35	95.72
/	6	0.30	96.02
asimetría	6	0.30	96.32
inespecíficas	6	0.30	96.62
inespecífico	6	0.30	96.92

Table 4: The most frequent negation markers in the radiology subcollection of the UHU-HUVR corpus.

3 Preliminary results

Table 5 illustrates the results obtained for Inter-Annotator Agreement (IAA) between the two annotators in terms of Dice's coefficient measure (Dice, 1945) as calculated by the CAT annotation tool. We are currently working on the adjudication process. The IAA for markers is high and almost equal in radiology (0.949) and anamnesis (0.947) reports, whereas the IAA for events is lower than for markers. The IAA for events in radiology reports is lower (0.729) than for anamnesis reports (0.840).

	Marker Level		Token Level	
	Markers	Events	Markers	Events
Anamnesis	0.947	0.840	0.947	0.896
Radiology	0.949	0.729	0.949	0.860

Table 5: Preliminary Inter-Annotator Agreement in terms of Dice's coefficient.

The lower scores for radiology reports can be explained by the fact that the average length for events in anamnesis reports is lower (2,10 words) than in radiology reports (2,81 words), and thus there are more tokens that can be differently annotated for the same event. Additionally, it is also the case that interpreting the information in the radiology reports requires more specialised knowledge than interpreting the information in the anamnesis reports. In the latter, the events are better delimited and there are certain frequent negated expressions that are repeated throughout the reports, such as *No* or *Niega* followed by an event, as in the examples below.

11. **No** [hábitos tóxicos]. (En. No toxic habits.)

12. **Niega** [alergias]. (En. Denies allergies.)

13. **No** [alergias medicamentosas]. (En. No drug allergies.)

In the radiology reports it is more difficult to agree on which is the event affected by the marker, as in Example (14).

14. **Sin** [hallazgos en la distribución visualización de asas y mesos intestinales y órganos pélvicos]. (En. No findings in the distribution visualization of intestinal loops and mesenteries and pelvic organs.)

Furthermore, in many radiology reports incorrect syntactic constructions are used, as in Example (15), where two negation markers are used after each other forming an agrammatical expression *ni tampoco*, which can be confusing for the annotators.

15. **No** se [realza] **ni tampoco** [las ramas tributarias segmentarias de lngula] y [LII]. (En. It is not enhanced neither the segmental tributary branches of lingula and LII.)

The IAA is similar to the IAA reported for the clinical collection of the BioScope corpus (Szarvas et al., 2008) in English, where the IAA between the two annotators was 0.907 for the markers and 0.762 for the scope. This shows that the difficulty of the task is independent of the language in which the documents are written and inherent to the type of texts.

Disagreement cases were analysed by the adjudicator who decides on the correct annotations. Most of the disagreements were the result of a human error, i.e., the annotators missed a word or included a word that did not belong either to the event or to the marker. However, other cases of disagreement can be explained by the difficulty of the task and the lack of clear guidance.

A great number of disagreements are related to the difficulty of detecting what is an event and which is the chain of words that express the event, as in Example (16), where the first annotator identified as event the words *Contraste I* while the second one annotated *Contraste I . V .*, but none annotated the correct span of the event.

16. **Sin** [Contraste I. V. de Senos Paranasales] (En. Without Contrast I.V. of paranasal sinuses.)

Cases in which negation is expressed by affixes were a source of disagreement due to the lack of initial guidelines. An example is the word *afebril* which expresses absence of fever. In these cases, the whole word should be marked both as an event and as a marker. In contrast, words such as incontinencia urinaria ('urinary incontinence'), which contain a negation affix (*in-* in this case) do not have to be annotated as negation markers since the clinical condition that they express is not negated.

More cases of disagreement involve a false double negation as shown in Example (17). One of the annotators identified two markers (*no* and *tampoco*) instead of one, which is the correct solution because this is not a case of double negation.

17. Nunca le han dicho que tuviera anemia porque **no** [ha acudido tampoco al médico]. (En. (S)he has never been told that (s)he had anemia because (s)he did not go to the doctor either.)

Another source of disagreement are cases in which one marker negates several events, as in Example (18). One annotator annotated the word *niega* as a marker and marked as event the rest of the sentence, while the other annotator also identified the word *niega* as a marker but annotated each event separately, which is the correct solution. This case represents an enumeration where several events are negated by the same marker. Therefore, each event had to be annotated and related to the marker independently, generating 4 negation relations.

18. **Niega** [HTA], [diabetes], [dislipemias] u [otros FR vascular]. (En. Denies AHT, diabetes, dyslipidemia or other FR vascular.)

Jiménez-Zafra et al. (2016) encountered the same type of disagreements when annotating a Spanish review corpus (Taboada et al., 2006) with negation information, i.e., lack of agreement between annotators about the scope, the event and the discontinuities. This type of errors amount to 63.26% of the total. They also mentioned semantic problems that arose from the interpretation of negation patterns in comparative and contrastive constructions.

4   Conclusions

We have presented the first version of the UHU-HUVR Corpus, a collection of 604 clinical documents written in Spanish and annotated with negation markers, negated events and their relations. The corpus contains two types of reports, anamnesis and radiology. In both of them negation is a frequent phenomenon that needs to be treated for natural language processing purposes. The high IAA obtained suggests that the task is well defined. As expected, the agreement is higher for negation markers than for negated events, and higher in the anamnesis reports than in the radiology reports. As future work we plan to perform a detailed disagreement analysis in order to improve the guidelines for future annotation projects and to gain insight into the complexity of the task.

Acknowledgments

References

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory Cooper, and Bruce G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*, 35(5):301–310.

Roberto Costumero, Federico Lopez, Consuelo Gonzalo-Martín, Marta Millan, and Ernestina Menasalvas, 2014. *An Approach to Detect Negation on Medical Documents in Spanish*, pages 366–375. Springer International Publishing, Cham.

Viviana Cotik, Darío Filippo, and José Casta no. 2016. An approach for automatic classification of radiology reports in spanish. *Stud Health Technol Inform*, 216:634–638.

Lee R. Dice. 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Salud María Jiménez-Zafra, Maite Martín, L. Alfonso Ureña Lopez, Toni Martí, and Mariona Taulé. 2016. Problematic cases in the annotation of negation in spanish. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM)*, pages 42–48, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. 2012. Cat: the celct annotation tool. In *In Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC 2012*, pages 333–338.

M. Antònia Martí, M. Teresa Martín-Valdivia, Mariona Taulé, Salud María Jiménez-Zafra, Montserrat Nofre, and Laia Marsó. 2016. La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural*, 57:41–48.

Roser Morante and Eduardo Blanco. 2012. *SEM 2012 Shared Task: Resolving the scope and focus of negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 265–274, Stroudsburg, PA, USA. Association for Computational Linguistics.

Maite Oronoz, Koldo Gojenola, Alicia Pérez, Arantza Díaz de Ilarraza, and Arantza Casillas. 2015. On the creation of a clinical gold standard corpus in spanish. *J. of Biomedical Informatics*, 56(C):318–332, August.

Vanesa Stricker, Ignacio Iacobacci, and Viviana Coti. 2015. Negated findings detection in radiology reports in spanish: an adaptation of negex to spanish. In *Proceedings of the IJCAI - Workshop on Replicability and Reproducibility in Natural Language Processing: adaptative methods, resources and software*, Buenos Aires, Argentina.

William Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, and James Pustejovsky. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.

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, June. Association for Computational Linguistics.

Maite Taboada, Caroline Anthony, and Kimberly Voll. 2006. Methods for creating semantic orientation dictionaries. In *Conference on Language Resources and Evaluation (LREC)*, pages 427–432.