NUBES: A Corpus of Negation and Uncertainty in Spanish Clinical Texts

Salvador Lima¹, Naiara Perez¹, Montse Cuadros¹, German Rigau²
¹SNLT group at Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia/San-Sebastián, 2009, Spain
²IXA group. HiTZ centre. University of the Basque Country UPV/EHU, Donostia/San-Sebastián, 2018, Spain
{slima, nperez, mcuadros}@vicomtech.org, german.rigau@ehu.es

Abstract
This paper introduces the first version of the NUBES corpus (Negation and Uncertainty annotations in Biomedical texts in Spanish). The corpus is part of an on-going research and currently consists of 29,682 sentences obtained from anonymised health records annotated with negation and uncertainty. The article includes an exhaustive comparison with similar corpora in Spanish, and presents the main annotation and design decisions. Additionally, we perform preliminary experiments using deep learning algorithms to validate the annotated dataset. As far as we know, NUBES is the largest publicly available corpus for negation in Spanish and the first that also incorporates the annotation of speculation cues, scopes, and events.

Keywords: negation, uncertainty, clinical texts, Spanish

1. Introduction
The aim of Natural Language Understanding is to capture the intended meaning of texts or utterances. However, until recently, research has predominantly focused only on propositional aspects of meaning. Truly understanding language involves taking into account many linguistic aspects which are usually overlooked. These linguistic phenomena are sometimes referred to as Extra-Propositional Aspects of Meaning (EPAM) (Morante and Sporleder, 2012). Some examples of EPAM include factuality, uncertainty, opinions, beliefs, intentions or subjectivity. Documents enriched with this kind of information can be of utmost importance. For instance, in a domain such as the biomedical, the implicit meaning of a sentence can be crucial to differentiate whether a patient suffers from a disease or not, or whether they should be taking or not a given drug.

One way to learn these nuances is through means of an annotated corpus. Unfortunately, there are not many corpora that cover these phenomena. Just a few consider negation, a key aspect of factuality—and, even fewer, uncertainty. Negation is understood as an element that modifies the truth value of an event or a statement, that makes explicit that an event is absent; uncertainty (also called speculation) occurs when a speaker is not sure whether an event or statement is true.

Usually, negation and uncertainty are annotated in two parts: on the one hand, the phrase that triggers the change of meaning (called ‘cue’, ‘trigger’ or ‘marker’) and, on the other hand, the words that are affected by them (called ‘scope’). For a higher level of granularity, there are other elements that can be annotated, such as the element most clearly affected by the cue (called ‘event’) or the element that reinforces or diminishes the meaning of the cue (called ‘polarity’). A typical annotation that includes all these elements is shown in example (1).

(1) La paciente ingresa en UCI con la sospecha de {possible} encefalitis
The patient is admitted to ICU under suspicion of possible encephalitis

This paper describes the NUBES corpus (Negation and Uncertainty annotations in Biomedical texts in Spanish), a new collection of health record excerpts enriched with negation and uncertainty annotations. To date, NUBES is one of the largest available corpus of clinical reports in Spanish annotated with negation, and the first one that includes the annotation of speculation cues, scopes, and events. Additionally, we also present an extension of the IULA-SCRC corpus (Marimon et al., 2017) enriched with uncertainty using the same guidelines developed for NUBES. In order to validate the annotated corpus, we present some experimentation using deep neural algorithms. NUBES, the extension of IULA-SCRC—under the name of IULA++, as well as the guidelines, are publicly available. The paper is structured as follows: first, a brief survey of related work and corpora is presented. Then, Section 3 explains the design decisions, annotation guidelines and the annotation process of NUBES. Next, we present an overview of the main characteristics of the corpus. Section 4 presents some preliminary experiments with NUBES. In Section 5 we discuss some of the difficulties faced during the development of the NUBES guidelines. Finally, Section 6 presents the conclusions reached and the future work to extend and improve NUBES.

2. Related Work
Negation is such a complex phenomenon that it has been studied from the perspective of multiple fields, ranging

1 In the biomedical domain, ‘event’ refers to any ‘medical entity’, not only to an action, happening, etc.

2 In the following examples, cues are marked in bold, scopes in italics, events are underlined and polarity items are enclosed between curly brackets; translations to English are given below each corresponding example.

https://github.com/Vicomtech/NUBes-negation-uncertainty-biomedical-corpus
Table 1: Comparison between existing biomedical negation and/or uncertainty corpora in Spanish and NUBES, adapted from Jiménez-Zafra et al. (2018a) and Martí and Taulé (2018). Marimon et al. (2017) report 3,194, but we counted 3,370 sentences in the publicly available corpus. 2354.677 emergency admission notes. 3513 radiology reports. 427.58% of the diseases annotated are negated. 56% of the “findings” annotated are negated. 61.90% of the diseases annotated are speculative.

|                | IxaMed-GSC | UHU-HUVR | IULA-SCRC | IULA+ | CL2017 | Co2017 | NUBES |
|----------------|------------|---------|-----------|-------|--------|--------|-------|
| negation cue   | no         | yes     | yes       | yes   | yes    | yes    | yes   |
| uncertainty cue| no         | no      | yes       | yes   | no     | yes    | yes   |
| scope          | no         | yes     | yes       | no    | no     | yes    | yes   |
| event          | yes        | yes     | no        | yes   | no     | yes    | yes   |
| sentences      | 5,410      | 8,412   | 3,194?    | 3,370?| ?2     | ?3     | 29,682 |
| with negation (#)?  | 2,298      | 1,093   | 957       | ?     | ?      | 7,567  |
| with negation (%)  | ?4         | 27.32   | 34.22     | 28.40 | ?5     | 25.49  |
| with uncertainty (#)?  | ?           | 0       | 182       | 0     | ?      | 2,219  |
| with uncertainty (%) | ?4         | 0       | 5.40      | 0     | ?     | 7.48   |

4Some of the the articles do not report all the details introduced in the table. Such cases have been marked with a question mark. CL2017 and Co2017 refer to the works by Campillos Llanos et al. (2017) and Cotik et al. (2017), respectively.

5https://spacy.io/
3.1. Annotation process

An initial draft of our guidelines was produced by extending IULA-SCRC’s to include uncertainty. After annotating IULA-SCRC with this initial draft, we decided to make further changes with respect to negation by annotating a) negations inside indirect speech (e.g., ‘The patient denies’); b) verbs that convey a change of state (e.g., ‘remove’); and, c) morphological negation (e.g., ‘incoherent’). Other minor changes to the guidelines had to be made in order to accommodate uncertainty annotations. These differences with IULA-SCRC and the other corpora are further described in Section 5.

After producing the second draft, two linguists worked independently on a first batch of documents of the NUBES corpus. Their results were compared and multiple questions and disagreements that arose were discussed. The team also consulted a medical expert who aided them with some difficult scenarios, which are also examined in Section 5. All this greatly contributed towards producing the final version of the guidelines.

Then, the two linguists annotated the same batch adhering to the final guidelines. The inter-annotator agreement was then calculated on the second draft annotations and the final guideline annotations. As Table 2 shows, the agreement (Cohen’s kappa, $\kappa$, and linearly weighted kappa, lw$\kappa$) improved after the discussion, particularly for cues. The low agreement in polarity items is explained by the fact that they occur very few times (15) and the number of possible tags is also small (2: either it is a polarity item or it is not), which distorts the $\kappa$ measurement. The percentage agreement in the 2nd round for this class is actually 99.95%.

![Table 2: Inter-annotator agreement between 2 annotators on the first batch (2,971 sentences). $N$ is the number of tag types considered. The best results are highlighted in bold.](image)

3.2. Annotation guidelines

NUBES includes three main annotated elements: negation cues, uncertainty cues and their scope. Moreover, polarity items and events are also annotated as part of the scope.

3.2.1. Negation cues

We define negation cues as elements that modify the truth value of a clause or specify the absence of an entity. Three different types of cues can be distinguished: syntactic, lexical and morphological.

**Syntactic negation cues.** These are mostly function words or adverbs that can accompany multiple categories of words. It is the simplest type of negation, as well as the most common, as it covers words such as ‘no’ (no) and ‘sin’ (without):

(2) Fiebre de 38.5 sin foco
38.5 degrees fever without a focus

Negative time adverbs, such as ‘nunca’ (never), can also act as syntactic cues.

**Lexical negation cues.** They are content words or multi-word expressions that convey negation depending on the context, including verbs, adjectives or noun phrases. These cues are harder to detect as the way in which they negate a phrase is usually subtler than that of syntactic cues. Some examples are ‘suspend’ (‘suspend’), ‘incapacidad para’ (‘inability to’) o ‘descartar’ (‘discard’):

(3) Desestiman actualmente la realización de endoscopia
At present they dismiss conducting an endoscopy

Noun phrases with negative determiners are also considered lexical cues:

(4) Ninguna de ellas de evolución aguda-subaguda
None of them of acute-subacute course

**Morphological negation cues.** Morphological negation refers to negation by means of affixes. Since NUBES is a medical texts corpus, we decided to limit the annotation of these cues to words that explicitly state the absence of symptoms (‘afebril’, afebrile) or that could be seen as negating a symptom or state (‘deshidratado’, dehydrated). Words that do not fulfill those conditions or that are part of a condition name are not annotated. In general, as long as a word could be reformulated as a negated sentence that would be annotated under those conditions, the word would be classified as a cue. For example, ‘insuficiencia’ (failure), as in example (6) was not annotated because ‘no suficiencia’ is ungrammatical.

(5) Afebril al ingreso
Afebrile at admission

(6) Presentó descompensacion de su insuficiencia cardiaca
[The patient] showed decompensation of their heart failure

Finally, it is worth mentioning that not all appearances of negation cues are annotated as such. There are two main cases. On the one hand, there are formulas that simply
change the polarity of a positive event without truly negating it (‘casi sin’, almost no; ‘no siempre’, not always). These formulas can be restated without the negative cue with no real change in meaning, so we did not consider them. On the other hand, there are negation cues that are actually part of an uncertainty cue, such as ‘no claro’ (not clear). This exception will be further developed in the next section.

3.2.2. Uncertainty cues

Similarly to negation, uncertainty cues can be separated into two groups: syntactic cues and lexical cues.

Syntactic uncertainty cues. Again, these are function words. The only instances of this class are the disjunctions ‘or’ (or) and ‘vs’. These were only annotated when they appeared by themselves in a context of uncertainty\(^7\) as they could also appear listing alternatives or as a way to reformulate a sentence or phrase\(^8\).

(7) \textit{Una complicación postCNG o una patología de origen digestivo}
A post-coronary angiography complication or a pathology of digestive origin

(8) En las intercrisis refiere sensación continua de mareo o inestabilidad
[The patient] mentions continuous dizziness or instability

Lexical uncertainty cues. As with lexical negation, these are content words that express uncertainty depending on the context. Some of the most used cues are ‘probable’, ‘posible’ or ‘sospecha de’ (see\(^9\)\(^10\)). Verbs in the conditional mood are also treated as uncertainty cues, including those that usually act as negation cues, as in example\(^11\).

(9) \textit{Sospecha de dehiscencia de suturas}
Suspicion of wound dehiscence

(10) \textit{Se pensó en un origen funcional de ambos síntomas}
A functional origin of both symptoms was considered

(11) \textit{Descartaría (de forma razonable) una arteritis de la temporal como causa de la clínica}
It would reasonably rule out temporal arteritis as the origin of the symptoms

Seemingly negative cues can also express uncertainty depending on the context they appear in. For example, a negated negative cue might be used to express uncertainty\(^12\) while words that express confidence are also classified as uncertainty when they are negated\(^13\). When the latter happens, it might be the case that the cue is discontinuous\(^14\).

(12) \textit{No se descarta (definitivamente) sangrado activo}
Active bleeding is not definitively ruled out

(13) \textit{No claro transtorno sensitivo}
No clear sensitive disorder

(14) \textit{Sin signos claros de isquemia aguda}
No clear signs of severe ischemia

Finally, negation can also happen together with uncertainty in the same sentence. There are two possible scenarios. If an uncertainty cue appears within the scope of a negation, the latter usually invalidates the meaning of the uncertainty. For example, in\(^15\) ‘sugestiva de’ stops indicating that the speaker is unsure of what they say when it is negated by ‘no’. In such cases, the uncertainty cue is not annotated. However, if an uncertainty cue is the one that appears first and scopes over a negation cue, the meaning is maintained, as in example\(^16\). This time, the negation cue as well as its own scope are annotated inside the uncertainty’s scope.

(15) \textit{No refiere clínica sugestiva de aura migrañosa}
[The patient] does not allude to symptoms suggestive of migraine aura

(16) \textit{Sospecha de ( posible) HSA no apreciada en el TAC}
Suspicion of a possible subarachnoid hemorrhage not detected in the CT

3.2.3. Scopes

The scope is the part of the sentence whose meaning is changed by a negation or uncertainty cue. We follow IULA-SCRC’s definition of the scope as “the maximal syntactic unit that is affected by the marker” (Marimon et al., 2017, p. 46). In NUBes, coordinated items are included within the scope, but cues are not. Subjects are only included when they appear in post-verbal position.

As the scope is always the maximal syntactic unit, it is sometimes longer than the actual part that is most prominently affected by negation or uncertainty. For that reason, we also annotate events inside the scope, as in\(^17\).

(17) \textit{No se aprecian lesiones estructurales}
No structural lesions are observed

When a sentence contains multiple noun phrases coordinated, each of them is annotated as an individual event inside a bigger scope, as in example\(^18\). However, if the modifiers of the same noun phrase are separated by coordination, they are still all treated as part of the same event\(^19\).

(18) \textit{Sin aparente TCE ni focalidad}
With no apparent TBI or [neurological] focus

(19) \textit{No clínica digestiva ni miccional}
No digestive nor voiding symptoms

Events are labelled with a set of medical entity tags adapted from IULA-SCRC’s interpretation of the SNOMED-CT classification\(^6\): Medical findings and Disorders, Medical Procedures, Chemicals and Body Substances, Body Structure. Other—for medical concepts outside of the previous categories; not in IULA-SCRC— and Phrase –used for general scopes and entities outside of the medical field. If the event and the scope match in span, the most specific label is used for the whole scope. Otherwise, the event is annotated inside a longer Phrase label.

\(^6\)In this example, the scope of the embedded cue is marked with a dotted underline.

\[^{http://www.snomed.org/}\]
The scope of a cue can sometimes be discontinuous. That is, a cue can affect multiple text spans that are separated. The most frequent structures that trigger discontinuous scopes are the following: (a) the cue appears within the affected phrase [20] causing the cue to be surrounded by its scope; (b) the object of a verb has been omitted or substituted by a pronoun —in [21] “them” substitutes “inhalers” and thus the latter is annotated as being part of the scope; (c) there is ellipsis of the verb, as in [22] where the verb “repeats” is omitted in the second sentence as it has already been used before. Thus, the first mention is annotated as being part of the scope. Discontinuous scopes also happen frequently in combination with discontinuous cues, as in example [23].

(20) Relación posible con incipientes cambios por otitis media crónica
Probable relation to early changes caused by chronic otitis media

(21) Refiere su Médico de Cabecera que le pautó inhaladores pero no los tolera
Her family doctor refers that she gave him inhalers but he does not tolerate them

(22) Repite palabras sencillas pero no frases
[The patient] repeats simple words but not sentences

(23) No pudiendo precisar si ha presentado o no pérdida de conciencia
[The patient] is not able to specify whether they lost consciousness or not

Finally, elements expressing polarity changes can also appear inside the scope. They are elements that reinforce the expressive power of the phenomena. Usually, these are pronouns or negative determiners, such as ‘alguno’ or ‘ninguna’ (any), but multiple cues of the same type appearing together are also treated as such if they were used to reaffirm the meaning of the first cue.

(24) Niega dolor a ningún nivel
[The patient] denies pain at any level

(25) Parece detectarse un posible deterioro cognitivo de (possible) origen vascular
A possible cognitive impairment of possible vascular origin has seemingly been detected

Nevertheless, there are some rare cases where it is also possible for negation or uncertainty to appear embedded inside another negation cue’s scope without actually reinforcing the meaning of the first cue (e.g. if the second cue is part of a modifier of the negated event or the scope of the first cue is a long embedded clause). In such cases, both cues and their scope are annotated separately, as in example [26].

(26) Imposibilidad para una bipedestación sin ayuda
Inability to stand without help

| documents | 7,019 |
| sentences | 29,682 |
| tokens | 518,068 |
| vocabulary size | 31,698 |
| negation |  |
| sentences affected | 7,567 |
| average cues per affected sentence | 1.25 ± 0.66 |
| discontinuous cues | 123 |
| average scope size in tokens | 4.01 ± 3.59 |
| uncertainty |  |
| sentences affected | 2,219 |
| average cues per affected sentence | 1.12 ± 0.38 |
| discontinuous cues | 95 |
| average scope size in tokens | 5.27 ± 4.97 |
| discontinuous scopes | 123 |

Table 3: Size of NUBes

3.3. Dataset statistics
The part of the corpus that has been annotated so far consists of 29,682 sentences, out of which 7,567 (25.49%) include negation and 2,219 (7.48%) include uncertainty. A general overview of the size of NUBes is described in Table 3. In many of the sentences there is more than one cue, and both phenomena might appear together and/or independently. Discontinuous cues and scopes seem to be much more frequent for uncertainty than for negation.

The distribution of both phenomena over the different medical report sections follows the same pattern. Unsurprisingly, negation and uncertainty are more frequent in sections that tend to be longer (i.e., Progress Notes, Diagnostic Tests, and Present Illness). Their distribution does not fit into the same pattern, however, when analysed over medical specialties. Neurology reports stand out in particular for their high usage of speculative expressions. Negation, on the other hand, is most frequent in Cardiology, General Surgery, Neurology, and Internal Medicine. Concerning the different cues that appear in the corpus, 345 unique negation and 297 unique uncertainty cues have been annotated. The most frequent cues sorted by type are shown in Table 4.

4. Experiments with NUBes
A set of experiments have been conducted in order to ascertain the validity of NUBes and establish a competitive baseline on this corpus. The task evaluated has been automatic negation and uncertainty cue and scope labelling with the BIO scheme [Ramshaw and Marcus, 1999].

---

Please note that the aim of this work is not to find the best possible algorithm or features for automatic negation detection in Spanish. For literature specific to the topic, please refer to Santiso et al. (2019), Loharja et al. (2018), Fabregat et al. (2018) or Koza et al. (2019), among others.
Table 4: The 5 most common cues by type

| syntactic negation                  | freq. |
|------------------------------------|-------|
| **no** *(no, not)*                | 4,058 |
| **sin (without)*                   | 2,518 |
| tampoco *(neither)*                | 40    |
| nunca *(never)*                    | 5     |
| excepto *(except)*                 | 4     |

| lexical negation                    |       |
|------------------------------------|-------|
| negativo *(negative, sg.)*         | 123    |
| negativos *(negative, pl.)*        | 99     |
| retirada de *(withdrawal of)*      | 96     |
| niega *(denies)*                   | 83     |
| suspender *(withhold)*             | 59     |

| morphological negation              |       |
|------------------------------------|-------|
| afebril *(afreibile)*              | 252    |
| asintomático *(asymptomatic, m.)*  | 241    |
| asintomatía *(asymptomatic, f.)*   | 150    |
| inespecífico *(non-specific)*      | 39     |
| asintomático *(sic)*               | 34     |

| syntactic uncertainty              |       |
|------------------------------------|-------|
| vs                                 | 13     |
| o *(or)*                           | 4      |
| versus                             | 1      |
| vs.                                | 1      |

| lexical uncertainty                |       |
|------------------------------------|-------|
| probable                           | 357    |
| posible *(possible)*               | 198    |
| compatible con *(compatible with)* | 188    |
| sospecha de *(suspicion of)*      | 144    |
| parece *(seems)*                  | 130    |

Table 5: Size of the corpus subset used in the experiments

|                  | train | dev. | test |
|------------------|-------|------|------|
| sentences        | 13,802| 1,840| 2,762|
| negation cues     | 6,976 | 919  | 1,423|
| negation scopes   | 6,379 | 847  | 1,322|
| uncertainty cues  | 1,866 | 263  | 400  |
| uncertainty scopes| 1,886 | 260  | 400  |

4.1. Data

The dataset used contains all the sentences with at least one negation or uncertainty annotation (9,202) plus as many sentences with no annotations whatsoever. This dataset has been shuffled and split into train (75%), development (10%), and test (15%) sets. Table 5 shows the size of these splits. Marker and scope labels have been simplified to 4 generic categories: negation or uncertainty marker, and negation or uncertainty scope. Scopes have been flattened to the biggest span possible, thus ignoring events, coordination and polarity particles.

4.2. Methodology

We use NCRF++ *(Yang and Zhang, 2018)*, an open-source toolkit built upon PyTorch to develop neural sequence labelling architectures. Out-of-the-box network configuration and hyperparameters have been kept, except for the batch size (16), the learning rate (0.05) and learning rate decay (0.1). Several groups of input features at token level have been tested, namely:

- **form**: affixes of 2 and 3 characters, and whether the token is a punctuation mark, a number or an alphabetic string.
- **morphsyn**: the token’s lemma, its part-of-speech tag, the type of dependency relation, and the lemmas of the dependent children to the right and left, all extracted with spaCy’s es-core-news-md 2.2.0 model.
- **brown**: Brown cluster *(Brown et al., 1992)* complete paths and paths pruned at lengths 16, 32, and 64. The clusters were learned with tan-clustering from the training set and the 11,278 sentences left out from the dataset split.
- **metadata**: the speciality and section the sentence has been extracted from.
- **window**: all the features of the neighbouring tokens in a ±2 window.

An ablation study has been performed by withdrawing one group of features each time. We have also trained a model with just the tokens as features. In total, then, 7 systems have been trained: one with all the features available, another with just tokens as features, and one per –ablated– feature group. For each system, we have kept the model that has obtained the best F1-score against the development split within 40 epochs.

The experiment has been run 5 times with random seed initialisation. We report the mean and standard deviation of the micro-averaged precision, recall and F1-score of the 5 runs for each system. We have also computed the statistical significance per the Bootstrap test *(Efron and Tibshirani, 1994)* of a) the differences between the results of each run, and b) the difference between each ablation model and the base model that uses all the features available.

4.3. Results

The results of the negation and uncertainty detection are shown in Tables 6a and 6b, respectively. Overall, a sharp difference between negation and uncertainty detection can be observed. Unsurprisingly, negation detection seems to be an easier task than uncertainty detection –marker F1-score 95.0 vs 83.2; scope F1-score 90.6 vs 78.5–, which can be explained by the fact that we have more examples of the former case. Moreover, speculation cues and scopes are more likely to be discontinuous and the
variability of speculation cues is also higher, which adds difficulty to their correct identification.

Regarding the impact of the different groups of features, removing individual groups seems to have little effect, the differences not being statistically significant in most of the occasions. Notwithstanding, removing all features yields significantly worse results for all categories except negation marker detection. The difference is sharper, again, for uncertainty marker and scope detection, which seems to benefit more from the features, particularly of window features. All in all, the experiments show that the corpus is useful for training models for negation and speculation detection. Nevertheless, there is ample room to improve the results, specially of uncertainty detection.

4.4. Error analysis

Most of the errors involve post-scope and discontinuous markers. Although the cues are properly detected, their preceding scopes are not. This happens in a variety of structures, such as relative clauses [27], postnominal adjectives [28] and phrases formatted with colon: [29]. Note that the examples in this section show incorrect annotations made by the trained taggers:

(27) Sangrado que desaparece
Bleeding that goes away

(28) Control en heces negativo
Negative stool test

(29) Examen anatomopatológico: no
Anatomopathological examination: no

This type of error is reduced somewhat when exploiting all the features described. Moreover, additional experiments without the biLSTM architecture have proven that it is beneficial in this regard, although insufficient. Another frequent error arises from the incorrect capitalisation and/or punctuation in the input texts, which leads to scopes ranging beyond sentence boundaries. Co-occurrence of several cues—the same [30] or different [31] type—within a short text span also introduces errors. Uncertainty markers starting with ‘no’ or ‘sin’ [32] are particularly tricky.

(30) Parece poco probable que […]
It seems hardly likely that […]

(31) No parece identificarse ningún factor
No factor seems to be identified

(32) Sin focos claros
No clear foci

Finally, sources of less common errors include: markers that occur seldom in the corpus and thus are hard to detect automatically; long and complex scopes that involve coordination and/or subordination of several clauses; and the inclusion of a preposition or complementiser as being part of the marker instead of the scope, or vice versa.
5. Discussion

In this section, we report some of the main difficulties faced during the creation of the NUBES corpus. Annotating a corpus with extra-propositional meaning requires a thorough linguistic analysis that led to many discussions before, during and even after the process. Aspects like how to demarcate the definition of negation and uncertainty and whether some examples were actually part of them proved to be a source of disagreement. On top of that, the idiosyncrasies of medical language also posed some complications, mostly vocabulary-related.

The first step of the corpus creation process was to reach an agreement on what the terms negation and uncertainty encompass. An overview of the existing literature both in English and Spanish, revealed the there is not a clear-cut definition of the phenomenon across corpora. As a consequence, each corpus has been annotated with a different criterion. The main differences between them have to do with what is accepted as negation and the way in which elements such as scope are annotated.

We ultimately considered that our definition of negation (see Section 1) should also encompass every word that implies that an entity is not occurring or has not occurred: either at all (‘imposibilidad para’, (impossibility to)) or anymore (‘retrirada de’ (removal of), ‘suspender’ (withhold)). Some authors such as Marimon et al. (2017) argue that they did not take into account the cues we have just mentioned because they express a “change of state” (ibid.) or, in the case of ‘negativo’ (negative), that it “is considered, in factual terms, an statement of what someone says”. However, we consider that figuring out whether a statement is actually indirect speech or not is a different task.

Another debatable example is the postnominal adjective ‘negativo’ (negative). The authors of UHU-HUVR (Cruz Díaz et al., 2017) only annotate this word for test results whenever the name of the test and that of the condition is the same, as it means that the patient does not have said condition; otherwise, it means that the test has taken place and the results are simply negative. This contrast is shown respectively in examples (33) and (34) taken from UHU-HUVR.

(33) Serología materna: [Toxoplasma]: Negativo
Maternal serology: Toxoplasma: Negative

(34) Técnicas de Z-N (normal y largo) negativo
Negative Z-N stain (normal and long)

In NUBES, the latter case (34) is also annotated as it still accommodates into our definition of negation. The only exception is when ‘negativo’ is part of an entity’s name, e.g., ‘bacterias Gram negativas’ (Gram-negative bacteria). In spite of our broad definition of negation, not all negative occurrences have been annotated. Negative polarity verbs have only been considered when they appear in performative utterances. That is, conditional constructions (35) or final adverbs (37) have not been annotated:

(35) Si fiebre alta que no cede [...] If [they have] high fever that doesn’t drop [...]
Interestingly, these expressions are difficult to classify because they express uncertainty at different levels (e.g. from almost certain to completely unsure). We are considering expanding the annotations to include the different levels of confidence and uncertainty in the future to better deal with these cases.

Finally, some of the instances that are categorised as negation by other corpora were annotated as uncertainty in NUBES due to the inclusion of this phenomenon. For example, given the sequence ‘sin clara’ (no clear), IULA-SCRC annotates ‘sin’ as a cue and ‘clara’ as part of the scope. In NUBES, ‘sin clara’ as a whole is considered an uncertainty cue.

6. Conclusions and Future Work

In this paper we have presented the NUBES corpus, a new collection of biomedical texts in Spanish annotated for negation and uncertainty. As far as we know, NUBES is the largest corpus of clinical reports in Spanish annotated with negation and the first one including the annotation of speculation cues, scopes, and events. We have explored the corpus from different perspectives: by its comparison with similar corpora, by justifying its design and by explaining the challenges faced during its creation. Furthermore, preliminary experiments have been conducted with the corpus in order to ascertain its validity and establish a competitive baseline. NUBES, IULA+, as well as the annotation guidelines are publicly available from the web.¹¹

As part of on-going work, we expect to improve the quality of NUBES. At the moment, ~10% of the corpus has been annotated by three people, while the rest has been produced by a single annotator. Another line of future work includes performing different and more exhaustive experiments with NUBES, such as testing other sequence labelling algorithms and architectures, or exploiting the relations between cues and scopes.

7. Acknowledgements

This work has been supported by Vicomtech and partially funded by the project DeepReading (RTI2018-096846-B-C21, MCIU/AEI/FEDER,UE).

8. Bibliographical References

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992). Class-based n-gram models of natural language. Comput Linguist, 18(4):467–479.

Campillos Llanos, L., Martínez, P., and Segura-Bedmar, I. (2017). A preliminary analysis of negation in a Spanish clinical records dataset. In Taller de NEGación en Español (NEGES).

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. (2001). A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. J Biomed Inform, 34(5):301–10.

Cotik, V., Filippo, D., Roller, R., Uszkoreit, H., and Xu, F. (2017). Creation of an Annotated Corpus of Spanish Radiology Reports. In Proceedings of WinLNP 2017.

¹¹ https://github.com/Vicomtech/NUBes-negation-uncertainty-biomedical-corpus

Cruz Díaz, N. P., Morante Vallejo, R., Maña López, M. J., Mata Vázquez, J., and Parra Calderón, C. L. (2017). Annotating Negation in Spanish Clinical Texts. In Proceedings of SemBeAr 2017, pages 53–58.

Efron, B. and Tibshirani, R. (1994). An Introduction to the Bootstrap. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Fabregat, H., Martínez-Romo, J., and Araujo, L. (2018). Deep Learning approach for Negation Cues Detection in Spanish. In Proceedings of NEGES 2018, pages 43–48.

Horn, L. R. and Wansing, H. (2020). Negation. In The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Huang, Y. and Lowe, H. J. (2007). Research Paper: A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports. J Am Med Inform Assoc, 14(3):304–11.

Jiménez-Zafra, S. M., Morante, R., Martin, M., and Ureña-López, L. A. (2018a). A review of Spanish corpora annotated with negation. In Proceedings of COLING 2018, pages 915–924.

Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., and Martí, A. M. (2018b). SFU ReviewSP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. LREJ, 52(2):533–569.

Kahneman, D. and Tversky, A. (1982). Variants of uncertainty. Cognition, 11:143–157.

Koza, W., Filippo, D., Cotik, V., Stricker, V., Muñoz, M., Godoy, N., Rivas, N., and Martínez-Gamboa, R. (2019). Automatic Detection of Negated Findings in Radiological Reports for Spanish Language: Methodology Based on Lexicon-Grammatical Information Processing. J Digit Imaging, 32(1):19–29.

Lima, S., Perez, N., García-Sardiña, L., and Cuadros, M. (2019). HitzalMed: Anonymisation of Clinical Text in Spanish. In Proceedings of LREC 2020.

Loharja, H., Padró, L., and Turno, J. (2018). Negation Cues Detection Using CRF on Spanish Product Review Texts. In Proceedings of NEGES 2018, pages 49–54.

Marimon, M., Vivaldi, J., and Bel, N. (2017). Annotation of negation in the IULA Spanish Clinical Record Corpus. In Proceedings of SemBeAr 2017, pages 43–52.

Martí, A. M. and Taulé, M. (2018). Análisis Comparativo de los Sistemas de Anotación de la Negación en Español. In Proceedings of NEGES 2018, pages 23–28.

Morante, R. and Sporleder, C. (2012). Modality and Negation: An Introduction to the Special Issue. Comput Linguist, 38(2):223–260.

Moreno, A. and Garrote, M. (2013). La anotación de la negación en un corpus escrito etiquetado sintácticamente. RIL, 8:45–60.

Moreno, A., López, S., Sánchez, F., and Grishman, R. (2003). Developing a Syntactic Annotation Scheme and Tools for a Spanish Treebank, pages 149–163. Springer.

Oronoz, M., Gojenola, K., Pérez, A., Ilarraza, A., and Casillas, A. (2015). On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. J Biomed Inform, 56:318–332.
Ramshaw, L. A. and Marcus, M. P. (1999). *Natural Language Processing Using Very Large Corpora*, chapter 9, pages 157–176. Springer Netherlands.

Santiso, S., Casillas, A., Pérez, A., and Oronoz, M. (2019). Word embeddings for negation detection in health records written in Spanish. *Soft Computing*, 23:10969–10975.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). BRAT: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at EACL 2012*, pages 102–107.

Vincze, V., Szarvas, G., Farkas, R., Móra, G., and Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinform.*, 9(Suppl 11):S9.

Vincze, V. (2014). Uncertainty Detection in Hungarian Texts. In *Proceedings of COLING 2014*, pages 1844–1853.

Yang, J. and Zhang, Y. (2018). NCRF++: An Open-source Neural Sequence Labeling Toolkit. In *Proceedings of the System Demonstrations at ACL 2018*, pages 74–79.