

NEGES 2019 Task: Negation in Spanish

Salud María Jiménez-Zafra¹[0000-0003-3274-8825], Noa Patricia Cruz
Díaz²[0000-0002-6685-6747], Roser Morante³[0000-0002-8356-6469], and
María-Teresa Martín-Valdivia¹[0000-0002-2874-0401]

¹ SINAI, Computer Science Department, CEATIC, Universidad de Jaén, Spain
{sjzafra,maita}@ujaen.es

² Centro de Excelencia de Inteligencia Artificial, Bankia, Spain
contact@noacruz.com

³ CLTL Lab, Computational Linguistics, VU University Amsterdam, The
Netherlands
r.morantevallejo@vu.nl

Abstract. This paper presents the 2019 edition of the NEGES task, Negation in Spanish, held on September 24 as part of the evaluation forum IberLEF in the 35th International Conference of the Spanish Society for Natural Language Processing. In this edition, two sub-tasks were proposed: Sub-task A: “Negation cues detection” and Sub-task B: “Role of negation in sentiment analysis”. The dataset used for both sub-tasks was the SFU Review_{SP}-NEG corpus. About 13 teams showed interest in the task and 5 teams finally submitted results.

Keywords: NEGES 2019 · negation · negation processing · cue detection · sentiment analysis.

1 Introduction

Negation is a complex linguistic phenomenon that has been widely studied from a theoretical perspective [16, 17], and less from an applied point of view. However, interest in the computational treatment of this phenomenon is of growing interest, because it is relevant for a wide range of Natural Language Processing applications such as sentiment analysis or information retrieval, where it is crucial to know when the meaning of a part of the text changes due to the presence of negation. In fact, in recent years, several challenges and shared tasks have focused on negation processing: the BioNLP’09 Shared Task 3 [20], the NeSp-NLP 2010 Workshop: Negation and Speculation in Natural Language Processing [25], the CoNLL-2010 shared task [12], the i2b2 NLP Challenge [30], the *SEM 2012 Shared Task [24], the ShARe/CLEF eHealth Evaluation Lab 2014 Task 2 [27], the ExProM Workshop: Extra-Propositional Aspects of Meaning in Computational Linguistics [26, 6, 4] and the SemBEaR Workshop: Computational Semantics Beyond Events and Roles [5, 1].

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). IberLEF 2019, 24 September 2019, Bilbao, Spain.

However, most of the research on negation has been done for English. Therefore, the aim of NEGES task⁴ is to advance the study of this phenomenon in Spanish, the second most widely spoken language in the world and the third most widely used on the Internet. The 2018 edition consisted of three tasks related to different aspects of negation [18]: Task 1 on reaching an agreement on the guidelines to follow for the annotation of negation in Spanish, Task 2 on identifying negation cues, and Task 3 on evaluating the role of negation in sentiment analysis. A total of 4 teams participated in the workshop, 2 for developing annotation guidelines and 2 for negation cues detection. Task 3 had no participants. In this edition, the objective is to continue bringing together the scientific community that is working on negation to discuss how it is being addressed, what are the main problems encountered, as well as sharing resources and tools aimed at processing negation in Spanish.

The rest of this paper is organized as follows. The proposed sub-tasks are described in Section 2, and the data used is detailed in Section 3. Evaluation measures are introduced in Section 4. Participating systems and their results are summarized in Section 5. Finally, Section 6 concludes the paper.

2 Task description

In the 2019 edition of NEGES task, Negation in Spanish, two sub-tasks were proposed as a continuation of the tasks carried out in NEGES 2018 [18].

- **Sub-task A:** “Negation cues detection”
- **Sub-task B:** “Role of negation in sentiment analysis”

The following is a description of each sub-task.

2.1 Sub-task A: Negation cues detection

Sub-task A of NEGES 2019 had the aim to promote the development and evaluation of systems for identifying negation cues in Spanish. Negation cues could be *simple*, if they were expressed by a single token (e.g., “no” [no/not], “sin” [without]), *continuous*, if they were composed of a sequence of two or more contiguous tokens (e.g., “ni siquiera” [not even], “sin ningún” [without any]), or *discontinuous*, if they consisted of a sequence of two or more non-contiguous tokens (e.g., “no...apenas” [not...hardly], “no...nada” [not...nothing]). For example, in sentence (1) the systems had to identify four negation cues: i) the discontinuous cue “No...nada” [Not...nothing], ii) the simple cue “no” [no/not], iii) the simple cue “no” [no/not] again, and iv) the continuous cue “ni siquiera” [not even].

- (1) **No**¹ tengo **nada**¹ en contra del servicio del hotel, pero **no**² pienso volver, **no**³ me ha gustado, **ni siquiera**⁴ las vistas son buenas.
I have nothing against the service of the hotel, but I do not plan to return, I did not like it, not even the views are good.

⁴ <http://www.sepln.org/workshops/neges2019/>

Participants received a set of training and development data consisting of reviews of movies, books and products from the SFU Review_{SP}-NEG corpus[19] to build their systems during the development phase. At a later stage, a set of tests were made available for evaluation. Finally, the participant’s submissions were evaluated against the gold standard annotations. It should be noted that the data sets used in this sub-task were manually annotated with negation cues by domain experts, following well-defined annotation guidelines [19, 23].

2.2 Sub-task B: Role of negation in sentiment analysis

Sub-task B of NEGES 2019 proposed to evaluate the impact of accurate negation detection in sentiment analysis. In this task, participants had to develop a system that used the negation information contained in a corpus of reviews of movies, books and products, the SFU Review_{SP}-NEG corpus [19], to improve the task of polarity classification. They had to classify each review as *positive* or *negative* using an heuristic that incorporated negation processing. For example, systems should classify a review such as (2) as *negative* using the negation information provided by the organization, a sample of which is shown in Figure 1.

- (2) El 307 es muy bonito, pero no os lo recomiendo. Por un fallo eléctrico te puedes matar en la carretera.

The 307 is very nice, but I don’t recommend it. An electrical failure can kill you on the road.

```
<sentence complex="no">
  <d wd="el" postype="article" pos="da0ms0" name="d" lem="el" num="s" gen="m"/>
  <z wd="307" pos="z" name="z" lem="307"/>
  <v wd="es" postype="semiauxiliary" pos="vsip3s0" name="v" lem="ser" person="3" num="s" tense="present" mood="indicative"/>
  <r wd="muy" pos="rg" name="r" lem="muy"/>
  <a wd="bonito" postype="qualificative" pos="aq0ms0" name="a" lem="bonito" num="s" gen="m"/>
  <f wd="," pos="fc" name="f" lem="," punct="comma"/>
  <c wd="pero" postype="coordinating" pos="cc" name="c" lem="pero"/>
  <neg_structure polarity="negative" value="neg" change="yes">
    <scope>
      <negexp>
        <r wd="no" postype="negative" pos="rn" name="r" lem="no"/>
      </negexp>
      <p wd="os" postype="personal" pos="pp2cp000" name="p" lem="os" person="2" num="p" gen="c"/>
      <p wd="lo" postype="personal" pos="pp3cna00" name="p" lem="lo" person="3" num="n" gen="c" case="accusative"/>
      <event>
        <v wd="recomiendo" postype="main" pos="vmip3s0" name="v" lem="recomendar" person="1" num="s" tense="present" mood="indicative"/>
      </event>
    </scope>
  </neg_structure>
  <f wd="." pos="fp" name="f" lem="." punct="period"/>
</sentence>
<sentence>
  <s wd="Por" postype="preposition" pos="sps00" name="s" lem="por" complex="no"/>
  <d wd="un" postype="indefinite" pos="di0ms0" name="d" lem="uno" num="s" gen="m"/>
  <n wd="fallo" postype="common" pos="ncms000" name="n" lem="fallo" num="s" gen="m"/>
  <a wd="eléctrico" postype="qualificative" pos="aq0ms0" name="a" lem="eléctrico" num="s" gen="m"/>
  <p wd="te" postype="personal" pos="pp2cs000" name="p" lem="te" person="2" num="s" gen="c"/>
  <v wd="puedes" postype="main" pos="vmip2s0" name="v" lem="poder" person="2" num="s" tense="present" mood="indicative"/>
  <v wd="matar" postype="main" pos="vmm0000" name="v" lem="matar" mood="infinitive"/>
  <s wd="en" postype="preposition" pos="sps00" name="s" lem="en" complex="no"/>
  <d wd="la" postype="article" pos="da0fs0" name="d" lem="el" num="s" gen="f"/>
  <n wd="carretera" postype="common" pos="ncfs000" name="n" lem="carretera" num="s" gen="f"/>
  <f wd="." pos="fp" name="f" lem="." punct="period"/>
</sentence>
```

Fig. 1. Review annotated with negation information.

3 Data

The SFU Review_{SP-NEG} corpus⁵ [19] was the collection of documents provided for training and testing the systems in *Sub-task A* and *Sub-task B*⁶. This corpus is an extension of the Spanish part of the SFU Review corpus [29] and it could be considered the counterpart of the SFU Review Corpus with negation and speculation annotations [21].

The Spanish SFU Review corpus [29] consists of 400 reviews extracted from the website Ciao.es that belong to 8 different domains: cars, hotels, washing machines, books, cell phones, music, computers, and movies. For each domain there are 50 positive and 50 negative reviews, defined as positive or negative based on the number of stars given by the reviewer (1-2=negative; 4-5=positive; 3-star review were not included). Later, it was extended to the SFU Review_{SP-NEG} corpus [19] in which each review was automatically annotated at the token level with fine and coarse PoS-tags and lemmas using Freeling [28], and manually annotated at the sentence level with negation cues and their corresponding scopes and events. Moreover, it is the first Spanish corpus in which it was annotated how negation affects the words within its scope, that is, whether there is a change in the polarity or an increase or decrease of its value. Finally, it is important to note that the corpus is in XML format and it is freely available for research purposes.

3.1 Datasets Sub-task A

The SFU Review_{SP-NEG} corpus [19] was randomly splitted into development, training and test sets with 33 reviews per domain in training, 7 reviews per domain in development and 10 reviews per domain in test. The data was converted to CoNLL format [7] where each line corresponds to a token, each annotation is provided in a column and empty lines indicate the end of the sentence. The content of the given columns is:

- Column 1: domain_filename
- Column 2: sentence number within domain_filename
- Column 3: token number within sentence
- Column 4: word
- Column 5: lemma
- Column 6: part-of-speech
- Column 7: part-of-speech type
- Columns 8 to last: if the sentence has no negations, column 8 has a “***” value and there are no more columns. Else, if the sentence has negations, the annotation for each negation is provided in three columns. The first column contains the word that belongs to the negation cue. The second and third columns contain “-”.

⁵ <http://sinai.ujaen.es/sfu-review-sp-neg-2/>

⁶ To download the data in the format provided for Sub-task A and Sub-task B go to <http://www.sepln.org/workshops/neges2019/> or send an email to the organizers

Figure 2 and Figure 3, show examples of the format of the files with different types of sentences. In the first example (Figure 2) there is no negation so the 8th column is “***” for all tokens, whereas the second example (Figure 3) is a sentence with two negation cues in which information for the first negation is provided in columns 8-10, and for the second in columns 11-13.

```

hoteles_yes_4_1_9 1 Muy muy rg - ***
hoteles_yes_4_1_9 2 cómodo cómodo aq0ms0 qualificative ***
hoteles_yes_4_1_9 3 llegar llegar vmn0000 main ***
hoteles_yes_4_1_9 4 en en sps00 preposition ***
hoteles_yes_4_1_9 5 metro metro ncms000 common ***
hoteles_yes_4_1_9 6 ( ( fpa - ***
hoteles_yes_4_1_9 7 parada parada ncfs000 common ***
hoteles_yes_4_1_9 8 Sevilla sevilla np00000 proper ***
hoteles_yes_4_1_9 9 ) ) fpt - ***
hoteles_yes_4_1_9 10 o o cc coordinating ***
hoteles_yes_4_1_9 11 andando andar vmg0000 main ***
hoteles_yes_4_1_9 12 desde desde sps00 preposition ***
hoteles_yes_4_1_9 13 el el da0ms0 article ***
hoteles_yes_4_1_9 14 centro centro ncms000 common ***
hoteles_yes_4_1_9 15 . . fp - ***

```

Fig. 2. Sentence without negation in CoNLL format.

```

hoteles_yes_5_2_12 1 El el da0ms0 article - - - - -
hoteles_yes_5_2_12 2 hotel hotel ncms000 common - - - - -
hoteles_yes_5_2_12 3 es ser vsip3s0 semiauxiliary - - - - -
hoteles_yes_5_2_12 4 muy muy rg - - - - -
hoteles_yes_5_2_12 5 bueno bueno aq0ms0 qualificative - - - - -
hoteles_yes_5_2_12 6 , , fc - - - - -
hoteles_yes_5_2_12 7 tiene tener vmip3s0 main - - - - -
hoteles_yes_5_2_12 8 todas todo di0fp0 indefinite - - - - -
hoteles_yes_5_2_12 9 las el da0fp0 article - - - - -
hoteles_yes_5_2_12 10 comodidades comodidad ncfp000 common - - - - -
hoteles_yes_5_2_12 11 típicas típico aq0fp0 qualificative - - - - -
hoteles_yes_5_2_12 12 de de sps00 preposition - - - - -
hoteles_yes_5_2_12 13 un uno di0ms0 indefinite - - - - -
hoteles_yes_5_2_12 14 cuatro 4 z - - - - -
hoteles_yes_5_2_12 15 estrellas estrella ncfp000 common - - - - -
hoteles_yes_5_2_12 16 , , fc - - - - -
hoteles_yes_5_2_12 17 pero pero cc coordinating - - - - -
hoteles_yes_5_2_12 18 no no rn negative no - - - - -
hoteles_yes_5_2_12 19 es ser vsip3s0 semiauxiliary - - - - -
hoteles_yes_5_2_12 20 nuevo nuevo aq0ms0 qualificative - - - - -
hoteles_yes_5_2_12 21 , , fc - - - - -
hoteles_yes_5_2_12 22 así_que así_que cs subordinating - - - - -
hoteles_yes_5_2_12 23 no no rn negative - - - no - - -
hoteles_yes_5_2_12 24 está estar vmip3s0 main - - - - -
hoteles_yes_5_2_12 25 todo todo pi3ms000 indefinite - - - - -
hoteles_yes_5_2_12 26 reluciente reluciente aq0cs0 qualificative - - - - -
hoteles_yes_5_2_12 27 . . fp - - - - -

```

Fig. 3. Sentence with two negations in CoNLL format.

The distribution of reviews and negation cues in the datasets is provided in Table 1: 264 reviews with 2,511 negation cues for training the systems, 56 reviews with 594 negation cues for the tuning process, and 80 reviews with 836 negation cues for the final evaluation.

Table 1. Distribution of reviews and negation cues in the datasets of Sub-task A.

	Reviews	Negation cues
Training	264	2,511
Development	56	594
Test	80	836
Total	400	3,941

3.2 Datasets Sub-task B

For this sub-task, we provided the SFU Review_{SP}-NEG corpus [19] with the original format (XML). The meaning of the labels found in the reviews are the following:

- <review polarity=“positive/negative”>. It describes the polarity of the review, which can be “positive” or “negative”.
- <sentence complex=“yes/no”>. This label corresponds to a complete phrase or fragment thereof in which a negation structure can appear. It has associated the *complex* attribute that can take one of the following values:
 - “yes”, if the sentence contains more than one negation structure.
 - “no”, if the sentence only has a negation structure.
- <neg_structure>. This label corresponds to a syntactic structure in which a negation cue appears. It has 4 possible attributes, two of which (*change* and *polarity_modifier*) are mutually exclusive.
 - polarity: it presents the semantic orientation of the negation structure (“positive”, “negative” or “neutral”).
 - change: it indicates whether the polarity or meaning of the negation structure has been completely changed because of the negation (*change*=“yes”) or not (*change*=“no”).
 - polarity modifier: it states whether the negation structure contains an element that nuances its polarity. It can take the value “increment” if there is an increment in the intensity of the polarity or, on the contrary, it can take the value “reduction” if there is a reduction of it.
 - value: it reflects the type of the negation structure, that is, “neg” if it expresses negation, “contrast” if it indicates contrast or opposition between terms, “comp” if it expresses a comparison or inequality between terms or “noneg” if it does not negate despite containing a negation cue.
- <scope>. This label delimits the part of the negation structure that is within the scope of negation. It includes both, the negation cue (<negexp>) and the event (<event>).
- <negexp>. It contains the word(s) that constitute(s) the negation cue. It can have associated the attribute *discid* if negation is represented by discontinuous words.
- <event>. It contains the words that are directly affected by the negation (usually verbs, nouns or adjectives).

The distribution of reviews in the training, development and test sets is provided in Table 2, as well as the distribution of the different negation structures per dataset. The total of positive and negative reviews can be seen in the rows named as *+ Reviews* and *- Reviews*, respectively.

Table 2. Distribution of reviews and negation cues in the datasets of Sub-task B.

	Training	Dev.	Test	Total
Reviews	264	56	80	400
+ Reviews	134	22	44	200
- Reviews	130	34	36	200
neg	2,511	594	836	3,941
noneg	104	22	55	181
contrast	100	23	52	175
comp	18	6	6	30

4 Evaluation measures

The evaluation script used to evaluate the systems presented in **Sub-task A** was the same as the one used to evaluate the *SEM 2012 Shared Task: “Resolving the Scope and Focus of Negation” [24]. It is based on the following criteria:

- Punctuation tokens are ignored.
- A True Positive (TP) requires all tokens of the negation element have to be correctly identified.
- To evaluate cues, partial matches are not counted as False Positive (FP), only as False Negative (FN). This is to avoid penalizing partial matches more than missed matches.

The measures used to evaluate the systems were Precision (P), Recall (R) and F-score (F1). In the proposed evaluation, FN are counted either by the system not identifying negation cues present in the gold annotations, or by identifying them partially, i.e., not all tokens have been correctly identified or the word forms are incorrect. FP are counted when the system produces a negation cue not present in the gold annotations and TP are counted when the system produces negation cues exactly as they are in the gold annotations.

For evaluating **Sub-task B**, the traditional measures used in text classification were applied: P, R, F1 and Accuracy (Acc). P, R and F1-score were measured per class and averaged using macro-average method.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2PR}{P + R} \quad (3)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

5 Participants

13 teams showed interested and 5 teams submitted results.

Sub-task A had 4 participants: Aspie96 from the University of Turin, the CLiC team from Universitat de Barcelona, the IBI team from Integrative Biomedical Informatics group of Universitat Pompeu Fabra, and the UNED team from Universidad Nacional de Educación a Distancia (UNED) and Instituto Mixto de Investigación-Escuela Nacional de Sanidad (IMIENS). The official results by domain are shown in Table 3, and overall results are presented in Table 4, both evaluated in terms of P, R and F1. For IBI and UNED teams the domain in which it was most difficult to detect the negation cues was that of cell phones reviews, while for Aspie96 and CLiC it was the domain of hotels and books reviews, respectively. In terms of overall performance, the results of Aspie96 were quite low compared to the other teams. CLiC, IBI and UNED team obtained similar precision. However, the CLiC team achieved the highest recall, reaching the first rank position.

Table 3. Official results by domain for Sub-task A.

Domain	Aspie96			CLiC			IBI			UNED		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Books	16.00	28.57	20.51	80.59	75.79	78.12	80.97	72.62	76.57	84.02	81.35	82.66
Cars	19.42	29.41	23.39	94.92	82.35	88.19	92.73	75.00	82.93	94.83	80.88	87.30
Cell phones	18.07	26.32	21.53	87.76	75.44	81.13	90.48	66.67	76.77	88.37	66.67	76.00
Computers	17.36	25.93	20.80	90.48	93.83	92.12	89.06	70.37	78.62	94.12	79.01	85.91
Hotels	10.59	15.25	12.5	87.5	71.19	78.51	97.67	71.19	82.35	93.62	74.58	83.02
Movies	20.53	33.13	25.35	88.67	81.60	84.99	90.30	74.23	81.48	89.86	81.60	85.53
Music	24.17	33.33	28.02	94.44	78.16	85.53	94.20	74.71	83.33	95.38	71.26	81.57
Washing machines	24.24	34.78	28.57	92.98	76.81	84.13	94.34	72.46	81.96	94.34	72.46	81.96

Aspie96 [15] presented a model based in a convolutional Recurrent Neural Network (RNN) previously used for irony detection in Italian tweets [13] at IronITA shared task [8]. In order to address the task at NEGES, the system was modified to take tokens and Spanish spelling into account. Each word was represented using a 50-character window in which non-word tokens were also

Table 4. Overall official results for Sub-task A.

Team	P	R	F1
CLiC	89.67	79.40	84.09
UNED	91.82	75.98	82.99
IBI	91.22	72.16	80.50
Aspie96	18.80	28.34	22.58

considered. The words were then fed into a GRU layer to expand the context. The GRU layer’s output was fed to a classifier that classified each word as not part of a negation cue, the first word of a negation cue or part of the latest started negation cue. A similar model was shown to be suitable for the classification of irony [13] and factuality [14], but for negation it is not. The results of the task are quite low compared to other competing systems.

The CLiC team [3] developed a system based on the Conditional Random Field (CRF) algorithm, inspired in the system of *Loharja et al. (2018)* [22] presented in NEGES 2018 [18], which achieved the best results. They used as features the word forms and PoS tags of the actual word, the posterior word and the previous 6 words. They also conducted experiments including two post-processing methods: a set of rules and a vocabulary list composed of candidate cues extracted from an annotated corpus (NewsCom). Neither the rules nor the list of candidates boost basic CRF’s results during the development phase. Therefore, they presented to the competition the CRF model without post-processing, achieving the first position in the rank.

The IBI team [9] experimented with four supervised learning approaches (CRF, Random Forest, Support Vector Machine with linear kernel and XGBoost) using shallow textual, lemma, PoS tags and dependency tree features to characterize each token. For Random Forest, Support Vector Machine with linear kernel and XGBoost they also used the same set of features for the three previous and three posterior tokens in order to model the context of the token in focus. The highest performance during the development phase was the one grounded by the CRF approach. Therefore, they chosen it to support their participation, reaching the third rank position in the competition.

The UNED team [10] participated in the sub-task with a system based on Deep Learning, which is an evolution of the system presented in the previous edition of this workshop [11]. Specifically, they proposed a BiLSTM-based model using words, PoS tags and characters embedding features, and a one-hot vector to represent casing information. Moreover, they included in the system a post-processing phase in which some rules were used to correct frequent errors made by the network. The results obtained represent an improvement in relation to those of the 2018 edition of NEGES [18] and place them in the second position this year.

Sub-task B had 1 participant: LTG-Oslo from University of Oslo. The official results per sentiment class (positive and negative) and overall results are

shown in Table 5. The results for the positive class are better than those of the negative class and, overall, they do not give a strong performance in absolute numbers, but the proposed approach is very interesting. LTG-Oslo [2] addressed the task using a multi-task learning approach where a single model is trained simultaneously to negation detection and sentiment analysis. Specifically, shared lower-layers in a deep Bidirectional Long Short-Term Memory network (BiLSTM) were used to predict negation, while the higher layers were dedicated to predicting sentiment at document-level.

Table 5. Official results for Sub-task B.

LTG-Oslo				
	P	R	F1	Acc
Positive class	68.90	70.50	69.70	-
Negative class	62.90	61.10	62.00	-
Overall	65.90	65.80	65.85	66.20

6 Conclusions

This paper presents the description of the 2019 edition of NEGES task, whose aim is to continue working on advancing the state-of-the-art of negation detection in Spanish. Exactly, this edition consisted of 2 of the 3 sub-tasks carried out in the previous edition: Sub-task A: “Negation cues detection” and Sub-task B: “Role of negation in sentiment analysis”, in both using the SFU Review_{SP}-NEG corpus [19] to train and test the systems presented.

Compared to the previous edition, this year the workshop has attracted more attention, with more teams interested in participating in it (15 vs. 10). In addition, despite including one less sub-task, the number of submissions has been higher: In the 2018 edition of NEGES, a total of 4 teams participated in the workshop, 2 for developing annotation guidelines and 2 for cues detection. The task of studying the role of negation in sentiment analysis had no participants. This year, 5 teams submitted results, 4 for identifying negation cues and 1 for studying the role of negation in sentiment analysis. The low number of submissions in the last sub-task may be due to the fact that in order to study the impact of accurate negation detection in sentiment analysis it is necessary to determine how to efficiently represent negation, in the case of machine learning systems, or how to modify the polarity of the words within the scope of negation in the case of lexicon-based systems.

Regarding the approaches followed to detect negation cues, it seems that the teams continue to opt indistinctly for both more traditional machine learning approaches and deep learning algorithms, confirming that the use of Conditional Random Fields obtains the best results in this sub-task.

Concerning the system errors and difficulties encountered in the identification of negation cues, we can say the following. Aspie96 reported that the low results of its system could be due to the fact that only the text of the documents had been taken into account, without incorporating features such as the lemma and the PoS tags of the words, which could be of help. In fact, the other teams used them and obtained good results. The CLiC team reported several types of errors: errors in identifying negation cues that do not express negation (e.g. “Ya estaba casi, no (B)?” [*It was almost there, wasn't it?*]); not correctly identifying continuous cues (e.g. “a no ser que” [*unless*], “a excepción de” [*with the exception of*], “a falta de” [*in the absence of*]); tagging elements such as “tan” [*so*], “tanto” [*so much*], “muy” [*very*] or “mucho” [*much*] in discontinuous cues; and not detecting discontinuous cues. The IBI team detected that the performance of the approaches tested drastically decreases when they deal with multi-token negation cues. The UNED team also found it more difficult to identify multiple-term negation cues.

As for the difficulties and errors in the evaluation of the role of negation in sentiment analysis, LTG-Oslo stated that given the fact that the task is performed at the document level, it is difficult to determine them exactly. However, it is concluded that the multi-task model (MTL) is better than the single-task sentiment model (STL) for this sub-task and that the training size and different domains complicate the use of deep neural architectures.

Future editions of the workshop will also focus on detecting negation in other domains such as biomedical and studying other components of negation, such as the scope. Moreover, authors will have to include an error analysis of the results presented.

Acknowledgements

This work has been partially supported by a grant from the Ministerio de Educación Cultura y Deporte (MECD - scholarship FPU014/00983), Fondo Europeo de Desarrollo Regional (FEDER), and REDES project (TIN2015-65136-C2-1-R) and LIVING-LANG project (RTI2018-094653-B-C21) from the Spanish Government. RM is supported by the Netherlands Organization for Scientific Research (NWO) via the Spinoza-prize awarded to Piek Vossen (SPI 30-673, 2014-2019).

References

1. Proceedings of the Workshop on Computational Semantics beyond Events and Roles. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). <https://doi.org/10.18653/v1/W18-13>, <https://www.aclweb.org/anthology/W18-1300>
2. Barnes, J.: LTG-Oslo Hierarchical Multi-task Network: The Importance of Negation for Document-level Sentiment in Spanish. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (2019)

3. Beltrán, J., González, M.: Detection of Negation Cues in Spanish: The CLiC-Neg System. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (2019)
4. Blanco, E., Morante, R., Saurí, R.: Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics (exprom). In: Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM) (2016)
5. Blanco, E., Morante, R., Saurí, R.: Proceedings of the workshop computational semantics beyond events and roles. In: Proceedings of the Workshop Computational Semantics Beyond Events and Roles (2017)
6. Blanco, E., Morante, R., Sporleder, C.: Proceedings of the second workshop on extra-propositional aspects of meaning in computational semantics (exprom 2015). In: Proceedings of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics (ExProM 2015) (2015)
7. Buchholz, S., Marsi, E.: CoNLL-X shared task on multilingual dependency parsing. In: Proceedings of the tenth conference on computational natural language learning. pp. 149–164 (2006)
8. Cignarella, A.T., Frenda, S., Basile, V., Bosco, C., Patti, V., Rosso, P., et al.: Overview of the evalita 2018 task on irony detection in italian tweets (ironita). In: Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'18). pp. 26–34 (2018), <http://ceur-ws.org/Vol-2263/paper005.pdf>
9. Domínguez-Mas, L., Ronzano, F., Furlong, L.I.: Supervised Learning Approaches to Detect Negation Cues in Spanish Reviews. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (2019)
10. Fabregat, H., Duque, A., Martínez-Romo, J., Araujo, L.: Extending a Deep Learning approach for Negation Cues Detection in Spanish. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (2019)
11. Fabregat, H., Martínez-Romo, J., Araujo, L.: Deep Learning Approach for Negation Cues Detection in Spanish at NEGES 2018. In: Proceedings of NEGES 2018: Workshop on Negation in Spanish, CEUR Workshop Proceedings. vol. 2174, pp. 43–48 (2018)
12. Farkas, R., Vincze, V., Móra, G., Csirik, J., Szarvas, G.: The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task. pp. 1–12 (2010)
13. Giudice, V.: Aspie96 at IronITA (EVALITA 2018): Irony Detection in Italian Tweets with Character-Level Convolutional RNN. In: Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'18). pp. 160–165 (2018), <http://ceur-ws.org/Vol-2263/paper026.pdf>
14. Giudice, V.: Aspie96 at FACT (IberLEF 2019): Factuality Classification in Spanish Texts with Character-Level Convolutional RNN and Tokenization. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (2019)
15. Giudice, V.: Aspie96 at NEGES (IberLEF 2019): Negation Cues Detection in Spanish with Character-Level Convolutional RNN and Tokenization. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS, Bilbao, Spain (2019)

16. Horn, L.R.: A natural history of negation. CSLI Publications (1989)
17. Horn, L.: The expression of negation. De Gruyter Mouton, Berlin (2010)
18. Jiménez-Zafra, S.M., Cruz Díaz, N.P., Morante, R., Martín-Valdivia, M.T.a.: NEGES 2018: Workshop on Negation in Spanish. *Procesamiento del Lenguaje Natural* (62), 21–28 (2019)
19. Jiménez-Zafra, S.M., Taulé, M., Martín-Valdivia, M.T., Ureña-López, L.A., Martí, M.A.: SFU Review SP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation* **52**(2), 533–569 (2018)
20. Kim, J.D., Ohta, T., Pyysalo, S., Kano, Y., Tsujii, J.: Overview of bionlp’09 shared task on event extraction. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*. pp. 1–9. Association for Computational Linguistics (2009)
21. Konstantinova, N., De Sousa, S.C., Díaz, N.P.C., López, M.J.M., Taboada, M., Mitkov, R.: A review corpus annotated for negation, speculation and their scope. In: *Proceedings of LREC 2012*. pp. 3190–3195 (2012)
22. Loharja, H., Padró, L., Turmo, J.: Negation Cues Detection Using CRF on Spanish Product Review Text at NEGES 2018. In: *Proceedings of NEGES 2018: Workshop on Negation in Spanish, CEUR Workshop Proceedings*. vol. 2174, pp. 49–54 (2018)
23. Martí, M.A., Taulé, M., Nofre, M., Marsó, L., Martín-Valdivia, M.T., Jiménez-Zafra, S.M.: La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural* (57), 41–48 (2016)
24. Morante, R., Blanco, E.: * SEM 2012 shared task: Resolving the scope and focus of negation. In: *Proceedings of the First Joint Conference on Lexical and Computational Semantics*. pp. 265–274 (2012)
25. Morante, R., Sporleder, C.: Proceedings of the workshop on negation and speculation in natural language processing. In: *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*. pp. 1–109 (2010)
26. Morante, R., Sporleder, C.: Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics. In: *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics* (2012)
27. Mowery, D.L., Velupillai, S., South, B.R., Christensen, L., Martinez, D., Kelly, L., Goeuriot, L., Elhadad, N., Pradhan, S., Savova, G., et al.: Task 2: Share/clef ehealth evaluation lab 2014. In: *Proceedings of CLEF 2014* (2014)
28. Padró, L., Stanilovsky, E.: Freeling 3.0: Towards wider multilinguality. In: *Proceedings of LREC 2012*. Istanbul, Turkey (May 2012)
29. Taboada, M., Anthony, C., Voll, K.D.: Methods for creating semantic orientation dictionaries. In: *Proceedings of LREC 2016*. pp. 427–432 (2006)
30. Uzuner, Ö., South, B.R., Shen, S., DuVall, S.L.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association* **18**(5), 552–556 (2011)