ISSN 2007-9737

LiSSS: A New Multi-annotated Multi-emotion Corpus of Literary Spanish Sentences

Luis-Gil Moreno-Jiménez[1], Juan-Manuel Torres-Moreno[1,2]

[1] LIA, Université d'Avignon, Avignon, France

[2] Polytechnique Montréal, Montréal (Québec), Canada

luis-gil-moreno-jimenez@alumni.univ-avignon.fr, juan-manuel.torres@univ-avignon.fr

Abstract. In this work we present LiSSS, a new multi-annotated and multi-emotion corpus of Literary Spanish Sentences. We consider that this corpus may be strongly useful in the area of Computational Creativity (CC) to evaluate emotions classifier algorithms. A large number of literary sentences were manually classified by 12 annotators in five emotions: Love, Fear, Happiness, Anger and Sadness/Pain. Some classical classifiers were tested on this corpus. LiSSS corpus is available to the community as a linguistic free resource.

Keywords. Emotion corpus, Spanish literary corpus, linguistic resources.

1 Introduction

Researchers in Natural Language Processing (NLP) focused in the emotions classification, have been systematically left aside the studies of literary corpus for the development and evaluation of their models, mainly because the complex level of literary discourse. Instead, the use of corpora constituted by encyclopedic documents (mainly Wikipedia), journals (newspapers or magazines) or specialized (legal, scientific or technical documents) has been more frequently employed in recent years. [10, 2, 8]. In this work we introduce a new literary emotion corpus in order to evaluate and validate the NLP algorithms in the literary emotions classification tasks.

This paper is structured as follows. In Section 2 we show some works related to development and analysis of Spanish corpora. In Section 3 we describe the corpus LiSSS and in Section 4 the learning corpus *CitasIn*. The test and validation process are described in Section 5, as well as their respective results. Finally in Section 7, we propose some ideas for future works before to conclude.

2 Related Works

Several corpora in Spanish have been built and made available to the scientific community [3] however, a few number of them have been classified considering categories of emotions. For example, the corpus SAB composed by tweets in Spanish was introduced in [6]. These tweets represent critics toward different commercial brands. For each tweet, the perceived emotion must be indicated. The corpus SAB consists of 4 548 annotated tweets using 8 predefined emotions: {*Trust, Satisfaction, Happiness, Love, Fear, Disaffection, Sadness, Anger*}.

Another data set concerning tweets is the corpus TASS [11]. It contains about 70 000 tweets classified using automatic methods into the following categories: {*Positive, Negative, Neutral, None*}. Tweets in TASS corpus concern different topics: Politics, Economy, Sport, Music, etc.

A polarity emotion analysis (at word level) is described in [1]. The corpus employed was built with lexicons in 40 languages, annotated into the categories: {*Positive, Negative*}.

Our LiSSS corpus consists only of literary texts, which gives it a particular characteristic more useful for studying the algorithms of automatic emotion classification and generation of literary text. Moreover, for the classification, five categories of emotions were defined, instead of a binary (positive-negative) classification. This characteristic of LiSSS could be useful for more complete analysis.

3 LiSSS Corpus

LiSSS corpus is a small but well-controlled corpus, exclusively composed of literary sentences in Spanish selected from universal literature and tagged manually by a pull of annotators.

3.1 Corpus Annotation

The LiSSS corpus was constituted manually using literary texts in Spanish from around 200 Spanish-speaking authors. We also include not Spanish-speaking authors (keeping only official or good quality translations) in order to enrich the emotion content, the vocabulary and the expressive sense of the corpus.

This corpus is constituted by a P number of "sentences". Sentences are considered in a large sense. Actually, each "sentence" in this corpus is a complex linguistic object compound of one or several sentences, phrases or paragraphs. Henceforth in this paper, we will call **sentence** this linguistic object. The sentences were taken from quotes, stories, novels and poems. The literary genre is homogeneous. P sentences were classified by n annotators. All annotators in this study have a university level education and they are Spanish native speakers. Each sentence was read and manually classified into five categories: $\{$Anger (**A**), Love (**L**), Fear (**F**), Happiness (**H**), Sadness/Pain (**S**)$\}$.

Since the sentences may belong to one or more emotions, the annotators could tag the sentences using all perceived emotions. The sentences were manually processed to create n text and XML files, one per annotator. In the text version, each file contains P lines, with information structured in three fields:

```
ID      Sentence      # Author
```

Each field is separated by a tab character. The ID field is composed of a sequential number (1,2,3,...) followed by a code (A, L, F, H, S) corresponding to perceived emotions. In the XML version, the same structure is preserved using suitable XML tags.

If a sentence is considered as multi-emotion, it will have as many codes as categories it belongs to. The sentences were selected in order to maintain a balance between the categories, but this is not always guaranteed.

As mentioned, our "sentences" are often composed by many short-sentences, giving mini-paragraphs. This allows to respect as much as possible the coherence and the corresponding emotions. For example, sentence 455 of the emotion Fear (F), by J.P. Sartre:

```
Todos los hombres tienen miedo.  El que no
tiene miedo no es normal.  No tiene nada
que ver con el coraje.
```

is a three-sentence paragraph:

— *Todos los hombres tienen miedo* (All men are afraid).

— *El que no tiene miedo no es normal* (The one who is not afraid is not normal).

— *No tiene nada que ver con el coraje* (It has nothing to do with courage).

Approximately 10% of sentences of LiSSS corpus are mini-paragraphs. The multi-annotated corpus has currently $P = 500$ literary sentences, one per line. The support sentences (composed by non-literary language) as well as those too short ($N \leq 3$ words) or too long ($N \geq 50$ words) were ignored. A complex and aesthetic vocabulary where certain literary figures like anaphora or metaphor can be observed in this corpus. The characteristics of LiSSS corpus are shown in Table 1.

We detected some sentences that were tagged with opposed emotions. This phenomenon derived from ambiguity is commonly observed in this genre of texts, becoming a difficult task for classification. Authors like Leon Tolstoy are known by their strong emotional style: for example, Tolstoy often writes about the contiguity between love and death. To

Table 1. LiSSS corpus of literary sentences classified in 5 emotions

Sentences	Paragraphs	Words per file	Total Words	Spanish Speaking authors	Translated authors	Annotators + vote
500	49 (10%)	9 401	112 812	37	164	12+2

best interpreting this ambiguity, the classification performed by the annotators was compiling using two voting strategies. Therefore, two voting strategies may be computed in order to produce an integrated classification (see Section 5).

We tried a characterization of LiSSS corpus using a pull of artificial "annotators" for an experimental classification, i.e. using a set of artificial random taggers without knowledge of emotions, distribution nor textual content. The idea is to have an extended test corpus in order to measure the impact of performances of baseline algorithms. We called the artificial annotate corpus LiSSS/Art, and it will be used in the experiments of Section 5.

The LiSSS corpus has the advantage of being homogeneous in terms of genre, containing only "literary sentences", but it is heterogeneous in terms of emotions classes. In other emotion corpora, the sentences are overloaded of support sentences: linguistic structures that give a fluency to the reading and provide the necessary relations between ideas expressed in literary sentences, this is a disadvantage when it is pretended to analyse and process literary texts. The corpora composed by tweets are not ideal to be used with literary goals due to the presence of *noise* like: symbols, special characters, cut phrases, pasted words, wrong syntax, etc. This *noise* was avoided by a repeated and carefully reading from LiSSS corpus.

However, LiSSS corpus has the disadvantage of having a reduced number of sentences. It makes it not suitable for training algorithms based on Machine Learning (ML). But the goal of LiSSS corpus is not to be employed in the learning process but in testing the quality and performance of literary or emotions analysis algorithms.

3.2 Agreement

We have defined the agreement $c_i(k, e)$ as a triplet of $k = 1...n$ annotators a_k, $e_j, j \in \{A, L, F, H, S\}$ emotions and i sentences, $i = 1...P$, as follows:

$$c_{i,k}(e_j) = \begin{cases} 1 & \text{if all emotions } e_j \text{ are equal} \\ \frac{1}{n} & \text{if emotions } e_j \text{ are partially equal} \\ 0 & \text{elsewhere} \end{cases} \tag{1}$$

where $0 \leq c_i \leq 1$ is a value that represents the emotion agreement between the k annotations corresponding to the phrase i. 0 means no agreement, 1 means perfect agreement.

Considering $k = 1...n$ annotators,

$$\langle c_i \rangle = \frac{1}{n} \sum_{k=1}^{n} c_{i,k}, \tag{2}$$

we calculate the agreement mean weighed over all P sentences as:

$$C = \frac{1}{P} \sum_{i=1}^{P} \langle c_i \rangle; i = 1...P. \tag{3}$$

For example, if $n = 8$ humans, $a_{k=1...8}$ have annotated the sentence #75:

`75 En sustancia, es una misma cosa odio y amor. # Giordano Bruno` [1] as follows: a_1=ALS, a_2=AL, a_3=AL, a_4=AL, a_5=AL, a_6=AL, a_7=L, a_8=F. Considering the Equation (1), for each emotion we have: $\sum_A = 6$, $\sum_L$=7, $\sum_F$=1, $\sum_H$=0, $\sum_S$=1. Therefore: $\sum_A$=0.750, $\sum_L$=0.875, $\sum_F$=$\sum_S$=0.125 and $\sum_H$=0. The weighed vote will be **AL**, and the agreement (partial) between annotators: $c_{75}(\mathbf{AL})$ = partial agreement(A) + partial agreement(L) $\rightarrow c_{75}(\mathbf{AL}) = \frac{1}{5}_A + \frac{1}{5}_L = 0.4$.

[1] In essence, it's one and the same thing, love and hate.

Human Voters

We obtain an agreement mean value $C =$ **82.2**$\%$ computed with $n = 12$ annotators over $P = 500$ sentences. Table 2 shows the agreement among the annotators (the matrix values) and the annotators' agreement in relationship with voting strategy (the last column). We can see that annotator a_{11} differs from others: his lowest agreement value is less than 59.5%, with annotator a_{12}. The higher agreement value is obtained with annotators a_2 and a_{10}, getting 89.5%. The annotators with the highest vote agree are a_2 and a_4, having near 91% values. The "worst" annotators (in the sense of the agreement) are a_1, a_{11} and a_{12}. These information are used in a pilot test classification described in Section 5.

There are several sentences having overlapped emotions. They were tagged by the voting algorithm, processing all annotator's classification. The columns **A/x, F/x, H/x, L/x, S/x** in Table 3 represent single emotions vs overlapped emotions. For example, the voters have tagged 89.5 sentences only as **L** and 48.5 sentences combining **L** with other emotions **x**. The voters have tagged 153 sentences of corpus as multi-emotion. An example of this kind of sentences is the number 329, tagged with an identifier belonging to emotions Anger **A** and Love **L**:

```
329AL   Del amor al odio, solo hay mas
amor.    # Mario Benedetti
         (From love to anger, there's only
more love)
```

The matrix on Table 4 shows the mean class distribution calculated by dividing the numbers of sentences tagged for each emotion by the number of annotators. So we have $\approx$ 18% multi-emotion sentences in LiSSS corpus. Then, for each class we obtain the overlapping degree considering, a = the mean of sentences mono-class, and b = the mean of sentences multi-class, so we calculate $b/(a + b)$. This represents the fraction of sentences combining one class with the others. This ambiguity is mainly observed in the pair of emotions Happiness–Sadness/Pain and Love-Sadness/Pain, with an overlap of **HS**=18.7 and **LS**=19.9 multi-emotion sentences, respectively. Literary complexity and multi-emotion combined represent a challenge for classification algorithms.

Artificial Voters

For the LiSSS/Art corpus, we have an average agreement C_{Art} (Equation 3) computed over the A_k voters, $k = 1..15$ (20 random draws).

Since the artificial annotators are equally likely, it is not necessary to show the complete agreement matrix, but only their average value, as showed in Table 5. We computed an average agreement of 53.2% for the voting strategy, against the **85.2**$\%$ for vote of $n = 12$ human annotators. These values indicate that random annotators do not have a real consensus (they only agree on half of their vote), on the other hand, the humans show a strong consensus.

3.3 Voting Strategies

We have defined two voting strategies in order to compare reasonably the information furnished by n annotators. Therefore, the sentences have at least one emotion: the annotations without emotions selected are therefore avoided. The first one is a simple **majority vote** and the second one is a more elaborated **democratic vote**.

Majority Vote

This is a *winner-take-all* like strategy: the output is computed as the most weighting emotion. The output class is therefore always mono-label.

Democratic Vote

We fixed a threshold $t = 0.5$ (50%). We keep the emotion(s) selected by at least a fraction of n voters $\geq t$. We computed the output as follows: we calculate the probability $p(e) = Count(e)/n; e \in \{A, L, F, H, S\}$ over all voters. If there are one or more emotions e having $p(e) \geq t$, the process is finished and output is the concatenation of emotions having $p(e) > 0.5$.

If not, the threshold is down to $t = 0.3$ and the output is re-computed. Finally, if there are not emotions below this threshold, we down now $t = 0.2$ to re-compute the output. $t = 0.2$ seems to

Table 2. LiSSS agreement among annotators and voting (values are in %)

#	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	**Voting**
a_1	•	71.4	69.7	69.9	72.1	66.9	70.5	73.3	69.2	69.9	62.8	60.1	*72.0*
a_2	71.4	•	87.3	88.1	84.6	75.0	85.2	83.0	87.1	**89.5**	63.2	61.4	91.0
a_3	69.7	87.3	•	85.5	83.9	72.4	83.9	82.1	83.9	85.9	64.6	61.6	88.2
a_4	69.9	88.1	85.5	•	84.9	76.3	84.8	82.1	86.1	85.5	64.2	61.9	**91.1**
a_5	72.1	84.6	83.9	84.9	•	74.3	82.4	83.6	82.7	83.6	65.1	61.8	87.7
a_6	66.9	75.0	72.4	76.3	74.3	•	74.5	74.3	74.1	73.3	62.2	62.4	77.6
a_7	70.5	85.2	83.9	84.8	82.4	74.5	•	79.6	84.9	84.4	64.7	62.6	88.1
a_8	73.3	83.0	82.1	82.1	83.6	74.3	79.6	•	79.2	81.5	65.6	62.1	85.0
a_9	69.2	87.1	83.9	86.1	82.7	74.1	84.9	79.2	•	85.6	63.8	61.8	88.9
a_{10}	69.9	**89.5**	85.9	85.5	83.6	73.3	84.4	81.5	85.6	•	63.7	61.8	88.7
a_{11}	62.8	63.2	64.6	64.2	65.1	62.2	64.7	65.6	63.8	63.7	•	59.5	*65.2*
a_{12}	60.1	61.4	61.6	61.9	61.8	62.4	62.6	62.1	61.8	61.8	59.5	•	*63.2*
Mean	*71.3*	**81.3**	80.0	80.7	79.9	73.8	79.8	78.9	79.9	80.4	*66.6*	*64.8*	**82.2**

Table 3. Voted LiSSS corpus: distribution of single emotion vs multi-emotions

A/x	F/x	H/x	L/x	S/x	Overlap %
74.1/31.3	**103.2/31.4**	**92.4/32.4**	**89/48.5**	**115.3/64.8**	**33%**

be a suitable threshold in the hypothetical condition where a human have annotated a sentence with all possible emotion. In this case, each emotion e has at least a probability $p(e) = 0.2$. Using a democratic vote the output may be multi-labelled.

4 Experimental Setup

4.1 Classification Algorithms

The LiSSS corpus was tested using several classification algorithms available in Weka libraries[2]. In particular, we have employed:

1. Naïve Bayes, classical implementation,
2. Naïve Bayes Multinomial (NBM), oriented to textual classification,
3. Support Vector Machine (SVM), using a polynomial kernel with a multinomial logistic regression calibrator,
4. A mixture of the 3 precedent classifiers.

[2] https://www.cs.waikato.ac.nz/ml/weka/

We decided to use the Naïve Bayes model because its wide implementation in several classification works. A different implementation of Naïve Bayes known as Multinomial Naïve Bayes was performed due to its high precision score on text mining tasks, considering the estimated frequency of terms [9]. We also tested with a standard implementation of SVM[3] to compare the performance of models on the LiSSS corpus.

These algorithms need to be trained to produce a classification model. The training process must be performed with an independent corpus from the test corpus. We decided to build a learning corpus suitable for this task, adapting it to the five categories of the LiSSS corpus.

4.2 Learning Corpus

For the training process, we built a learning corpus. The `https://citas.in` website contains several thousand documents in Spanish, (mostly literary documents), sentences, paragraphs, quotes, phrases, etc. A large number of documents

[3] https://weka.sourceforge.io/doc.stable/weka/classifiers/functions/LibSVM.html

Table 4. Voted LiSSS corpus: Multi-emotion mean distribution (in %)

Emotion	A	L	F	H	S	Overlap %
A	74.1	12.1	7.9	2.3	9.0	29.7
L	12.1	89.5	5.7	10.8	19.9	27.7
F	7.9	5.7	103.2	0.6	17.2	35.5
H	2.3	10.8	0.6	92.4	18.7	38.9
S	9.0	19.9	17.2	18.7	115.3	35.9

Table 5. LiSSS/Art agreement between artificial annotators (values are in %)

#	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	**Vote**
Mean	59.5	59.2	59.1	59.2	59.1	59.8	59.3	59.8	59.7	59.7	59.5	59.4	59.4	59.1	59.8	**53.2**

belonging to different categories[4] (friendship, lovers, beauty, success, happiness, laughter, enmity, deception, anger, fear, etc.) were recovered from this website[5]. Documents were manually clustered into the five classes of the LiSSS corpus, from their own categories (last column of Table 6).

The result is *CitasIn* corpus, with an adequate size to be used on training phase for classifiers[6].

Table 6 shows some features of *CitasIn* corpus.

We pre-processed the *CitasIn* corpus before the training phase. The texts were coded in utf-8 format, we removed the special symbols, as well as the stop words using the Weka libraries and stop lists. We normalized the words by transforming the capital letters into small letters. Finally, a tokenisation process was applied using Weka. Of course, for the learning process, we have eliminated from the *CitasIn* corpus, the common sentences with the LiSSS corpus.

5 Results and Discussion

We characterized the LISSS corpus through two different experiments. The first one is a test using the LiSSS corpus with the classes defined by the two voting strategies considering the classification of all annotators. The second one is a pilot test using a sub-set of LiSSS corpus, where atypical annotators were suppressed in order to measure the impact of these inconsistent classifications. In both cases, the *CitasIn* was employed as learning corpus. We measured classical Precision, Recall and F-score values[8].

5.1 Classification using All Annotators

We experimented using the algorithms presented in Section 4 to validate the performances of human annotators and the performance of an artificial "mean" annotator (the mean output of 15 artificial annotators) Section 3. The LiSSS corpus, re-annotated by the two voting strategy ($n = 12$ humans annotators) was employed as test corpus. Table 7 shows the average F-score obtained for each human annotator and the mean of artificial annotators, taking as references the democratic and majority vote.

We can see that is more difficult to obtain an agreement between annotators (human or artificial) using the democratic vote. Therefore, the test was realized using only the majority class output (the majority emotion label per sentence). Table 8 shows the average F-score obtained for classifier algorithms on each emotion.

The best result was obtained by NBM algorithm with F-score=59.84. This seemingly mediocre result, shows the difficulty of classifying emotions

[4] `https://citas.in/temas/`

[5] All documents were downloaded, with the editor authorization, on March 25, 2020.

[6] A version of *CitasIn* corpus with snippet sentences is available in our website[7]. The reader should not have problem reconstituting the corpus *CitasIn* using these snippets and the correspondence between class.

[8] An harmonic combination of Precision p and Recall R: F-score=$2 \cdot p \cdot R/(p + R)$.

Table 6. *CitasIn*: Sentences from several categories mapped into 5 emotions

CitasIn	Sentences	Words	Words per sentence	Categories
Emotion	**72 790**	**1 352 810**	**18.6**	`https://citas.in/temas/`
L	14 738	264 339	29.2	alma, amantes, amistad, amor, belleza, beso, esperanza, pasión
H	13 647	256 697	18.8	felicidad, amistad, diversión, sonrisa, risa, motivación, victoria, éxito, optimismo
A	15 043	280 784	18.7	egoísmo, enemistad, engaño, envidia, venganza, guerra, infierno, mentira, guerra, odio, muerte, infierno, mentira
F	14 773	275 059	18.6	necesidad, miedo, dolor, fracaso indecisión, problema, soledad, suicidio
S	14 589	275 931	18.9	despedida, tristeza, pena, enfermedad, fracaso, pérdida, sufrimiento, olvidando, llorar, lágrima

in literary corpus. We detected two main problems in the classification of this type of texts. Firstly, the complexity of lexicon appreciated in the corpus. Secondly, the ambiguity: the mass of 30% of multi-emotion sentences provokes confusion in classification methods. This behaviour can be proved observing the results for *Sadness* emotion, with the higher overlapping (33.03%) score, and the lowest F-scores, between 4.71 for Naïve Bayes and 16.33 for SVM. Finally, the mixture of algorithms, NB+SVM+NBM obtains the second best performance, with a mean F-score value = 54.51.

5.2 Pilot Test using Selected Combinations of Annotators

For the second experiment, the idea was to verify how much the classification results could be altered eliminating the annotators having the lowest or the highest agreement values in the voting strategy. However, we think that there are not "bad" or "good" human annotators in this subjective classification task, but only consistent or inconsistent emotion perceptions. Also, we study the impact on the performance for classification algorithms using the LiSSS/Art artificial annotated corpus (see Section 3).

Therefore, we have employed 3 supplementary test corpora. The first one, LiSSS/– excludes the 3 annotators having the lowest agreement on the vote (a_1, a_{11} and a_{12}).

The second one, LiSSS/+, excluding the 3 annotators having the highest agreement on the vote (a_2, a_4 and a_9); and the last corpus LiSSS/Art, corresponds to voting strategy using all 15 artificial annotators.

In this experiment, we tested only the algorithm that obtained the best F-score performance, i.e. the NBM algorithm (see Table 8). The results per emotion are showed in Table 9.

It could be observed that suppression of "inconsistent" annotators (LiSSS/– test) impact slightly the F-score of NBM algorithm (it pass from 59.79 to 59.83) and emotions **A** and **L** are slightly best classified. On the other hand, the suppression of "consistent" annotators will fall the performances to 58.41 (LiSSS/+ test).

Finally, the F-score performances measured on LiSSS/Art corpus are the lowest of all experiments, as expected. These results confirm the real complexity of this classification task, and also that could be a good idea, to verify the annotators' agreement in order to constitute a more coherent testing set.

Table 7. LiSSS F-score performance of humans and artificial annotators

Annotator	Majority vote	Democratic vote
a1	62.71	31.85
a2	**93.41**	57.93
a3	89.96	50.23
a4	*92.86*	**73.57**
a5	85.92	65.28
a6	79.11	45.62
a7	88.16	64.15
a8	82.56	64.62
a9	90.49	*69.74*
a10	89.04	56.16
a11	49.24	21.75
a12	52.57	18.00
⟨HUMAN⟩	**79.64**	**51.57**
⟨ARTIFICIAL⟩	**24.76**	**7.08**

Table 8. voted LiSSS corpus: Evaluation of F-score models' performance per class (majority vote)

Algorithm	A	F	H	L	S	F-score mean
SVM	**55.59**	55.34	57.44	50.65	4.71	44.74
NB	43.08	60.08	53.68	60.29	16.33	46.69
NBM	54.05	**66.34**	**77.78**	67.73	**33.03**	**59.79**
NB+SVM+NBM	51.76	60.43	67.35	**70.05**	22.68	54.45

6 Corpus Availability

The version 0.5x12 of LiSSS corpus (distributed in several files encoded utf8, Linux EOL, $n = 12$ annotators) is available on our website[9] under GPL3 public license:

— n files containing: ID emotion codes, the sentence and the author in plain text and XML.

— $2n$ files POS tagged (2 formats) of annotated files using Freeling 4.1.

— 2 files containing the output of n voters (democratic and majority vote): ID, emotion(s), sentence and author in plain text and XML.

— 4 files containing the POS tagged version of votes output using Freeling 4.1.

[9] http://juanmanuel.torres.free.fr/corpus/LiSSS/

7 Conclusion and Future Work

We have introduced the LiSSS corpus, a new multi-annotated and multi-emotion literary corpus in Spanish. The manual multi-classification have allowed to establish a suitable voting strategy. The results obtained show that the multi-emotion classification of this kind of documents is a very difficult task (for both machines and humans): the low F-score value of annotators in the democratic vote ($\approx 51\%$) seems to confirm it.

We have tested some classical classifiers on the LiSSS corpus. The sentences often belong to two or more classes. The overlap between the sentences of the different classes prevents the systems a better classifying of this literary corpus. We think that automatic classifiers could be enriched through the integration of linguistic and stylistic characteristic or rich representations like word embedding, to achieve a better classification [12, 4, 7, 5]; but this study is out of scope of

Table 9. F-score values obtained by NBM algorithm on pilot corpora (majority vote)

Test corpus	A	F	H	L	S	F-score mean
LiSSS/–	53.33	66.67	77.42	67.21	34.55	**59.83**
LiSSS/+	52.17	65.69	73.10	71.04	30.08	*58.41*
LiSSS/Art	30.27	27.68	16.22	9.33	9.84	18.66

this paper. The purpose of the LiSSS corpus is *to evaluate* the efficiently of classification and ML algorithms on a specialized corpus, *not to train* such-as algorithms.

Future work must be accomplished in order to enrich the corpus with a more important number of sentences and more annotators. The scientific community can contribute to modify or distribute this corpus under the GPL3 license.

Acknowledgements

This work is funded by Consejo Nacional de Ciencia y Tecnología (Conacyt, Mexico), grant number 661101 and partially by the Université d'Avignon/Laboratoire Informatique d'Avignon (LIA), France. We thank the admin of the site `https://cite.in` for the database for our experiments. Also, authors thank Carlos-Emiliano González-Gallardo for their comments and all anonymous annotators that have participated in this project.

References

1. **Chen, Y. & Skiena, S. (2014).** Building sentiment lexicons for all major languages. *52nd Annual Meeting of the ACL*, volume 2, pp. 383–389.
2. **da Cunha, I., Cabré, M. T., SanJuan, E., Sierra, G., Torres-Moreno, J.-M., & Vivaldi, J. (2011).** Automatic specialized vs. non-specialized sentence differentiation. *CICLing*, pp. 266–276.
3. **da Cunha, I., Torres-Moreno, J.-M., & Sierra, G. (2011).** On the development of the RST Spanish treebank. *5th Linguistic Annotation Workshop*, ACL, pp. 1–10.
4. **Edalat, A. (2017).** Self-attachment: A holistic approach to computational psychiatry. In **Érdi, P., Sen Bhattacharya, B., & Cochran, A.**, editors, *Computational Neurology and Psychiatry. Series in Bio-/Neuroinformatics*, volume 6. Springer, pp. 273–314.
5. **Moreno-Jiménez, L.-G., Torres-Moreno, J.-M., & Wedemann, R. (2020).** Literary natural language generation with psychological traits. *25th NLDB*. Accepted.
6. **Navas-Loro, M., Rodríguez-Doncel, V., Santana-Pérez, I., & Sánchez, A. (2017).** Spanish corpus for sentiment analysis towards brands. *ICSP*, pp. 680–689.
7. **Siddiqui, M., Wedemann, R. S., & Jensen, H. J. (2018).** Avalanches and generalized memory associativity in a network model for conscious and unconscious mental functioning. *Physica A: Statistical Mechanics and its Applications*, Vol. 490, pp. 127–138.
8. **Sierra, G. (2018).** *Introducción a los Corpus Lingüísticos*. UNAM Mexico.
9. **Su, J., Shirab, J. S., & Matwin, S. (2011).** Large scale text classification using semi-supervised multinomial naive Bayes. *ICML-11*, pp. 97–104.
10. **Torres-Moreno, J.-M. (2014).** *Automatic Text Summarization*. Wiley, London.
11. **Villena-Román, J., Lana-Serrano, S., Martínez-Cámara, E., & González-Cristóbal, J. C. (2013).** TASS - workshop on sentiment analysis at SEPLN. *PLN*, Vol. 50, No. 0, pp. 37–44.
12. **Wedemann, R. S. & Carvalho, L. A. V. d. (2012).** Some things psychopathologies can tell us about consciousness. In **Villa, A. E. P., Duch, W., Érdi, P., Masulli, F., & Palm, G.**, editors, *ICANN 2012*, volume 7552. Springer, Heidelberg, pp. 379–386.

Article received on 15/06/2020; accepted on 22/07/2020. Corresponding author is Juan-Manuel Torres-Moreno.