

# Overview of the Task on Multimodal Stance Detection in Tweets on Catalan #1Oct Referendum

Mariona Taulé<sup>1</sup> [0000-0003-0089-940X], Francisco Rangel<sup>2,3</sup> [0000-0002-6583-3682], M. Antònia Martí<sup>1</sup> [0000-0002-8849-6826], and Paolo Rosso<sup>2</sup> [0000-0002-8922-1242]

<sup>1</sup> CLiC-UBICS, Universitat de Barcelona, Spain  
{mtaule, amarti}@ub.edu,

<sup>2</sup> PRHLT Research Center, Universitat Politècnica de València, Spain  
proso@dsic.upv.es

<sup>3</sup> Autoritas Consulting, S.A., Spain  
francisco.rangel@autoritas.es

**Abstract.** In this paper we present the MultiModal Stance Detection in tweets on Catalan #1Oct Referendum (MultiStanceCat) task at IberEval 2018 evaluation campaign. We describe the datasets created, the way in which they have been annotated, the evaluation metric used, the results obtained by the submitted approaches, and an evaluation of these approaches. The systems presented here had to detect from a multimodal perspective the authors stances -in favor, against or neutral- with respect to the Catalan first of October Referendum (2017) in tweets written in Spanish and Catalan. Four teams participated, submitting a total of seven runs in Catalan and nine in Spanish. The results obtained show that the performance of these systems improved when the context information was taken into account.

**Keywords:** Stance detection, Twitter, Multimodality, Catalan, Spanish.

## 1 Introduction

The aim of the MultiModal Stance Detection in tweets on Catalan #1Oct Referendum task at IberEval 2018 (*MultiStanceCat*)<sup>4</sup> is to detect the authors stances -in favor, against or neutral- with respect to the Catalan first of October Referendum (2017) in tweets written in Spanish and Catalan from a multimodal perspective. This is a multimodal task because both the text of the tweet and the information included in the link (when this information occurs), as well as the images from the authors timelines, are taken into account when determining their stance.

The task of stance detection is related to sentiment analysis, but unlike the latter, in which the systems detect the positive, negative or neutral polarity of the

<sup>4</sup> <http://www.autoritas.net/MultiStanceCat-IberEval2018/>

text, in stance detection the systems detect whether a text message is favorable or unfavorable to a topic of discussion, which is usually controversial, and which may or may not be explicitly mentioned in the text message [3, 9, 5, 6]. The stance detection task is also related to textual inference since a tweeter’s position is often expressed implicitly; and must be inferred in many cases. Stance detection is particularly interesting for studying controversial political debates. Therefore, for this task, we decided to focus on a specific, controversial political issue: the Catalonia self-determination Referendum (2017), which, in a nutshell, was approved by the Catalan Parliament, convened by the Generalitat of Catalonia on 6 September 2017, declared illegal by the Spanish Government the day after and later suspended by the Constitutional Court of Spain<sup>5</sup>. Therefore, there was a heated debate between those in favor and those against the referendum, which was considered to be legitimate for the former and an illegal referendum for the latter.

Stance detection in microblogging texts was carried out for the first time on texts in English at the SemEval-2016<sup>6</sup> workshop [7]. This task was performed for the first time on texts written in Catalan and Spanish, *StanceCat* task<sup>7</sup>, at the IberEval-2017 workshop -with the participation of 10 groups from 5 countries who performed a total of 31 runs [9]. The innovative aspect of the current task is its use of multimodality for detecting the stance of the tweet and, concretely, the use of images from the author’s timeline. Another important difference with respect to the IberEval 2017 task is that contextual information is also provided (i.e., the tweet before and after the tweet under evaluation are also taken into account). Therefore, more information is considered in order to detect the stance of the author’s message, based on the assumption that the more information, the better the results. The results obtained will be of interest not only for sentiment analysis but also for socio-political studies.

The rest of the overview is structured as follows. In Section 2 the task description is given. In Section 3 the dataset is described together with the way it was annotated and the measures used in the evaluation. In Sections 4 and 5 we describe the approaches and discuss the results. Finally, in Section 7 we draw some conclusions.

## 2 Task Description

The aim of the *MultiStanceCat* task is to detect the authors stances with respect to the Catalan 1st October Referendum (2017) in tweets written in Spanish and Catalan from a multimodal perspective, that is, taking into account both textual and image information. Given a message, the aim of the task is to determine the stance and to assign one of three possible, mutually exclusive tags: FAVOR, AGAINST and NEUTRAL. Participants are allowed to send up to five runs depending on the information used to assign the stance. That is, the systems can

<sup>5</sup> [https://en.wikipedia.org/wiki/Constitutional\\_Court\\_of\\_Spain](https://en.wikipedia.org/wiki/Constitutional_Court_of_Spain)

<sup>6</sup> <http://alt.qcri.org/semeval2016/task6/>

<sup>7</sup> <http://stel.ub.edu/Stance-IberEval2017/>

take into account: a) only the information appearing in the tweet under evaluation; b) the information included in the message and the contextual information, and c) the message and contextual information, as well as images downloaded from the authors timeline. All this information is provided in different files to the participants in the task (see Section 3.4). In addition to this information, participants can use any other type of resources for developing their systems.

Regarding the language, the *MultiStanceCat* task was carried out separately for Spanish and Catalan. The systems do not need to detect the language, because tweets are provided in two different datasets, one including the Catalan tweets and the other the Spanish tweets. Although we encouraged the teams to address the full multilingual task, the task could be performed for only one language or for both languages, Catalan or/and Spanish.

### 3 Evaluation Framework

In this section, we present the building of the *TW-10 Referendum* corpus: how it was collected, annotated and distributed.

#### 3.1 The TW-10 Referendum corpus

We used the #1oct, #1O, #1oct2017 and #1oct16 hashtags to select the tweets to be included in the *TW-10 Referendum* corpus. These hashtags were the most widely used (especially the first two) in the debate on the right to hold a unilateral referendum on Catalan independence from Spain.<sup>8</sup> A total of 87,449 tweets in Catalan and 132,699 tweets in Spanish were collected<sup>9</sup> from 20 September to the day before the Referendum was held on (30 September 2017). We used these data to build the *TW-10Referendum* corpus, which consists of 11,398 tweets-5,853 written in Catalan (the *TW-10Referendum\_CA* corpus) and 5,545 written in Spanish (the *TW-10Referendum\_ES* corpus). 80% of the *TW-10 Referendum* corpus was used for training purposes, while the remaining 20% was used for testing.

#### 3.2 Annotation Scheme

The possible stance labels are **FAVOR**, **AGAINST** and **NEUTRAL**, where:

- **FAVOR** identifies tweets with a positive stance towards the 1st October Referendum (examples (1) and (2)).
- **AGAINST** indicates tweets with a negative stance towards the 1st October Referendum (examples (3) and (4)).

<sup>8</sup> Concretely, the question asked in the Catalonia self-determination Referendum 2017 was: "Do you want Catalonia to be an independent country in the form of a republic?" and two possible answers were available 'yes' or 'no'.

<sup>9</sup> The dataset was collected with the Cosmos tool by Autoritas (<http://www.autoritas.net>).

- **NEUTRAL** indicates tweets with a neutral stance towards the 1st October Referendum (5), including basically informative or reporting tweets (that is, tweets that limit themselves to reporting what other people or media say), and tweets in which the stance cannot be inferred (6).

Below, we present some examples of annotated tweets:

1. **Tweet:** Res ni ningú, ens aturarà #Votarem #DretaDecidir #1Oct #CatalunyaLliure #defensemlademocracia <http://t.co/PgVLYH8AgN>  
 Language: Catalan  
 Stance: FAVOR  
 'Nothing and nobody will stop us #Votarem #DretaDecidir #1Oct #CatalunyaLliure #defensemlademocracia <http://t.co/PgVLYH8AgN>'
2. **Tweet:** Mientras tanto en #España se espera una REPRESIÓN para todo público este #1Oct Tan democráticos ellos... <https://t.co/gw7QIfrijHk>  
 Language: Spanish  
 Stance: FAVOR  
 'Meanwhile in #España a REPRESSION is expected by the general public this #1Oct Very democratic them... <https://t.co/gw7QIfrijHk>'
3. **Tweet:** Adeu #1octubreARV #1octubrenovotare <http://t.co/x3dXO3v7np>  
 Language: Catalan  
 Stance: AGAINST  
 'Bye bye #1octubreARV #1octubrenovotare <http://t.co/x3dXO3v7np>'
4. **Tweet:** Más q votos creo q estais usando personas jugando con sus sentimientos SABIAIS q el #1Oct ES ILEGAL <https://t.co/1SJewn7LHd>  
 Language: Spanish  
 Stance: AGAINST  
 'You know that more than votes you are using persons playing with their sentiments YOU KNOW that the #1Oct IS ILLEGAL <https://t.co/1SJewn7LHd>'
5. **Tweet:** Voteu! #1Oct Crees que la respuesta del Estado al desafío independentista catalán está siendo adecuada? <https://t.co/LlZrkd20gh> vía @20m  
 Language: Catalan+Spanish  
 Stance: NEUTRAL  
 'Vote! #1Oct Do you think that the States response to the Catalan independence challenge is appropriate? <https://t.co/LlZrkd20gh> va @20m'
6. **Tweet:** Necesito alguien con quien comentar #1octL6  
 Language: Spanish  
 Stance: NEUTRAL  
 'I need someone to comment on #1octL6 with'

Tweets are very short pieces of text but complex in their internal structure (including text, hashtags, links, @participants), and often their interpretation requires contextual information and even pragmatic knowledge. We consider that for detecting the stance we cannot restrict ourselves exclusively to the information included in the tweet because it is insufficient. The key of revealing the stance can be given by the text, but also by the link or the image, and often by the sum of all these sources of information.

### 3.3 The Annotation Process

The annotation of the *TW-10Referendum* corpus was carried out in two stages. In the first stage, we automatically annotated a subset of tweets as favorable or unfavorable from a list of preselected authors depending on the stance towards the first of October Referendum. The stance of the authors, who were politicians, journalists and other well-known people, was clearly known. However, the number of automatically annotated tweets turned out to be a very small subset, only 0.32% of the total annotated tweets. In the second stage, two trained annotators, supervised by three senior researchers, carried out the whole manual annotation of the *TW-10Referendum* corpus. This manual annotation was performed as follows:

- First, 500 tweets in each language were tagged by the trained annotators in parallel following the guidelines.
- Then, a first inter-annotator agreement test (IAA) was conducted on the 500 tagged tweets in order to detect and solve inconsistencies and disagreements (Table 1).
- After reviewing the cases of disagreement and resolving doubts, the annotators tagged 1,300 more tweets in each language and a second IAA test was carried out (Table 2).
- Finally, the whole corpus was tagged by the annotators individually. Being aware of the difficulty of the task, the annotators and the senior researchers met once a week to discuss problematic cases and solve them by consensus.

In order to ensure the quality of the annotation, two IAA tests were conducted for each language: Table 1 and Table 2 show the results obtained in the first and in the second IAA tests respectively. In the second and fourth columns we show the results of the observed agreement and in the third and fifth columns the corresponding Cohens kappa score for each language. The third row shows the results obtained when the annotators only took into account the textual information included of the tweet; whereas the last row shows the results obtained when the annotators took into account both the textual information and the information in the link of the tweet. We asked the annotators to tag the same tweet twice, first considering only the text and then the text and link in order to measure the importance of considering both types of information. The annotator agreement increased 4.4% in Catalan and almost 15% in Spanish when the information in the link was taken into account in the first IAA test (Table 1) and 2.5% in Catalan and 15% in Spanish in the second IAA test (Table 2). This double annotation was performed only in the IAA tests, not in the annotation of the whole corpus. The observed agreement (89.4% in Catalan and 83.3% in Spanish) and the Cohens kappa score (0.82 and 0.65 in Catalan and Spanish respectively) obtained in the second test are better than the results of the first test. Following Landis and Koch (1977), the results reached show a substantial and almost perfect agreement rate between the annotators. The better results in the Catalan IAA test can be explained by the fact that 87% of the Catalan

tweets are in favor of the Referendum, whereas in the Spanish dataset the data are more balanced and the stance is often less clearly expressed.

It is worth noting that we took into account all the information available for the annotation of tweets with stance. That is, we used all the information included in the tweet: the text of the message, the information included in the link (when applicable) and the information taken from the images on the authors timeline. However, we also took into account the information included in the authors profile (for instance, the profile image). Needless to say, we also took into account the pragmatic information, our knowledge about this topic. That meant that it was important for the annotators to be familiar with the context in which the debate was taking place. Moreover, in order to ensure the consistency of the annotation, we also made sure that tweets written by the same author were consistently annotated in relation to their stance.<sup>10</sup>

**Table 1.** Results of the first IAA test (N=500)

Stance	TW-10Referendum-CA (N=500)		TW-10Referendum-ES (N=500)	
	% Agreement	Kappa	% Agreement	Kappa
Text	81.8%	0.63	67.3%	0.54
Text+Link	86.2%	0.76	81.2%	0.68

**Table 2.** Results of the second IAA test (N=1300)

Stance	TW-10Referendum-CA (N=1300)		TW-10Referendum-ES (N=1300)	
	% Agreement	Kappa	% Agreement	Kappa
Text	86.9%	0.73	68.1%	0.57
Text+Link	<b>89.4%</b>	<b>0.82</b>	<b>83.3%</b>	<b>0.65</b>

Regarding disagreements in the annotation, the most problematic cases were related to the assignment of the neutral tag, especially with ambiguous tweets. For instance, one annotator tagged tweet (7) as neutral, whereas the other annotator tagged it as against. This tweet can be interpreted, in fact, as being both in favor or against the referendum, therefore we finally decided to tag the tweet with the neutral tag. Other disagreements were related to irony, such as (8). One annotator tagged this tweet as being against the referendum and the other as being in favor. The clue for the ironical interpretation of the tweet is the photo

<sup>10</sup> In the annotation of the *TW-StanceCat* corpus used in the StanceCat task (IberEval 2017), we only took into account the textual information of the tweet but not the information in the link or the images posted by the author, not even the authors profile information.

included in the link, in which a group of peaceful, elderly people appear. After collectively discussing this case, we agreed to tag the tweet with the favor tag because it had to be interpreted ironically.

7. **Tweet:** Coscubielibers! El nostre idol està La Sexta! Parlarà del Daniel? #1octL6  
 Language: Catalan  
 Stance: NEUTRAL  
 'Coscubielibers! Our idol is on La Sexta (TV Channel). Will he talk about Daniel? #1octL6'
8. **Tweet:** Els RADIKALS abduits i antidemocràtics que provoquen el TUMULTO certament fan bastanta por... #referendumCA #1O...  
<https://t.co/nlEa8rkXTT>  
 Language: Catalan  
 Stance: FAVOR  
 'These brainwashed,anti-democratic RADIKALS who caused this TUMULT certainly generate fear...' #referendumCA #1O...  
<https://t.co/nlEa8rkXTT>

### 3.4 Format and Distribution

We provided participants with 80% of the *TW-10Referendum* corpus (4,684 tweets in Catalan and 4,437 tweets in Spanish) for training their systems. The remaining 20% (1,169 tweets in Catalan and 1,108 tweets in Spanish) were used for testing their systems. The tweets were provided in two independent files per language.

The training set provided contained two files with the ground truth for each language: *truth-ca.txt* for tweets written in Catalan and *truth-es.txt* for tweets written in Spanish. Each of these files contained exclusively the *ID* of the tweet and its corresponding, manually annotated, stance in the following format: *id::stance* (see Section 3.2 for the possible stance values). In addition, two more *xml* files were provided with the contents: *ca.xml* and *es.xml* for Catalan and Spanish respectively. These files contained the following information in *xml* format as shown in Figure 1:

- The *ID* of the tweet;
- the text of the tweet to be evaluated;
- the contextual information, that is, the tweet before and after the tweet under evaluation; and
- the name of the image (up to 10 images) obtained from the author’s timeline.

The images were stored in the photos subfolder. The language was encoded both in the file name and in the *xml* files. The test set provided for evaluation contained the *xml* files for each language but the truth values were not included in the *txt* files.

```

<MultiStanceCat lang="ca">
<documents>
  <document id="cce366ddeb5d654c4669a05f35f9261">
    <text>
      <![CDATA[ @emaspons @ABecharaGhobril @UnitsCat
        @Espinakka @rccarbo Publicada fa dies, abans que el
        posicionament #1oct, aquí... K2]]>
    </text>
    <prev>
      <![CDATA[RT @alexmiropujadas: 1/3 #BarcelonaenComú vol
        crear una moneda local #etfelicitofill http://t.co/ASscmKYFDa
        http://t.co/abEjmvYCjg]]>
    </prev>
    <next>
      <![CDATA[Som en una crisi dEstat. Editorial de
        @lavanguardia: "Crida a la serenitat"
        https://t.co/5P6jyWsN9y]]>
    </next>
    <photos>
      <photo>fe48d4aace536d6610b0cf1bdae835cf.jpeg</photo>
      <photo>9a07da8bf46a9b27c3777fb3d66fef77.jpeg</photo>
      <photo>791e6171a7f393ee751f07182a593cc3.jpeg</photo>
    </photos>
  </document>
  ...

```

Fig. 1. XML files format.

Table 3 presents the distribution of stance in both the training and the test sets in each language.

Table 3. Distribution of the stance labels for Catalan and Spanish)

	CATALAN			SPANISH		
	TRAINING	TEST	TOTAL	TRAINING	TEST	TOTAL
Favor	4,085	1,021	5,106	1,680	419	2,099
Against	120	29	149	1,785	446	2,231
Neutral	479	119	598	972	243	1,215
<b>TOTAL</b>	<b>4,684</b>	<b>1,169</b>	<b>5,853</b>	<b>4,437</b>	<b>1,108</b>	<b>5,545</b>

### 3.5 Evaluation Measures

The evaluation was performed according to standard metrics. In particular, we used the macro-average of *F*-score (FAVOR), *F*-score (AGAINST), and *F*-score



(NEUTRAL) to evaluate stance, in accordance with the metric proposed at SemEval 2016 - Task 6. A majority-class baseline has been provided for the sake of comparison.

## 4 Overview of the Submitted Approaches

Four teams participated in the shared task by sending up to nine runs for Spanish and seven runs for Catalan. In Table 4 we show the participating teams and the modalities they took part in, i.e. text (T), context (C), images (I), and their combinations, for the two languages: Spanish (ES) and Catalan (CA).

All the teams participated in the stance subtask in Spanish and three of them participated in the Catalan subtask. The team that participated only in Spanish approached the task only with textual features, whereas the remaining participants also used the context. Only one participant used the images on the authors timelines. Three teams sent a working note describing their systems<sup>11</sup>. We will analyze their approaches from three perspectives: *preprocessing*, *classification method*, and the *features* used.

**Table 4.** Teams participating to *MultiStanceCat* at IberEval 2018

Team	CA	ES
Casacufans	T, T+C, T+C+I	T, T+C, T+C+I
CriCa [1]	T, T+C	T, C
ELiRF [4]	-	T
uc3m [8]	T, T+C	T, T+C

The Casacufans team approached the task using all the modalities: text, text + context, and text + context + images. To preprocess and represent texts they used Hashing Vectorized from the scikit-learn toolkit and linear Support Vector Machines to learn the model. With respect to images, the participants trained a Convolutional Neural Network to detect Spanish or Catalan flags. Regrettably, the authors did not send a working note explaining their approach in a greater detail.

The CriCa team approached the task considering texts and contexts. They also used the scikit-learn toolkit to preprocess and obtain features from the texts. Concretely, they obtained a bag-of-words representation with the stem of the words, weighted by tf-idf. Then, a linear Support Vector Machine was used as a classification algorithm.

The ELiRF team approached the task only with textual features. They lowercased the texts, removed special characters such as accents and dieresis, and normalized Twitter elements such as hashtags, user mentions and urls. They sent

<sup>11</sup> Regrettably, the authors of the Casacufans team did not send a working note explaining their participation.

two runs. In the first one (run1), they used word embedding and Convolutional Neural Networks, whereas in the second one (run2) they used different sets of character and word  $n$ -grams with linear Support Vector Machines.

The uc3m team approached the task with textual features and also took advantage of the context. To represent the texts they used a bag-of-words weighted with tf-idf. The authors tested several machine learning algorithms and opted for the linear Support Vector Machine.

## 5 Evaluation and Discussion of the Submitted Approaches

In this section we present and discuss the official results of the shared task. We also analyze how contextual features and images may improve the performance of the systems. Finally, an error analysis is presented.

### 5.1 Stance Results

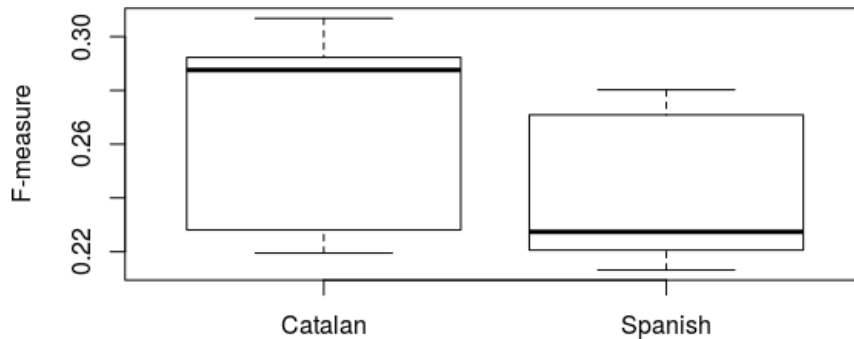
Four teams participated in the shared task, presenting seven runs in Catalan and nine in Spanish. In Table 5, the F-scores achieved by all runs are shown, as well as the baseline. At the bottom of the table some basic statistics are provided: minimum (min), maximum (max), mean, median, standard deviation (stdev), first quartile (q1) and third quartile (q3).

**Table 5.** Evaluation results for Stance in Catalan and Spanish ( $F$ -score).

Catalan				Spanish			
Position	Team	Run	F	Position	Team	Run	F
1	CriCa	T+C	0.3068	1	uc3m	T+C	0.2802
		<i>baseline</i>	<i>0.3050</i>	2	CriCa	T+C	0.2715
2	Casacufans	T+C	0.2933	3	Casacufans	T+C+I	0.2709
3	Casacufans	T+C+I	0.2913	4	Casacufans	T+C	0.2698
4	uc3m	T+C	0.2876	5	ELiRF	T (run1)	0.2274
5	CriCa	T	0.2315	6	uc3m	T	0.2247
6	Casacufans	T	0.2247	7	CriCa	T	0.2206
7	uc3m	T	0.2195	8	Casacufans	T	0.2194
		min	0.2195	9	ELiRF	T (run2)	0.2132
		q1	0.2281			<i>baseline</i>	<i>0.1913</i>
		median	0.2876			min	0.2132
		mean	0.2650			q1	0.2206
		stdev	0.0378			median	0.2274
		q3	0.2923			mean	0.2442
		max	0.3068			stdev	0.0278
						q3	0.2709
						max	0.2802

On average, the results for Catalan (26.50%) and for Spanish (24.42%) are very similar (there is no statistical significance between both means). However, as can be seen in Figure 2, the interquartile range is higher in the case of Catalan (6.42% vs. 5.03%), with higher values for most of the systems than in Spanish (the medians are 28.76% and 22.74% for Catalan and Spanish respectively).

It is worth mentioning that, due to the greater imbalance of the data in the case of Catalan, most of the participants' runs are below the majority-class baseline (30.50%). Only the CriCa team improved on this baseline, with an F-measure of 30.68%, which is not statistically significant.



**Fig. 2.** Distribution of results (F-score) for the stance subtask.

## 5.2 Improvement over Textual Features

In Table 6 we show the result achieved with textual features only, when adding contextual information, and the percentage of improvement. As can be seen, results were always improved when contextual information is taken into account, especially in the case of Catalan, where the improvement is about 30%. We applied the Bayesian Signed-Rank test [2] (results are shown at the bottom of Table 6). This test shows that the probability that the textual approach will obtain better results than when combined with context is 0%, while the probability of obtaining similar results is close to 1% in both languages. We can therefore conclude that the improvement achieved by adding contextual information is statistically significant.

**Table 6.** Percentage of improvement when using contextual information.

Catalan				Spanish			
Team	T	T+C	% Impr.	Team	T	T+C	% Impr.
Casacufans	0.2247	0.2933	30.53%	Casacufans	0.2194	0.2698	22.97%
CriCa	0.2315	0.3068	32.53%	CriCa	0.2206	0.2715	23.07%
uc3m	0.2195	0.2876	31.03%	uc3m	0.2247	0.2802	24.70%
P(T>T+C)	0%			P(T>T+C)	0%		
P(T=T+C)	1.18%			P(T=T+C)	1.18%		
P(T<T+C)	98.82%			P(T<T+C)	98.82%		

Only one participant (Casacufans) approached the task by also taking advantage of images. They trained a CNN with Spanish and Catalan flags with the intuition that people against or in favor of Catalan independence would use them. In case of Spanish they improved (27.09%) on the results obtained only with texts (21.94%) and with context (26.98%), albeit without reaching statistical significance. Furthermore, in the case of Catalan, their results (29.13%) were lower than those achieved using only text and context (29.33%), although again without reaching statistical significance. We can conclude that the use of images by this team did not contribute positively to stance identification.

### 5.3 Error Analysis

In this section we analyze errors in stance detection. We observed two kinds of errors: i) the participants interpreted a stance as being "in favor" when the real value was "against" (F->A); and ii) the participants interpreted a stance as being "against" when it was actually "in favor" (A->F). We analyzed the error rate for these two kinds of error. As can be seen in Table 7, the highest rate occurred with error F->A in Catalan (85.22%), whereas the rate of error A->F, also in Catalan, is almost null (0.08%). This is due to the imbalance in the corpus and shows that the systems did not model the task properly (they were biased towards the majority class, as it is supported by the results obtained by this baseline). In the case of Spanish, the highest error rate was with error type A->F (35.77%), though the difference with respect to type F->A (14.72%) is lower.

**Table 7.** Percentage of error types.

Catalan		Spanish	
F ->A	A ->F	F ->A	A ->F
85.22%	0.08%	14.72%	35.77%

Tables 8 and 9 show the top 5 most often wrongly classified tweets. The error rate is the result of dividing the number of runs that this particular error occurred in by the total number of errors. For example, in case of or type F->A in Catalan, there were 6 errors, with the first two occurring twice and the last two occurring once. Taking into account these error rates, the percentages obtained mean the following: in the case of type A->F in Catalan, the percentage of 33.33% means that two runs failed, whereas 16.67% means that only one run failed. The remaining percentages (in Tables 8 and 9) mean that all runs failed with the presented tweets<sup>12</sup>.

It is not always possible to infer the reasons for the misclassification of tweets, but we will now try to outline some possibilities. In the case of tweets in Catalan (Table 8), the assignment of the favor label instead of against can be explained by the fact that 87% of Catalan tweets are in favor of the 1 October Referendum, therefore the systems are biased towards the majority class, as mentioned previously. The two first tweets in Table 8, which were tagged with the value against instead of favor, can probably be explained by the fact that the hashtags included in the message are written in Spanish and usually used by authors tweeting against the referendum and, moreover, because they are written with irony. Finally, in the last two tweets in Catalan tagged with against instead of favor the detection of the stance is made more difficult by the fact that their interpretation depends on pragmatic information.

In the case of tweets in Spanish (Table 9), those classified as against the Referendum instead of in favor, could be explained because they all contained the hashtag #1octL6. This hashtag was related to the television debate about the Referendum that took place on 'La Sexta', a Spanish TV channel. The majority of tweets in which the #1octL6 hashtag appeared were classified as neutral or against the Referendum. Therefore, the systems learned that the most probable stance was against (though that was not the case in these tweets). The tweets in Spanish erroneously labeled as being in favor rather than against are more difficult to explain. The first two are ironic tweets and stance detection is therefore more difficult. The last one is ambiguous and can be interpreted literally and tagged as being against or interpreted ironically and tagged as being in favor. In this case, the annotators opted for a literal interpretation. Finally, the third and fourth tweets do not contain enough information to be tagged with an in favor stance and should have been tagged as neutral. The fact that in these cases the annotators assigned the wrong stance tag serves to highlight the difficulty of this task, even for human annotators.

---

<sup>12</sup> For instance, to obtain 4.05% in Table 8 in type F->A, we divided 7 (the number of runs in Catalan that failed with this particular tweet) by 173 (the total number of errors of this type).

**Table 8.** Tweets more frequently misclassified in Catalan.

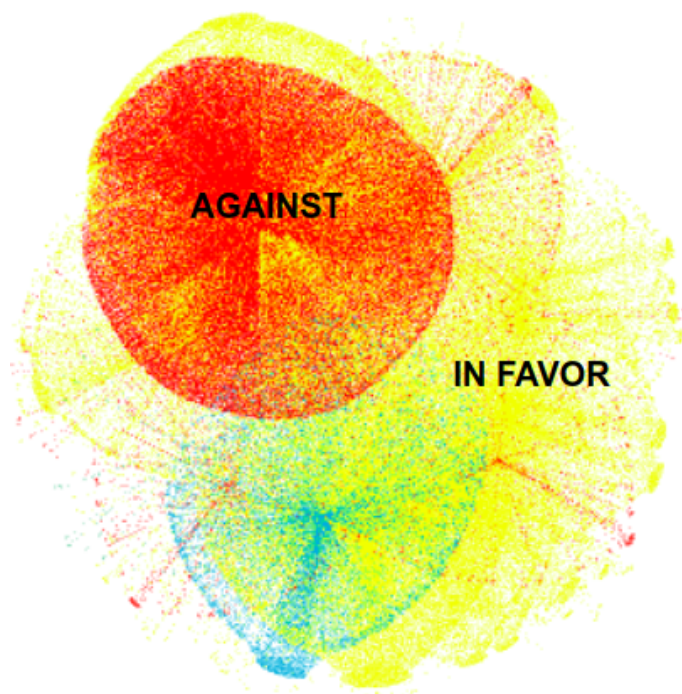
%	<b>Against -&gt;Favor</b>
33.33%	So true ! Què recomfortant, no estem sols !! #1octubreARV #acosoindependentista #hispanofobia #hispanoMola... <a href="https://t.co/nwxEr0coRz">https://t.co/nwxEr0coRz</a> 'So true! How comforting; we aren't alone!! #1octubreARV #acosoindependentista #hispanofobia #hispanoMola... <a href="https://t.co/nwxEr0coRz">https://t.co/nwxEr0coRz</a> '
33.33%	Després de l'@arqueoleg, ara és @jordievole qui se suma al #ProcésChallenge... #SeñorLlévamePronto #1O <a href="https://t.co/0wxBw0al7U">https://t.co/0wxBw0al7U</a> 'After the @arqueoleg, now it's @jordievole who's joining the #ProcésChallenge... #SeñorLlévamePronto #1O <a href="https://t.co/0wxBw0al7U">https://t.co/0wxBw0al7U</a> '
16.67%	Porto tota la setmana al gimnàs. Haig d'estar en forma per saltar la tanca el diumenge. #1O #Votarem 'I've been at the gym all week. I've got to get fit to leap over the barricades on Sunday. #1O #Votarem'
16.67%	@JRNadal @JRNadal a #Sueca governa el teu partit. És molt lamentable i decepcionant el posicionament de @compromis respecte #1O '@JRNadal @JRNadal your party is in power in #Sueca. The position of @compromis on #1O is unfortunate and disappointing'
%	<b>Favor -&gt;Against</b>
4.05%	Pepa Bueno entrevista monicaoltra: Els problemes s'han de resoldre amb diàleg i democràcia <a href="https://t.co/yBTCxD4GI">https://t.co/yBTCxD4GI</a> #1O ... 'Pepa Bueno interviewing monicaoltra: Problems must be solved through dialogue and democracy <a href="https://t.co/yBTCxD4GI">https://t.co/yBTCxD4GI</a> #1O ...'
4.05%	CarlaAntonelli: RT abalosmec: Davant la fatalitat hauria d'haver esperança, abans i després de #1O La sortida és ... <a href="https://t.co/TkVa8UQA0z">https://t.co/TkVa8UQA0z</a> 'CarlaAntonelli: RT abalosmec: Against fatalism there must be hope, before and after #1O The way forward is... <a href="https://t.co/TkVa8UQA0z">https://t.co/TkVa8UQA0z</a>
4.05%	@AdaColau Si us plau!!!. No fotis!!! #TotsSomBarcelona #1O <a href="https://t.co/X9ZREWgoiC">https://t.co/X9ZREWgoiC</a> '@AdaColau Please!!!. You must be kidding!!! #TotsSomBarcelona #1O <a href="https://t.co/X9ZREWgoiC">https://t.co/X9ZREWgoiC</a> '
4.05%	Tan fals com aquest. Gràcies a tots per assumir que en aquest #1O circulen cartells i afirmacions falses ???????? <a href="https://t.co/pCOqTqNpSU">https://t.co/pCOqTqNpSU</a> 'As false as this one. Thanks to everyone for assuming that fake posters and statements are circulating in this #1O ???????? <a href="https://t.co/pCOqTqNpSU">https://t.co/pCOqTqNpSU</a> '
4.05%	Davant la fatalitat hauria d'haver esperança, abans i després de #1O La sortida és diu dialleg, reforma de la constitució i més autogovern 'Against fatalism there must be hope, before and after #1O The way forward is through dialogue, reforming the constitution and more self-government'

**Table 9.** Tweets more frequently misclassified in Spanish.

%	Against ->Favor
0.67%	#1octL6 gracias a los catalanes podremos reformar la constitución,el derecho a decidir, la monarquía...q se mojen los partidos,ya está bien '#1octL6 thanks to the Catalans well be able to reform our constitution, the right to decide, the monarchy, that's enough, it's time for parties to take a stand'
0.67%	#1octL6 Hablan y hablan de democracia, pero... democracia sin leyes? Vamos cambiando leyes a nuestro antojo? '#1octL6 They're always talking about democracy, but... democracy without laws? Should be change the law on their whims?'
0.67%	#1octL6 para variar Iceta y los socialistas no saben nada. Por eso se ha llegado a esa situación. Marxem???????? '#1octL6 for a change Iceta and the socialists don't get the picture. That's why they're in this situation. Time to leave????????'
0.67%	#1octL6 Todos los no nacionalistas quieren reforma de la constitución para darle más poder. Los que van a salvar ESP van a ser los de la CUP '#1octL6 All of the non-nationalists want to change the constitution to make it more powerful. The ones who are going to save Spain are the CUP'
0.67%	#1octL6 El problema es q en Spain se puede ser independentista de pensamiento, pero no en la práctica, y eso no funciona así. '#1octL6 The problem is that in Spain you can be an independentist in your thoughts but not in practice, and that doesn't work that way.'
%	Favor ->Against
1.52%	Tensión en las calles de Barcelona, Catalunya vol.1 #1o #1Oct #Cataluna #catalunya #CatalunyaNoEstasSola #Barcelona... <a href="https://t.co/TmmlNovXer">https://t.co/TmmlNovXer</a> 'Tension in the streets of Barcelona, Catalonia vol.1 #1o #1Oct #Cataluna #catalunya #CatalunyaNoEstasSola #Barcelona... <a href="https://t.co/TmmlNovXer">https://t.co/TmmlNovXer</a> '
1.52%	#1octL6 albiol y la mayoría silenciosa que hoy se ha quedado en casa??? Estos hoy no se cuentan '#1octL6 albiol and the silent majority have stayed at home today??? They're not being counted today'
1.52%	@InesArrimadas si estuviera Rajoy en los demas países, posiblemente ocurriría lo que ocurre a día de hoy en catalunya. #1octL6 '@InesArrimadas if Rajoy was in any other country, quite possibly what would happen is the same as what's happening in Catalonia today.'
1.52%	PSOE, el partido 'obrero' #10oct #1oct2017 #Referendum1oct #catalunya 'PSOE, the "workers"' party #10oct #1oct2017 #Referendum1oct #catalunya'
1.52%	#1octL6 Ja Ja Albiol ya mostró TODO lo k no se kiere negociar. La Constitución ni tocarla!!Claro,Cs y PP de acuerdo.IMPOSIBLE DIALOGO '#1octL6 Ha ha Albiol already demonstrated that he doesn't want to negotiate. Hands off the constitution!! Of course, Cs and PP agree. DIALOGUE IMPOSSIBLE'

#### 5.4 Social Network Analysis

As a preliminary study, with an interest in investigating the echo chamber effect in the Catalan #1Oct Referendum, we have randomly selected a sample of users who tweeted against or in favor of Catalan independence. We then downloaded all the users followed by the previous sample. We annotated whether these users were followed by a pro-independence, an anti-independence or both. Figures can be seen in Table 10.



**Fig. 3.** Stance-based social network communities.

In Figure 3, the three communities of users are shown. In red, representing 71.30% of the total, is the community of users that are followed by users against the Catalan 1 October Referendum. In yellow, with 28.04% of the total, is the community of users that are followed by users in favor of the Referendum. Finally, in blue, with 0.67% of the total, is the community of users that are followed by both users in favor and against the Referendum. We can draw two main conclusions. Firstly, the anti-referendum community is much larger (71.30% vs. 28.04%), though this might depend on the selected sample. Secondly, the lack of interest in communicating with the other community and understanding the other viewpoint: the blue community is only 0.67% of the total.



**Table 10.** Number of users in the sample and their network.

Stance	Sample	Network	%
In favor	25	9,091	28.04%
Against	25	23.119	71.30%
Both	-	216	0.67%
Total	50	32.426	100%

## 6 Conclusions

We have described the second edition of the shared task on detecting the stance towards Catalan Independence. This year we have introduced multimodality to the task. Namely, we have provided the contexts of the tweets and ten images from the authors' timeline. As in the previous year, the task was carried out in Catalan and Spanish, the two languages used by users directly involved in the political debate. We encouraged participants to address both languages (Catalan and Spanish) together with all the possible modalities. Three modalities (text, text + context, text + context + images) have been finally addressed by the participants.

Regrettably, we only had four participants and only one of them took advantage of the extra information that images could provide. Despite this limitation, we can draw some interesting conclusions. As in the previous year, the results confirm that stance detection in micro-blogging texts is challenging, and there is plenty of room for improvement. In the case of Catalan, due to the imbalance of the data, most systems performed below the majority-class baseline. Even in Spanish, the best performing system obtained less than 30% of the F-measure and performed only about 9% higher than the baseline. However, it is very interesting to note how the systems improved their performance when the context of the tweets was added. All of the systems improved more than 30% in Catalan and more than 20% in Spanish. In both cases, the significance test shows that these systems always improved their results when context was taken into account.

Our preliminary study on the echo chamber effect in the Catalan #10Oct Referendum showed that there is a lack of interest in communicating with the other community and understanding their viewpoint since only 0.67% of users communicated across communities.

We hope that the dataset made available as part of the *MultiStanceCat* task<sup>13</sup> will foster further research on this topic, especially in under-resourced languages such as Catalan.

<sup>13</sup> <http://www.autoritas.net/MultiStanceCat-IberEval2018/corpus/>

## Acknowledgements

The work has been carried out in the framework of the SOMEMBED project (TIN2015-71147), funded by Ministerio de Economía y Competitividad, Spain. The work of the second author has been partially funded by Autoritas Consulting.

## References

1. Almendros-Cuquerella, C., Cervantes-Rodríguez, C.: Crica team: Multimodal stance detection in tweets on catalan 1oct referendum (multistancecat). In: Notebook Papers of 3rd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval), Sevilla, Spain, September 18, CEUR Workshop Proceedings. CEUR-WS.org, 2018 (2018)
2. Benavoli, A., Mangili, F., Corani, G., Zaffalon, M., Ruggeri, F.: A Bayesian Wilcoxon signed-rank test based on the Dirichlet process. In: Proceedings of the 30th International Conference on Machine Learning (ICML 2014). pp. 1–9 (2014), <http://www.idsia.ch/~alessio/benavoli2014a.pdf>
3. Bosco, C., Lai, M., Patti, V., Rangel, F., Rosso, P.: Tweeting in the debate about catalan elections. In: Proceedings of the International Workshop on Emotion and Sentiment Analysis (co-located with LREC 2016). ELSA, Portoroz, Slovenia (2016)
4. González, J.A., Hurtado, L.F., Pla, F.: Elirf-upv at multistancecat 2018. In: Notebook Papers of 3rd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval), Sevilla, Spain, September 18, CEUR Workshop Proceedings. CEUR-WS.org, 2018 (2018)
5. Lai, M., Farías, D.I.H., Patti, V., Rosso, P.: Friends and enemies of clinton and trump: using context for detecting stance in political tweets. In: Mexican International Conference on Artificial Intelligence. pp. 155–168. Springer (2016)
6. Lai, M., Patti, V., Ruffo, G., Rosso, P.: Stance evolution and twitter interactions in an italian political debate. In: International Conference on Applications of Natural Language to Information Systems. pp. 15–27. Springer (2018)
7. Mohammad, S.M., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: Semeval-2016 task 6: Detecting stance in tweets. In: Proceedings of the International Workshop on Semantic Evaluation. pp. 31–41. SemEval '16, ACL, San Diego, California (June 2016), <http://aclweb.org/anthology/S/S16/S16-1003.pdf>
8. Segura-Bedmar, I.: Labdas early steps toward multimodal stance detection. In: Notebook Papers of 3rd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval), Sevilla, Spain, September 18, CEUR Workshop Proceedings. CEUR-WS.org, 2018 (2018)
9. Taulé, M., Martí, M.A., Rangel, F.M., Rosso, P., Bosco, C., Patti, V.: Overview of the task on stance and gender detection in tweets on catalan independence at ibereval 2017. In: 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017. vol. 1881, pp. 157–177. CEUR-WS (2017)