WORDUP! at VaxxStance 2021: Combining Contextual Information with Textual and Dependency-Based Syntactic Features for Stance Detection

Mirko Lai[1,⋆], Alessandra Teresa Cignarella[1,2,⋆], Livio Finos[3,⋆], and Andrea Sciandra[4,⋆]

[1] Università degli Studi di Torino, Italy
[2] Universitat Politècnica de València, Spain
[3] Università degli Studi di Padova, Italy
[4] Università degli Studi di Modena e Reggio Emilia, Italy

Abstract. In this paper we describe the participation of the WORDUP! team in the *VaxxStance* shared task at *IberLEF 2021*. The goal of the competition is to determine the author's stance from tweets written both in Spanish and Basque on the topic of the *Antivaxxers movement*. Our approach, in the four different tracks proposed, combines the Logistic Regression classifier with diverse groups of features: stylistic, tweet-based, user-based, lexicon-based, dependency-based, and network-based. The outcomes of our experiments are in line with state-of-the-art results on other languages, proving the efficacy of combining methods derived from NLP and Network Science for detecting stance in Spanish and Basque.

Keywords: Stance Detection · Spanish and Basque · MDS · Contextual Features · Network Information · Syntax · Universal Dependencies · NLP

1 Introduction

In the last five years there has been a noticeable growth of interest in determining whether the author of a social media text is in favor, against, or neutral towards a statement or targeted event, person, organization, government policy or movement. The research area investigating such matter has been defined in literature as *Stance Detection* (SD) [31].

Investigating on this topic could have a huge impact on different aspects of everyday life such as policy-making, security choices and public administration strategies. A practical application of SD techniques, in fact, could support the automatic identification of people's extremist tendencies on the one hand (e.g.,

⋆ The contribution of each author is detailed in the Acknowledgments section.

IberLEF 2021, September 2021, Málaga, Spain.

religious extremism [17]), but could be employed by authoritarian governments to better control their citizens too. All in all, the amount of academics and companies dedicated to the computational study of polarized debates, today, is bigger than ever. This is also witnessed by the increase of scientific publications on the topic as recently surveyed by Küçük and Can [21] and by the ever-growing organization of shared tasks in different languages on a manifold of controversial events and topics. In order of appearance, the first shared task entirely dedicated to SD was held for English at SemEval in 2016 (i.e., *Task 6 "Detecting Stance in Tweets"* [30]) for detecting stance towards six different targets of interest: "Hillary Clinton", "Feminist Movement", "Legalization of Abortion", "Atheism", "Donald Trump", and "Climate Change is a Real Concern". After that, in 2017 a competition for SD systems was proposed at IberEval 2017 for both Catalan and Spanish: *StanceCat 2017*, where the target was uniquely the "Independence of Catalonia" [37]. The following year, the organizers proposed a follow-up edition in Catalan and Spanish, encouraging stance detection with multimodal approaches toward the target "Catalan 1st of October Referendum" (i.e., *MultiStanceCat*) [39]. Later on, in the second half of 2020, the first SD task for Italian has also been organized (i.e., *SardiStance*) proposing two different settings – *textual SD* and *contextual SD* – inviting, among other things, the exploration of contextual features based on the interactions with tweets, on the profile of users and also their social media network [8].

Along with the above mentioned shared tasks, which are representatives of the so-called TARGET-SPECIFIC STANCE CLASSIFICATION, there have been others dedicated to a different fashion of categorization, i.e., OPEN STANCE CLASSIFICATION. This second type of SD task is often mentioned with the acronym SDQC, by referring to the four categories[5] exploited for indicating the attitude of a message with respect to a rumour [2,42]. The most relevant events following this second kind of categorization are *SemEval-2017 Task 8* [11] and *SemEval-2019 Task 7* [16].

Although stance detection is an NLP problem still in its emerging stage, within these competitions – but not uniquely – there has a considerable body of conducted research on the topic, exploring a great variety of methodologies. In the most part, it is a common practice to exploit various classifiers and compare their results. According to recent related work, the most employed techniques for detecting stance are: rule-based algorithms; supervised algorithms like SVM, naïve Bayes, boosting, decision tree and random forest, Hidden Markov Models (HMM) and Conditional Random Fields (CRF); graph algorithms such as MaxCut, and other approaches such as Integer Linear Programming (ILP) and Probabilistic Soft Logic (PSL) [21]. The most exploited deep learning methods are Recurrent Neural Network (RNN)-based system such as such as Long Short-Term Memory (LSTM) [41], and Convolutional Neural Networks (CNN) [7,15,18,40]. Alongside with the traditional classification algorithms, the most used features are character n-grams, word n-grams, and features based on PoS tags, hashtags, and sentiment dictionaries. On the other hand, the approaches

[5] Support (S), Deny (D), Query (Q) and Comment (C).

based on deep learning, mostly exploit word embeddings, such as *word2vec* and *fastText* as additional features [19,29].

In this paper we combine insights gained from some previous work of two different research groups[6] [8,13,23]. In particular, we explore diverse groups of features such as: stylistic, tweet-based, user-based, and lexicon-based, also carrying on new research on dependency-based syntax and network-based data augmentation techniques (see Section 2).

1.1 The Task

The aim of *VaxxStance @ IberLEF 2021* [1,32] is to encourage the research community in working on SD in two languages: Basque and Spanish. In particular, the computational goal is for an automatic system, to determine whether a given tweet expresses an AGAINST, FAVOR or NEUTRAL stance towards the target: *vaccines*. The task organizers encouraged the participation, for both languages, in three different tracks, one of which also contained two different settings:

1. Close Track
   (a) Textual
   (b) Contextual
2. Open Track
3. Zero-Shot Track

For more information regarding the details of each track and setting, please refer to Agerri et al. [1] and to the official task webpage.[7] Our team participated in all three tracks proposed, by submitting 8 runs for Spanish and 8 runs for Basque. In the following section we outline the main features that we engineered.

2 Our Proposal

Our team – WordUp! – is composed by four researchers that have already dealt with SD in their previous work. In particular, this work is a joint research with some of the organizers of the *SardiStance @ EVALITA* shared task [8] and one of its participating teams (i.e., TextWiller) [13].

SardiStance and *VaxxStance* present some similarities. Indeed, *SardiStance* explored for the first time the setting of CONTEXTUAL STANCE DETECTION in Italian tweets with the addition of information on the tweet itself (e.g., the number of retweets, the number of favors and the date of posting) and contextual information about the author (e.g., as follower count, location, user's biography, and their social media network). Furthermore, some of the authors have acquired experience in predicting stance in Spanish tweets participating in the *StanceCat* shared task at IberEval 2017 [24]. Therefore, we propose a supervised approach which consists in determining stance towards the *Antivaxxers*

[6] The authors joined forces after the participation in the shared task *SardiStance 2020*.
[7] https://vaxxstance.github.io.

movement employing different types of features inherited from previous work. Additionally, we have introduced some novel types of features that have been specifically conceived for this task, and with the aim of augmenting data for Spanish and Basque.

In the sections below we list, in the most accurate way possible, all the diverse features that have been implemented for this work. For each feature, we also propose an acronym to be used later in the description of the 16 submitted systems. In fact, not all the features listed below have been ultimately employed in the submissions, but the best features for each track (and setting) were rather selected on the basis of the results obtained performing a 5-fold cross-validation on the training set (see Table 3 for details on the submitted systems).

2.1 Stylistic features

The first type of features we propose are commonly used in sentiment analysis [33]. They are based on the bag-of-words model, and they are namely:

- Bag of Words (**BoW**): a binary feature selecting 1-3 word n-grams of the textual content of the tweet.
- Bag of Chars (**BoC**): a binary feature selecting 3-5 character n-grams of the textual content of the tweet.

We also propose some features that try to capture the author's style of writing. They include the use of punctuation marks, the recourse to uppercase words (commonly used in social media for shouting) [5], the presence of the phonosymbolism of laughter (sometimes used for humiliating the interlocutor's position) [12], and the use of percentage numbers.

- Punctuation Marks (**PM**): a 6-dimensional feature that includes the frequency of exclamation marks, question marks, periods, commas, semicolons, and finally the sum of all the punctuation marks mentioned before.
- Uppercase Words (**UpW**): this 4-dimensional feature refers to the amount of upper-cased words of at least two chars, the number of words starting with a capital letter, the number of lowercase words containing at least two uppercase characters, and the ratio between uppercase and lowercase words.
- Laughter (**Lau**): a binary 1-dimensional feature that checks the presence of laughter (e.g., ahahah or jajaja).
- Percentage Numbers (**PN**): a 5-dimensional feature that counts the number of percent sign (%), real percentage numbers, of real percentage numbers greater than 50%, of real percentage numbers lower than 50%, and of real percentage numbers greater than 90%.

The last stylistic feature we propose estimates the values of lexical complexity, which might be related to the level of education of a user and its proficiency in writing [22].

– Lexical Complexity (**LC**): a 9-dimensional feature measuring lexical diversity (3 features) and readability (6 features) as proxies for textual complexity. We selected among several metrics [6] those that were found to discriminate stances the most, by means of multinomial regressions and Krusal-Wallis tests with Nemenyi's non-parametric all-pairs comparisons (Spanish: Bormuth.MC, Coleman, Coleman.C2, Dale.Chall, Danielson.Bryan.2, FOG, TTR, R, I; Basque: ARI, Bormuth.MC, Dale.Chall.old, Danielson.Bryan.2, Flesch, FOG, R, I, D).

2.2 Lexica-based features

The second type of features that we implemented is based on lexical resources. Relying on positive results obtained in previous research in SD on other languages such as English [4,14], we manually created a dictionary containing lemmas that refer to attitudes or states of mind (Open Cue Words). Additionally, we created a second dictionary which contains lemmas that we considered as pragmatically rich for the purpose of detecting stance (Open Linguistic Words). Both dictionaries have been created in English and then automatically translated in Spanish and Basque.

– Open Cue Words (**OpenCW**): a binary representation counting the presence/absence of words related to the following categories: *belief, denial, doubt, fake, knowledge, negation, question, report* [4];
– Open Linguistic Words (**OpenLW**): a binary representation counting the presence/absence of words related to the categories of: *assertives, bias, factives, implicatives, hedges, report verbs* [14];

2.3 Twitter-based features

The organizers released some metadata about the tweets and their authors in addition to the textual content of the tweet itself. A `Tweet_object`, derived form Twitter's APIs, has a long list of attributes, including fundamental ones such as *id*, *created_at*, and *text*. It also includes the `User_object` of the author that contains, in turn, other attributes such as *created_at*, *follower_count*, and *statuses_count*. We thus propose the following features for representing a tweet as a numerical vector:

– Closed Tweet's info (**CloseTinfo**): a 4-dimensional feature that takes into account the number of retweets and favorites that the tweet received, and the year, the month, and the hour of publication.
– Closed Tweet source tag (**CloseTSource**): a one-hot encoding representation of the source used for posting the tweet (e.g., Android, iOS).
– Closed User's info (**CloseUinfo**): a 7-dimensional feature that represents a user from the number of statuses posted, the number of followers and friends, the number of lists which the user is a member of, the year and month in which the account was created, and finally, the ratio of tweets posted per day by the user.

With the aim of exploring data augmentation, we also recovered the *description* field from the `User_objects` employing the Twitter's API GET users/show[8]. This attribute contains a string that is used by the account for describing himself. We propose the following feature:

– Augmented Bag of Description (**ABoD**): a binary feature selecting the word 1-grams of the textual content of the user's description (bio).

2.4 Word embeddings

Word embeddings are a type of word representation, based on distributional semantic theories, that allows words with similar meaning to have a similar vector representation. Despite we do not propose features exclusively based on word embeddings, we use them for refining some of the best features that we describe in the next paragraphs. To the best of our knowledge, a word embedding trained *specifically on tweets* for Spanish or Basque is not publicly available. Therefore, we decided to create ourselves two word embeddings models for both languages.

First, we collected about 907,000 tweets in Spanish and 853,000 tweets in Basque from January 2018 until April 2021 using Twitter's Academic Full Search API[9]. In order to obtain a random sample, we split one year in 1,460 timestamps (4365) t spaced from each other by 6 hours. Then, we request 10 tweets (*max_results*) for each timestamp t shifting the query parameter *end_time* of a random value between -6 and +6 hours from t. We employ the *lang* and " * " operators for retrieving tweets in each language. Using the same method, but filtering with the words "vacun*" and "txert*" (*vaccine*, respectively in Spanish and Basque), we then collected about 894,000 random tweets in Spanish and 12,000 random tweets in Basque. We also collected about 2,000,000 Facebook's messages in Spanish and 2,700 in Basque [9] respectively containing the words "vacuna*" and "txerto*", posted between October 2020 to April 2021 for the Spanish language and from April 2020 to April 2021 for Basque[10]. We finally included the content of all Wikipedia's pages in Spanish (1,689,000 pages) and Basque (375,000 pages).

The whole two corpora (composed by Twitter's posts, Facebook's messages, and Wikipedia's pages) have been used for training two word embeddings models of size 100 (ES_EMBEDDINGS and EU_EMBEDDINGS) employing *fastText* (the module we used is included in the python's library *Gensim*). We chose *fastText* because it allows to query for words that do not appear in the training data. This characteristic is very useful in social media domains, in order to represent unknown hashtags that are composed of known substrings.

8 https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-users-show

9 https://developer.twitter.com/en/docs/twitter-api/tweets/search/api-reference/get-tweets-search-all.

10 https://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-and-researchers.

2.5 Dependency-based features

The availability of morphological and syntactic knowledge is crucial for engineering the last group of features, which relies mainly on dependency syntax, encoded through the format of *Universal Dependencies*[11] (UD). Therefore, to obtain a UD representation of the texts of both training and test set, we apply the *UDPipe* pipeline (for tokenization, PoS-tagging and parsing) to them.[12] For doing so, we train two different models (one for Basque and one for Spanish) on all the available treebanks for those two languages: AnCora [38], GSD [27], and BDT [3]. After this procedure, we obtain a representation of texts like the one explicited in Figure 1 and Table 1.

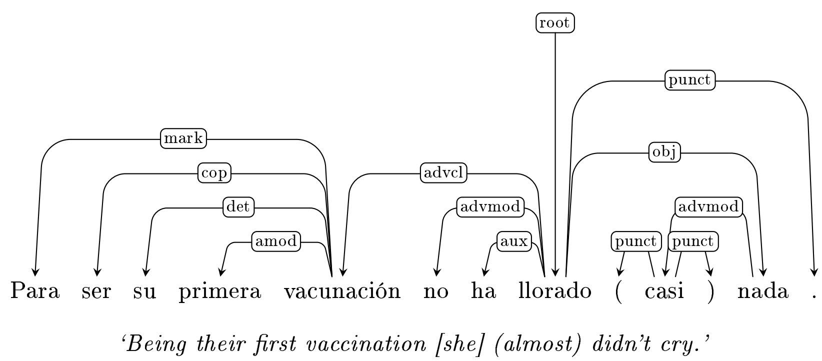

Fig. 1: Dependency-based syntactic tree of a Spanish tweet.

The features we design on the availability of morphology and dependency syntax are then following:

– upos (**upos**): a binary feature selecting PoS tags and creating a bag of 1-3 PoS n-grams. For instance: *[*ADP*,* AUX*,* DET*,* ADJ*,* NOUN*,* ADV*,* VERB*,* PUNCT*...]*

– deprelneg (**drN**): we consider the presence of negation in the text, relying on the morphosyntactic cues present in the UD format. When a negation was present, we append the correspondent dependency relation in the feature vector. For instance in Figure 1, we spot a negation in *[... primera vacunación* ***no*** *ha llorado ...]*, the dependency relation of "no" is `advmod`, therefore, we append it in the feature vector;

[11] https://universaldependencies.org/.

[12] See: http://ufal.mff.cuni.cz/udpipe and the Pyhton library: https://pypi.org/project/spacy-udpipe/.

id	token	lemma	upos	xpos	feats	head	deprel
1	Para	para	ADP	ADP	Prep	5	`mark`
2	ser	ser	AUX	AUX	Inf	5	`cop`
3	su	su	DET	DET	Sing|3|Poss|Prs	5	`det`
4	primera	primero	ADJ	ADJ	Fem|Sing|Ord	5	`amod`
5	vacunación	vacunación	NOUN	NOUN	Fem|Sing	8	`advcl`
6	no	no	ADV	ADV	Neg	8	`advmod`
7	ha	haber	AUX	AUX	Ind|Sing|3|Pres|Fin	8	`aux`
8	llorado	llorar	VERB	VERB	Masc|Sing|Tense=Past|Part	0	`root`
9	(	(	PUNCT	PUNCT	Ini|PBrck	10	`punct`
10	casi	casi	ADV	ADV	_	12	`advmod`
11	)	)	PUNCT	PUNCT	Fin|Brck	10	`punct`
12	nada	nada	PRON	PRON	Sing|Neg	8	`obj`
13	.	.	PUNCT	PUNCT	Peri	8	`punct`

Table 1: CoNLL-U representation of the Spanish tweet in Figure 1.

– deprel (**dr**): we build a bag of words of 5-grams, 6-grams and 7-grams of dependency relations as occurring in the linear order of the sentence from left to right; e.g. *[mark cop det amod advcl advmod aux root punct advmod punct obj punct]*;

– relationformVERB (**rV**): we create a feature vector with all the tuples of tokens that are connected with a dependency distance = 1, by starting from a verb and at the same time we blank the verb itself. For instance, in the example a verb is "*llorado*" and some of the tuples of tokens connected through this verb are, e.g. *[noVERBnada, noVERBha, noVERB., haVERBno, haVERBnada...]*;

– relationformNOUN (**rN**): we apply the same procedure of the feature above but considering nouns as starting points for collecting tuples;

– relationformADJ (**rA**): in the same fashion of the two features above, we repeat the same procedure for adjectives too;

– Sidorovbigramsform (**sF**): we create a bag of word-forms (tokens), considering the 2-grams that can be collected following the syntactic tree structure (rather than the 2-grams that can be collected reading the sentence from left to right).[13] Such that: e.g. *[llorado ha, llorado no, llorado vacunación, llorado nada, ... vacunación para, vacunación ser, ..., nada casi, ...]*;

– Sidorovbigramsupostag (**sT**): as the feature above, we create a bag of part-of-speech tags following the syntactic tree structure, starting from the root;

[13] Please refer to [36] and [35] for more details on this regard.

– Sidorovbigramsdeprel (**sDR**): as the two features above, we creat a bag of words based on dependency relations (*deprels*) following the syntactic tree structure, starting from the root.

We then propose two additional features that represent a tweet by using the dependency relations between the target of interest (for instance, the NOUN "vaccine" or VERB "vaccination"), in both Basque and Spanish, and the connected TOKENS in the dependency tree.

– Target Context Level 1 (**TC1**): this 200-dimensional feature consists in concatenating the word embeddings representations of the *previous* and *next* with respect to the target in the dependency tree.
– Target Context Level 2 (**TC2**): this 400-dimensional feature integrates the feature **TC1** including the word embeddings representations of the second level of *previous* and *next* words connected to the target in the dependency tree.

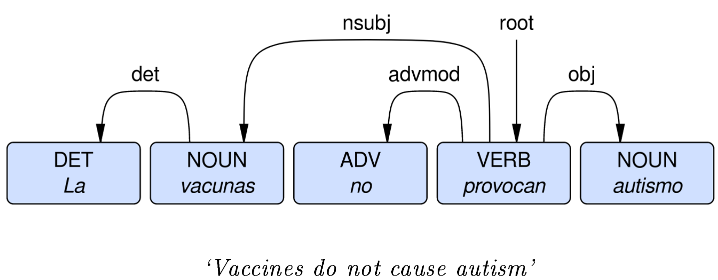

Fig. 2: The dependency tree of the sentence "La vacunas no provocan autismo".

Figure 2 shows the dependency tree of the sentence "Las vacunas no provocan autismo" (Vaccines do not cause autism). In this sentence, the target is the NOUN *vacunas* (vaccines). We can observe that the target has a dependency relation of type `det` (determiner) with the definite article *las* (the) and a relation of type `nsubj` (nominal subject) with the VERB *provocan* ([they] cause). Therefore, the feature **TC1** consists in the concatenation of the word embeddings representation of the words *las* and *provocan*.

```
concatenate([ model('las'), model('provocan') ])
```

Where the function $model(X)$ (ES_EMBEDDINGS or EU_EMBEDDINGS, depending on the language) returns the word embeddings representation of size 100 of the word X.

Then, the feature **TC2** includes the word embeddings representations of the second level of *previous* and *next* words connected to the target in the dependency tree. In this example there are no words related with the definite article *las*, but there are two words *no* (no) and *autismo* (autism) are connected with the VERB *provocan*. We fill the feature with zeros when a related word is missing, and we average the word embeddings when two or more words exist at the same level of the dependency tree. The feature **TC2** represents the example as follows:

```
concatenate([ 100*[0],
              model('las'), model('provocan'),
              average([model('no'), model('autismo')])
            ])
```

The two features are padded with zeros when the target is not mentioned in the text and we averaged the word embeddings representations when the target is mentioned two or more times.

2.6 Network-based features

A peculiarity of this NLP task is its willingness to explore the interplay of online social networks and users' stance. Indeed, the organizers released the *user_id* of every user the author of the tweet is following[14] (otherwise known as their "friend") and the *user_id* of every original tweet the author retweeted in their `User_timelines`.[15] We then create two FRIENDS NETWORK directed graphs - one for each language - where the nodes are users and an edge between two users exists if one follows the other. Similarly, we create two RETWEETS NETWORK directed graphs where the nodes are users and an edges between two users exists if one retweeted the other.

With the aim of exploring data augmentation - taking advantage of Twitter's Academic Full Search API[9] - we also retrieve the list of *user_id* retweeted by the author and the list of *user_id* that retweeted the author.[16] We obtain two AUGMENTED RETWEETS NETWORK directed graphs - one for each language - where the nodes are users and an edge between two users exists if one retweeted the other. The order (number of nodes) and the size (number of edges) of the two networks are shown in Table 2.

Spanish		Basque	
size	*order*	*size*	*order*
15,263,128	1,509,403	1,452,748	155,057

Table 2: Size and order of the AUGMENTED RETWEETS NETWORK

[14] https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-friends-ids.
[15] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/timelines/api-reference/get-statuses-user_timeline.
[16] We retrieved additional data from January 1st, 2019 until May 1st, 2021.

Network data allow us to propose three types of features that respectively measure the centrality of each node, the distance between all pairs of them, and the mixing of the networks on stance towards vaccination (i.e., assortative mixing).

The first type of network feature is based on node-level indices focusing on the "relevance" of nodes. Each user represents a node in the network and the ties between them are of two types: "friendship" (or following/followed) and "retweeting". Among several measures of network centrality available, we consider: indegree, outdegree, Kleinberg's hub and authority scores, closeness, betweenness, eigenvector, Bonacich's power, and Google PageRank [10]. First, we analyze the correlations among the indices for each language and for each type of network: when a pair of indices show a correlation higher than 0.8, we choose the one that was simplest in terms of computation and interpretation (e.g., indegree is preferred to PageRank). Then, we select the centrality indices that prove to discriminate most effectively among the stances, according to multinomial regressions and Krusal-Wallis tests with Nemenyi's non-parametric all-pairs comparisons. Following the results of these analyses, we select these feature groups:

- Network Friend Centralities (**NetFC**): a 4-dimensional feature measuring user centrality for the Spanish FRIENDS NETWORK (indices: indegree, outdegree, authority, and closeness), and a 3-dimensional feature measuring user centrality for the Basque FRIENDS NETWORK (indices: indegree, betweenness, and eigenvector).
- Network Retweet Centralities (**NetRC**): a 4-dimensional feature measuring user centrality for the RETWEETS NETWORK (Spanish indices: indegree, outdegree, hub, and authority; Basque indices: outdegree, authority, closeness, and eigenvector).
- Network Augmented Retweet Centralities (**ANetRC**): a 4-dimensional feature measuring user centrality for the AUGMENTED RETWEETS NETWORK, exploiting the same indices selected for the Spanish and Basque Retweets networks from the train and test set (**NetRC**).

NetFC and **NetRC** features are computed by combining the graphs of the train and test sets. In case some users are disconnected in the networks' graphs of friends and/or retweets, they are assigned a value of 0 for each of the computed centrality indices.

The second type of network feature is based on the distances among users in the networks. For each of them, a distance matrix among subjects is computed. The distance is defined as the shortest path, forcing the graph to be undirected. The Distance Matrix is then projected into a euclidean space trough a Multidimensional Scaling (MDS) [20]. Since we expect the users to be strongly polarized in clusters within the network, we also expect the largest dimension to discriminate among the stances. Therefore, we retain the first four dimensions for each of the four networks. This expectation is confirmed by Exploratory Data Analysis. The scatter plots of the first two dimension is shown in Figure 3. In almost all panels we can observe a separation of users with different polarization. This is perhaps more evident for Spanish users.

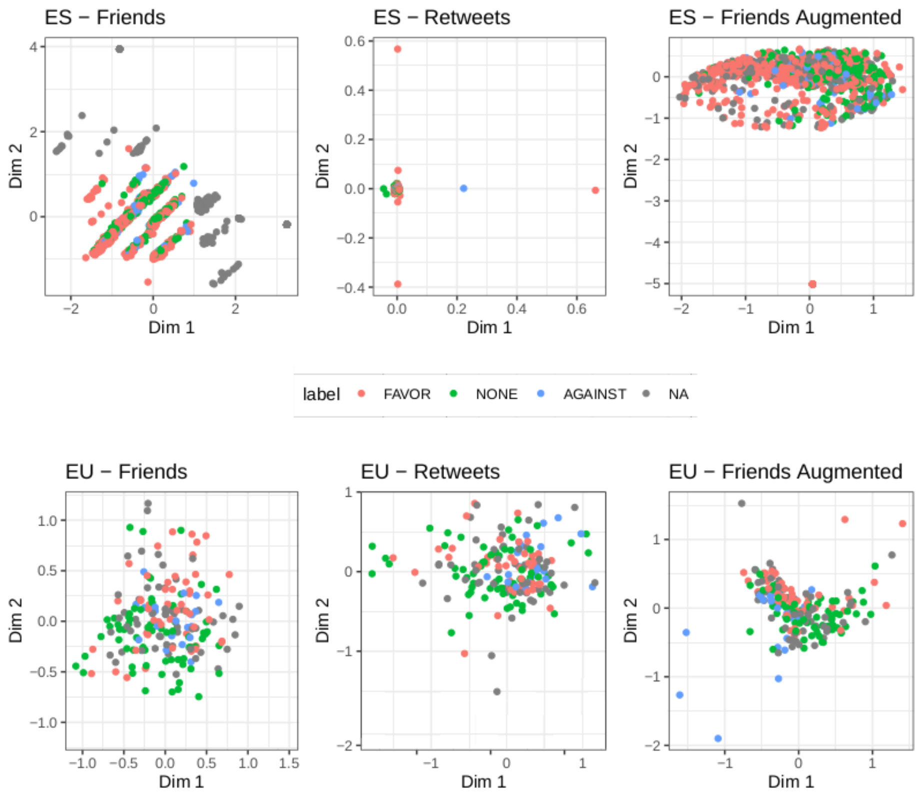

Fig. 3: First and second dimension of MDS for different languages and networks.

- Network Friend MDS (**NetFM**): A 4-dimensional feature has been extracted based on the first four dimensions of the MDS based on the FRIENDS NETWORK.
- Network Retweet MDS (**NetRM**): A 4-dimensional feature extracted from the RETWEETS NETWORK, as explained above.
- Network Augmented Retweet MDS (**ANetRM**): The same method has been used to extract the 4-dimensional feature form the AUGMENTED RETWEETS NETWORK.

The third type of network feature is based on the *homophily principle* which states that social networks contacts among similar users occur at a higher rate than among dissimilar ones [28]. We thus propose a 6-dimensional feature that counts the number of directed relations among the author of the tweet and the other users, grouping them by stance. More precisely, the feature considers the numbers of directed relations towards neighbors labeled as AGAINST (AGAINST$_{out}$ and AGAINST$_{in}$), FAVOR (FAVOR$_{out}$ and FAVOR$_{in}$), or NONE (NONE$_{out}$

and NONE$_{in}$). We then propose one feature for each employed network combining the users belonging to the train and the test set:

- Friends' Stance (**NetFS**): this feature counts the number of friendship relations among the author of the tweet and the other users, grouping them by stance, in the FRIENDS NETWORK.
- Retweets' Stance (**NetRS**): this feature counts the number of retweet relations among the author of the tweet and the other users, grouping them by stance, in the RETWEETS NETWORK.
- Augmented Retweets' Stance (**ANetRS**): this feature exploits the augmented retweets' network for counting the number of retweet relations, in the AUGMENTED RETWEETS NETWORK.

3 Experiments and Results

The organizers provided a training dataset of 2,400 tweets written in Spanish and 1,073 in Basque labeled with stance (AGAINST, FAVOR, and NEUTRAL) toward the topic of the *Antivaxxers movement*. The dataset also include additional information on the tweet itself, and contextual information about the authors (1,261 for Spanish, and 149 for Basque). The last contextual information available consists in the relations of friends and retweets among the authors and the other Twitter's users. The organizers have also made available the `Home_timeline` of the Basque authors, but we did not employ this information in any feature.

As mentioned earlier, the task is divided in four different tracks that establish different constraints on the use of data and external resources.

1. Close Track:
   (a) *Textual*: only the textual content of each tweet can be used to train a model. Therefore we resort to all the *Stylistic* features (**BoW**,**BoC**, **PM**, **UpW**, **Lau**, **PN**, **LC**) and to all *Dependency*-based features (**upos**, **drN**, **dr**, **rV**, **rN**, **rA**, **sF**, **sT**, **sDR**, **TC1**, **TC2**).
   (b) *Contextual*: models can be trained with any information provided by the organizers. Here, we take advantage of the features used in the previous track in addition to some *Twitter*-based features (**CloseTinfo**, **CloseTSource**, and **CloseUinfo**). We also employ some *Network*-based features (**NetFC**, **NetRC**, **NetFM**, **NetRM**, **NetFS**, and **NetRS**).
2. Open Track: it is the least restrictive track. Indeed, any kind of data can be used. We employ all the previous features and we also introduce features based on data augmentation (**ABoD**, **OpenCW**, **OpenLW**, **ANetRC**, **ANetRM**, and **ANetRS**).
3. Zero-Shot Track: in this track any kind of data can be used with the exception of the textual content of the training tweets. We employ all the previous features with the exception of *Stylistic*, *Lexica*-based, and *Dependency*-based ones.

We finally submitted 16 runs (2 for each track and for each language) choosing among the best combination of features that we retrieve by performing a feature selection. We employed a 5-fold cross validation over the train set via Logistic Regression.

		Basque		Spanish	
Track	Run	Features	F1 Macro	Features	F1 Macro
Close-Textual	01	BoW, BoC, PM, UpW, Lau, PN, drN dr, rV, sF, TC2	66.66	BoW, BoC, PM, UpW, Lau, PN, drN, dp, rV, sF, TC2	**76.54**
	02	BoW, BoC, LC, upos, rV, rN, sF	**68.82**	BoW, BoC, LC, upos, rV, rN, sF, TC2	76.39
Close-Contextual	01	Lau, PN, NetFC, NetFS, NetFM	**98.35**	BoW, BoC, NetFS	82.55
	02	BoW, BoC, PM, UpW, Lau, PN, drN, dr, rV, sF, TC2, NetFS, NetFM	57.49	BoW, BoC, PM, UpW, Lau, PN, drN, dp, rV, sF, TC2, NetFS	**83.04**
Open	01	NetFC, NetFS, NetFM, ANetRC, ANetRS, ANetRM,	**98.69**	BoW, BoC, NetFS, NetFM	83.23
	02	BoW, BoC, PM, UpW, Lau, PN, dpN, dr, rV, sF, TC2, NetFS, ANetRM	58.18	BoW, BoC, PM, UpW, Lau, PN, srN, dr, rV, sF, TC2, NetFS, AugNetR	**83.42**
Zero-Shot	01	NetFC, NetFS, NetFM, ANetRC, ANetRS, ANetRM	**98.69**	NetFC, NetFS, NetFM, ANetRC, ANetRS, ANetRM, ABoD	**78.54**
	02	ABoD	94.68	ABoD	76.19

Table 3: Results obtained with a 5-fold cross-validation on the training set by combining the diverse configurations of features.

Table 3 shows the combination of features submitted for each run. We include the F1 macro-average score of two classes FAVOR and AGAINST achieved with 5-fold cross validation over the training set.[17] The first thing that catches the eye is that three runs for Basque reach about 98 F1 macro-average (in the Close-Contextual, Open, and Zero-Shot tracks). A possible explanation is to be encountered in the fact that the social network based on friendship relations exhibits *homophily by stance*. In fact, users tend to follow people who have similar opinions to theirs [28]. The *Network*-based features have been conceived to provide this informative cue too. This hint seems to be very strong in the Basque social network, which is very small (only 6,451 nodes). This type of features achieves high performances in Spanish too. Indeed, by adding them to features based on the textual content of the tweet results are notably increased.

Another feature that seems to be very promising is **ABoD**, which profiles users by their *descriptions*. Indeed, the description often includes users' professions and interests that are very useful for predicting stance towards vaccinations (e.g., some of them are doctors and nurses). Example of a user description:

> Enfermera. 👩‍⚕️🇪🇸 Mención en Urgencias y Emergencias sanitarias. 🚑🏥
> (*Nurse.* 👩‍⚕️🇪🇸 *Specialization in ER and Health Emergencies.* 🚑🏥)

In the example above the user identifies herself as a nurse, using also the emoji 👩‍⚕️ identified by the shortcode `:woman_health_worker:` that combines the *woman* emoji (👩) and the *medical symbol* emoji (⚕️). In future work it would be interesting to exploit the emojis contained in user descriptions as feature, as well as word-embeddings representations in order to capture all the professions that are in any kind of relation to medical/health field.

Although using contextual features allows us to achieve high level of averaged F1 macro, also using *Stylistic* and *Dependency-based* features helps to improve results in both Basque and Spanish. Our findings, in fact, provide a meaningful support to the hypothesis that morphosyntactic knowledge extracted from treebanks can be usefully exploited for addressing the stance detection task. In particular, they pave the way for a further investigation where the combination of a dependency-based syntactic approach and state-of-the-art neural models can be explored. Thanks to dependency syntax it is possible to grasp connections among words that are not captured by n-grams or word embeddings standing alone. If we had only considered approaches such as those that take into account only the words that are in the immediate proximity of each other, the deeper pragmatic meaning of a sentence, might have been lost. Indeed, with dependency-based features is possible to capture the information in which words are syntactically related to each other also if they have a long-distance relation.

[17] All values in the result tables have been multiplied by 100 to enhance readability and to be consistent with official rankings provided by the organizers (see Table 4).

3.1 Official Rankings

In order to assess the performance of the participating systems, a test set of 694 and 312 unlabeled tweets were provided respectively for Spanish and Basque. The four tracks have been evaluated separately for each language. Three teams participated to Close-Textual and Close-Contextual tracks and we ranked as the first position in both sub-tasks for Basque and Spanish. The difference of our results from those of other teams is particularly evident.[18] Furthermore, our team is the only one that participated to the Open and Zero-shot tracks. Table 4 shows the official results, as provided on the task website, sorted by track and language setting.

	Basque				Spanish			
Track	Run	AGAINST	FAVOR	F1 Macro	Run	AGAINST	FAVOR	F1 Macro
Close-Textual	eu_01	57.69	56.99	**57.34**	es_01	75.54	82.58	79.06
	eu_02	55.03	54.27	54.65	es_02	78.36	83.47	**80.92**
Close-Contextual	eu_01	0.00	0.08	0.04	es_01	88.97	86.56	87.77
	eu_02	82.95	72.46	**77.71**	es_02	91.17	87.09	**89.13**
Open	eu_01	64.47	68.12	66.30	es_01	90.39	88.01	89.20
	eu_02	82.29	72.12	**77.21**	es_02	90.87	88.07	**89.47**
Zero-Shot	eu_01	64.47	68.12	**66.30**	es_01	88.03	46.13	**67.08**
	eu_02	55.70	39.74	47.72	es_02	18.63	62.77	40.70

Table 4: Official results sorted by track and language setting.

If we compare the scores of Table 3 with the scores of Table 4, there is a general drop in performance, with the exception of Close (textual and contextual) and Open Track for Spanish. Indeed, official results are higher than experimental ones in these cases. We can see right away that the run Close-Contextual eu_01 obtains a very low result (0.04 F1 Macro).

The features used in this particular track and setting are: **Lau**, **PN**, **NetFC**, **NetFS**, **NetFM** and they seem not to be relevant at all. The features are mostly based on the network of users, but the authors of the tweets belonging to the test set are not very connected to the authors of the train set. For this reason the system that uses only *Network*-based features is not able to detect a stance and tends to associate NONE labels. In any case, *Network*-based features, employed in addition to *Stylistic* and *Dependency*-based ones, reach 77.71 and 89.13 F1 macro-average respectively for Basque and Spanish. These results overcome those achieved in the Close-Textual track (57.34 and 80.92 F1 macro-average respectively for Basque and Spanish). The same situation can be observed in the Spanish setting: the use of contextual features (in particular the *Network*-based

[18] For the complete rankings refer to: https://vaxxstance.github.io/#results.

ones) improves the performance of our systems. The are no clear improvements in performance exploring data augmentation in the Open Track for both Spanish and Basque.

4 Error Analysis and Discussion

In order to gain more insights on the performance of our models we carry out two types of error analysis. On the one hand we look the tweets that have been misclassified and we observe them linguistically. On the other hand we carry out a test on the process of feature selection with an automatic tool.

4.1 Linguistic Analysis

In this subsection we compare the two outputs predicted with the two best performing models in Spanish (Close-Textual es_01 and Close-Contextual es_02). We focus only on the Spanish language, because, regrettably none of the authors is fluent in Basque.

We investigate the predictions especially of those two models because the they differ one from the other only by the employment of one feature. The Close-Textual es_01 model, in fact, employs the following features: *Bag of Words*, *Bag of Characters*, *Punctuation Marks*, *Uppercase Words*, *Laughter*, *Percentage Numbers*, *deprelNeg*, *deprel*, *relationVERBS*, *Sidorovbigramsform* and *Target Context Level 2*. The Close-Contextual es_02 model employs the same features as the other system, plus the addition of *Network Friends' Stance* (NetFS). Due to this reason, we believe their comparison might lead to interesting discoveries. In fact, by the single addition of one last feature, results are boosted up +10 points in terms of F1 Macro in Spanish (from 79.06 to 89.13, see Table 4). The same feature (together with *NetFM*), tested on the Basque dataset, induces a boost of +20 points of Macro F1 (from 57.34 to 77.71, see Table 4).

In Table 5 we report the confusion matrices of the labels predicted by the two models compared against the gold test set.

		PREDICTED		
		AGAINST	FAVOR	NONE
GOLD	AGAINST	105	22	13
	FAVOR	21	294	44
	NONE	12	37	146

(a) Close-Textual (es_01).

		PREDICTED		
		AGAINST	FAVOR	NONE
GOLD	AGAINST	129	5	6
	FAVOR	7	307	45
	NONE	7	34	154

(b) Close-Contextual (es_02).

Table 5: Confusion matrices of the errors in two different tracks.

From Table 5, if we compare the left table (a) with the right table (b), it can be seen that both precision and recall increase for the class AGAINST. The results suggest that the Close-Contextual model, benefiting from the *NetFS* feature, shows an improvement in the detection of highly polarized labels (AGAINST vs. FAVOR), while on the other hand, it continues to fail to successfully discern between less polarized choices (e.g., FAVOR vs. NONE or AGAINST vs. NONE). Furthermore, the *Network Friends' Stance* feature is highly proficient in detecting correctly especially the tweets that are labeled as AGAINST. This highlights the fact that authors that follow (and are followed) by other users who are skeptical about vaccinations, tend to be AGAINST the vaccines. Similar findings within polarized debates have been found also by Lai et al. [25,26]).

Additionally, we observe that 12 tweets that have been correctly predicted by the Close-Textual es_01 model were misclassified by the Close-Contextual es_02. On the opposite end, there are other 57 tweets that were misclassified from the Close-Textual es_01 model, but that were classified correctly by the Close-Contextual es_02 model. This procedure might lead to better understanding and plausible explanations of the usefulness of the **NetFS** feature.

Furthermore, we observed the impact of the *Network Friends' Stance* feature and its contribution in stance detection. We recall that such feature counts for each user the number of directed friendship relations (IN or OUT) grouped by stance. For instance, the tweet below has been wrongly classified in the Textual Track (AGAINST), but was labeled correctly in the Contextual Track (FAVOR):

> Cuando tienes la sensación de que la sociedad va hacia atrás en vez de avanzar, y todo por las malas decisiones del ser humano...Ni a sus hijos ni a sus perros: el nuevo peligro del movimiento antivacunas son las mascotas https://t.co/AvLXpf8G91 en @elpais_espana
> *(When you have the feeling that society is going backwards instead of moving forward, and all because of the bad decisions of human beings...Not their children or their dogs: the new danger of the anti-vaccine movement is pets* https://t.co/AvLXpf8G91 *in @elpais_espana)*
>
> GOLD: FAVOR
>
> CLOSE-TEXTUAL ES_01: AGAINST
> CLOSE-CONTEXTUAL ES_02: FAVOR

Interestingly, we observed 7 friendship relationships with users labeled as AGAINST (3 IN, 4 OUT) versus 12 friendship relationships with users labeled as FAVOR (4 IN, 8 OUT). Here, the use of the **NetFS** feature contributed to the correct classification of the tweet.

Although, using this approach might lead also to the opposite deduction. In fact, some tweets have been correctly classified by the Textual model, but were wrongly labeled by the Contextual model. For instance, in the following tweet, the **NetFS** feature pointed the Contextual Model towards the label AGAINST, since it observed 21 friendship relationships with users labeled as AGAINST (10 in, 11 out) versus 12 friendship relationships with users labeled as favor (5 in, 7 out):

> 🚨 Un alcalde y líder de las Juventudes del PSOE de 29 años en Valencia se vacunó del coronavirus el primer día 🚨Rojos haciendo cosas de rojos...#ADisfrutarDelPucherazohttps://t.co/Dllq3pRVAl
> (🚨 *A 29-year-old mayor and leader of the PSOE Youth in Valencia was vaccinated against the coronavirus on the first day* 🚨*Reds doing red things...#LetsEnjoyThePout*https://t.co/Dllq3pRVAl)
>
> gold: none
>
> close-textual es_01: none
> close-contextual es_02: against

For the author of the tweet above we did not detect any relationships with users labeled as none, hence we can speculate that the **NetFS** feature led to a misclassification.

4.2 Feature Analysis

In this subsection, we aim at including some analysis on both languages, since unfortunately Basque was overlooked in the previous manual analysis. We exploit a univariate feature selection by sampling the best 100 features based on the ANOVA F-value for the test set.

According to the ANOVA test, the best features exploited in the run Close-Textual es_01 mainly include n-grams and char-grams belonging to the BoW and BoC groups of features. If on the one hand we encounter n-grams such as "*somoslaresistencia*" (we are the resistance) and "*vacuna contra*" (against vaccines), on the other hand, we find several features attributable to the presence of URLs in the text. This confirms similar findings from earlier studies [23,24,34] that highlighted the significance of a feature based on the analysis of URLs for detecting stance towards the Catalan Independence in tweets written in Spanish and Catalan, and for disclosing substantive actions for sustainable development in tweets written in Spanish. The Close-Contextual es_02 run differs from Close-Textual es_01 only for the presence of the *Network Friends' Stance* feature (**NetFS**). Indeed, the 100 highest scoring features include most of the features detected in the previous case, but they also include three features belonging to the **NetFS** group of feature (none$_{out}$, none$_{in}$, and favor$_{in}$). It confirms the soundness of this group of features for detecting stance towards vaccination.

We also find similar results inspecting the highest scoring features in the runs submitted for Basque. The best 100 features of the run Close-Textual eu_01 includes n-grams such us "*zientzia*" (science) and char-grams attributable to the word "*txertoa*" (vaccine) belonging to the BoW and BoC groups of features. We also find several n-grams and char-grams features attributable to the presence of URLs in the text, similarly to what was observed in the analysis of the run for Spanish. The Close-Contextual eu_02 run differs from Close-Textual eu_01 only for the presence of the *Network Friends' Stance* (**NetFS**) and *Network Friend MDS* (**NetFM**) features. The features based on social media networks prove to be also relevant in Basque, in particular the best performing features include 5 out of 6 **NetFS** features (NONE_{out}, AGAINST_{in}, AGAINST_{out}, FAVOR_{in}, and FAVOR_{out}) and 2 out of 3 dimensions of the **NetFM** feature.

5 Conclusions

In this paper we presented an overview of the WordUp! submission for the *VaxxStance* task at IberLEF 2021. We participated in all the proposed tracks by submitting 16 different runs in the detection of author's stance towards the target '*vaccines*' for tweets in Basque and Spanish. Our approach, mainly employing *Stylistic*, *Dependency*-Based and *Contextual* features, proved to be highly successful concerning the task of stance in both languages. We ranked as the first position among three participating teams in all tracks in both languages. The results show that the addition of contextual features such as *Network*-Based ones, produced a significant contribution to the stance detection task. For instance, the exploitation of the *Network Friends' Stance* feature (NetFS) induced a boost of +10 points in terms of F1 Macro for Spanish, and together with *NetFM* +20 points for Basque. Therefore, we might interpret that contextual information and the network of users are the richest exploitable information in this SD task, and, that the combination of linguistic information with contextual features leads to more explainable results.

In the future, we plan to tailor the **ABoD** feature (Augmented Bag of Description) exploring in an even finer grained manner the content of the user's description. We also aim at exploring the contribution of the *Dependency*-based features for predicting stance in an unsupervised framework.

6 Availability of Materials

The code, the models and the resources used in this study are freely available online, allowing for an easy replication of the presented results. They can be found in the following repository: https://github.com/mirkolai/WordUp.

Acknowledgments

Detailed contribution of each author:
⋆ *Mirko Lai:* Conceptualization, Data Augmentation, Implementation of Stylistic Features, Target Content Features and Twitter-based Features, Formal Analysis, Methodology, Models Architecture, Supervision, Error Analysis, Writing.
⋆ *Alessandra Teresa Cignarella:* Conceptualization, Implementation of Syntactic Dependency-Based Features and Lexica-Based Features, Formal Analysis,

Methodology, Supervision, Error Analysis, Writing.
⋆ *Livio Finos:* Implementation of Network-based Features, Methodology, Formal Analysis, User Labeling, Writing.
⋆ *Andrea Sciandra:* Data Augmentation, Implementation of Network-based Features and Lexical Complexity Features, Formal Analysis, Error Analysis, Writing.

The work of the first two authors is partially funded by the project *Be Positive!* (under the 2019 "Google.org Impact Challenge on Safety" call).

References

1. Agerri, R., Centeno, R., Espinosa, M., Fernandez de Landa, J., Rodrigo, A.: VaxxStance@IberLEF 2021: Going Beyond Text in Crosslingual Stance Detection. Procesamiento del Lenguaje Natural 67(0) (2021)
2. Aker, A., Derczynski, L., Bontcheva, K.: Simple open stance classification for rumour analysis. In: Proceedings of RANLP 2017. INCOMA Ltd. (2017)
3. Aranzabe, M.J., Atutxa, A., Bengoetxea, K., de Ilarraza, A.D., Goenaga, I., Gojenola, K., Uria, L.: Automatic conversion of the Basque Dependency Treebank to Universal Dependencies. In: Proceedings of the fourteenth international workshop on treebanks an linguistic theories (TLT14) (2015)
4. Bahuleyan, H., Vechtomova, O.: UWaterloo at SemEval-2017 Task 8: Detecting Stance Towards Rumours with Topic Independent Features. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). ACL (2017)
5. Bazzanella, C.: Oscillazioni di informalità e formalità: scritto, parlato e rete. Formale e informale. La variazione di registro nella comunicazione elettronica (2011)
6. Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A.: quanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software **3**(30) (2018)
7. Bethard, S., Cer, D.M., Carpuat, M., Jurgens, D., Nakov, P., Zesch, T. (eds.): Proceedings of the 10th International Workshop on Semantic Evaluation. SemEval, ACL (2016)
8. Cignarella, A.T., Lai, M., Bosco, C., Patti, V., Rosso, P.: SardiStance@EVALITA2020: Overview of the Stance Detection in Italian Tweets. In: Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020) (2020)
9. CrowdTangleTeam: CrowdTangle (2021), list ID: [1533300, 1535071, 1533303, 1533293, 1533305, 1533302, 1535076, 1533304, 1533299, 1533307]
10. Csardi, G., Nepusz, T.: The igraph software package for complex network research. InterJournal, complex systems **1695**(5) (2006)
11. Derczynski, L., Bontcheva, K., Liakata, M., Procter, R., Wong Sak Hoi, G., Zubiaga, A.: SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. In: Proceedings of SemEval 2017. ACL (2017)
12. Dynel, M., Poppi, F.I.M.: Arcana imperii*: The power of humorous retorts to insults on Twitter. Journal of Language Aggression and Conflict **8**(1) (2020)
13. Ferraccioli, F., Sciandra, A., Da Pont, M., Girardi, P., Solari, D., Finos, L.: TextWiller @ SardiStance, HaSpeede2: Text or Con-text? A smart use of social network data in predicting polarization. In: Proceedings of the 7th Evaluation Campaign

of Natural Language Processing and Speech Tools for Italian (EVALITA 2020). CEUR-WS.org (2020)
14. Ghanem, B., Cignarella, A.T., Bosco, C., Rosso, P., Rangel Pardo, F.M.: UPV-28-UNITO at SemEval-2019 Task 7: Exploiting Post's Nesting and Syntax Information for Rumor Stance Classification. In: Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2019). ACL (2019)
15. González, J.Á., Pla, F., Hurtado, L.F.: Elirf-upv at ibereval 2017: Stance and gender detection in tweets. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) (2017)
16. Gorrell, G., Kochkina, E., Liakata, M., Aker, A., Zubiaga, A., Bontcheva, K., Derczynski, L.: SemEval-2019 Task 7: RumourEval, determining rumour veracity and support for rumours. In: Proceedings of SemEval 2019. ACL (2019)
17. Hogan, B.: The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. Bulletin of Science, Technology & Society **30**(6) (2010)
18. Igarashi, Y., Komatsu, H., Kobayashi, S., Okazaki, N., Inui, K.: Tohoku at SemEval-2016 Task 6: Feature-based Model versus Convolutional Neural Network for Stance Detection. ACL (2016)
19. Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., Mikolov, T.: Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651 (2016)
20. Klimenta, M., Brandes, U.: Graph drawing by classical multidimensional scaling: New perspectives. In: Didimo, W., Patrignani, M. (eds.) Graph Drawing. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
21. Küçük, D., Can, F.: Stance detection: A survey. ACM Computing Surveys **53**(1) (2020)
22. Lahmann, C., Steinkrauss, R., Schmid, M.S.: Factors Affecting Grammatical and Lexical Complexity of Long-Term L2 Speakers' Oral Proficiency. Language Learning **66**(2) (2016)
23. Lai, M., Cignarella, A.T., Farías, D.I.H., Bosco, C., Patti, V., Rosso, P.: Multilingual stance detection in social media political debates. Computer Speech & Language **63**(101075) (2020)
24. Lai, M., Cignarella, A.T., Hernandez Fariás, D.I.: iTACOS at IberEval2017: Detecting Stance in Catalan and Spanish Tweets. In: Proceedings of the 2nd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR-WS.org (2017)
25. Lai, M., Patti, V., Ruffo, G., Rosso, P.: #Brexit: Leave or Remain? The Role of User's Community and Diachronic Evolution on Stance Detection. Journal of Intelligent & Fuzzy Systems **39**, 2341–2352 (2020)
26. Lai, M., Tambuscio, M., Patti, V., Ruffo, G., Rosso, P.: Stance polarity in political debates: A diachronic perspective of network homophily and conversations on Twitter. Data & Knowledge Engineering **124**, 101738 (2019)
27. McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., et al.: Universal dependency annotation for multilingual parsing. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (2013)
28. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology **27**(1) (2001)
29. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)

30. Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: SemEval-2016 Task 6: Detecting Stance in Tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). ACL (2016)
31. Mohammad, S.M., Sobhani, P., Kiritchenko, S.: Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT) **17**(3) (2017)
32. Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Álvarez Carmona, M.A., Álvarez Mellado, E., Carrillo-de Albornoz, J., Chiruzzo, L., Freitas, L., Gómez Adorno, H., Gutiérrez, Y., Jiménez Zafra, S.M., Lima, S., Plaza-de Arco, F.M., Taulé, M.: In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021). CEUR-WS.org (2021)
33. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1-2) (2008)
34. Sciandra, A., Surian, A., Finos, L.: Supervised Machine Learning Methods to Disclose Action and Information in "U.N. 2030 Agenda" Social Media Data. Social Indicators Research (2020)
35. Sidorov, G.: Should syntactic n-grams contain names of syntactic relations?. International Journal of Computational Linguistics Applications **5**(2) (2014)
36. Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., Chanona-Hernández, L.: Syntactic dependency-based n-grams: More evidence of usefulness in classification. In: International Conference on Intelligent Text Processing and Computational Linguistics. pp. 13–24. Springer (2013)
37. Taulé, M., Martí, M.A., Pardo, F.M.R., Rosso, P., Bosco, C., Patti, V.: Overview of the Task on Stance and Gender Detection in Tweets on Catalan Independence. In: Proceedings of the 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017). CEUR-WS.org (2017)
38. Taulé, M., Martí, M.A., Recasens, M.: AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In: Proceedings of the Sixth Language Resources and Evaluation Conference (LREC 2008). European Language Resources Association (ELRA) (2008)
39. Taulé, M., Pardo, F.M.R., Martí, M.A., Rosso, P.: Overview of the Task on Multimodal Stance Detection in Tweets on Catalan #1Oct Referendum. In: Proceedings of the 3rd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018). CEUR-WS.org (2018)
40. Wei, W., Zhang, X., Liu, X., Chen, W., Wang, T.: pkudblab at SemEval-2016 Task 6 : A Specific Convolutional NeuralNetwork System for Effective Stance Detection. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (2016)
41. Zarrella, G., Marsh, A.: Mitre at Semeval-2016 Task 6: Transfer Learning for Stance Detection. arXiv preprint arXiv:1606.03784 (2016)
42. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., Procter, R.: Detection and Resolution of Rumours in Social Media: A Survey. ACM Computing Surveys **51**(2), 32 (2018)