

Do Dependency Relations Help in the Task of Stance Detection?

Alessandra Teresa Cignarella¹, Cristina Bosco¹ and Paolo Rosso²

1. Dipartimento di Informatica, Università degli Studi di Torino, Italy

2. PRHLT Research Center, Universitat Politècnica de València, Spain

alessandrateresa.cignarella@unito.it

crisrina.bosco@unito.it

prossso@dsic.upv.es

Abstract

In this paper we present a set of multilingual experiments tackling the task of *Stance Detection* in five different languages: English, Spanish, Catalan, French and Italian. Furthermore, we study the phenomenon of stance with respect to six different targets – one per language, and two different for Italian – employing a variety of machine learning algorithms that primarily exploit morphological and syntactic knowledge as features, represented throughout the format of Universal Dependencies. Results seem to suggest that the methodology employed is not beneficial *per se*, but might be useful to exploit the same features with a different methodology.

1 Introduction and Related Work

The task of monitoring people’s opinion towards particular targets in political topics or public life debates has grown in the last decade, thus leading to the creation of a specific area of investigation in NLP named *Stance Detection* (SD). Research on this topic, indeed, might have an impact on different aspects of everyone’s life such as public administration, policy-making, advertisement, marketing strategies and security. In fact, through the constant monitoring of people’s opinion, desires, complaints and beliefs on political agenda or public services, administrators could better meet population’s needs (Küçük and Can, 2020).

SD, as a task, shares various similarities with Sentiment Analysis (SA), and, exactly like Sentiment Analysis, also SD has been applied in several domains. For instance, to discover the reputation of an enterprise, what is the general public thought regarding a political reform, if costumers of a fashion brand are happy about the customer service, etc... Nevertheless, whereas the aim of SA is categorizing texts according to a notion of polarity (positive, negative or neutral), the aim of SD consists in classifying texts according to the attitude they express

towards a given target of interest (Mohammad et al., 2017).

The first shared task entirely dedicated to SD was held for English at SemEval in 2016, i.e., *Task 6 “Detecting Stance in Tweets”* (Mohammad et al., 2016). In the following years, many more shared tasks followed tackling the same issue in different languages and regarding different targets: Chinese (Xu et al., 2016), Spanish and Catalan (Taulé et al., 2017, 2018), Italian (Cignarella et al., 2020b), and lastly Spanish and Basque (Agerri et al., 2021).

Provided that several approaches based on different types of linguistic knowledge have been applied to address the SD task, in this paper we investigate the contribution of syntactic information and in particular that provided by dependency relations. Therefore, we exploit some of the datasets made available in the above-mentioned evaluation campaigns in different languages. In particular, we aimed at answering the following research question:

RQ: *Do features derived from morphology and syntax help automatic systems to address the task of stance detection?*

Indeed, some research already explored different kinds of syntactic features and their interaction in several NLP tasks, showing their effectiveness. For example, Sidorov et al. (2012) exploited syntactic dependency-based n-grams for general-purpose classification tasks, Socher et al. (2013) investigated sentiment and syntax with to the development of a sentiment treebank, and Kanayama and Iwamoto (2020) showed a pipeline method that makes the most of syntactic structures based on *Universal Dependencies* (UD¹), achieving high precision in sentiment analysis for 17 languages. Morphology and syntax have also been proved useful in a number of other tasks, such as rumour detection (Ghanem et al., 2019), authorship attribution (Posadas-Duran et al., 2014; Sidorov et al.,

¹<https://universaldependencies.org/>.

language	target and source	train				test			
		AGAINST	FAVOUR	NONE	TOTAL	AGAINST	FAVOUR	NONE	TOTAL
English	Hillary Clinton (Mohammad et al., 2016)	393	118	178	689	172	45	78	295
Spanish Catalan	Independence Referendum (Taulé et al., 2017)	335	1,446	2,538	4,319	84	361	636	1,081
		131	2,648	1,540	4,319	32	663	386	1,081
French	Emmanuel Macron (Lai et al., 2020)	244	71	109	424	64	20	22	106
Italian	Constitutional Referendum (Lai et al., 2020)	389	129	148	666	97	34	36	167
	Sardines Movement (Cignarella et al., 2020b)	1,028	589	515	2,132	742	196	172	1,110

Table 1: Benchmark datasets used for target-specific SD.

2014) and humor recognition (Liu et al., 2018). To the best of our knowledge there is no prior work exploiting dependency-based syntactic features for addressing the task of Stance Detection.

2 Methodology

The main goal of the experiments presented in this work consists in evaluating the contribution of syntax-based linguistic features extracted from the datasets described above to the task of SD. Therefore, we performed a set of experiments where several models were implemented exploiting classical machine learning algorithms and state-of-the-art language models implemented with the Python libraries *scikit-learn* and *keras*. The methodology we propose here, in which a multilingual setting is proposed and neural models are evaluated together with dependency-based features, recalls the idea that dependency based syntax might be useful in a variety of language scenarios for the task of SD and with a manifold of algorithms and architectures.

2.1 Datasets and pre-processing

Mirroring our previous work regarding irony detection in (Cignarella et al., 2020a), from which we replicate the methods and techniques used, we propose here to address SD as a multi-class classification task, testing an automatic system on five languages: English, Spanish, Catalan, French and Italian. Furthermore, with respect to Italian, we were able to test the approach on two different datasets with two different targets of interest, namely: the *Constitutional Referendum* (Lai et al., 2020) and the *Sardines Movement* (Cignarella et al., 2020b). In the multilingual experimental setting, we took advantage of three benchmark datasets made available during the last few years within evaluation campaigns, i.e., *SemEval 2016 - Task 6* (Mohammad et al., 2016), *StanceCat at IberEval 2017* (Taulé et al., 2017) and *SardiStance at EVALITA 2020* (Cignarella et al., 2020b), and two datasets

created *ad hoc* in the research group where we work, for previous studies on SD (with target Emmanuel Macron and Constitutional Referendum (Lai et al., 2020) and are freely available online.²

In Table 1, for each dataset, we report the language, the target of interest, the name of the shared task (or research) in which it was released through their paper reference, the number of tweets for each class (AGAINST, FAVOUR, NONE) and the total number of instances, for both training set and test set. The aim of our task is, thus, to determine the stance expressed by the user with respect to a given target.

In order to extract the information that is crucial for performing the experiments, we needed to apply also a layer of morpho-syntactic annotation to the corpora that are annotated only for SD. For this purpose, we selected the *standard de facto* Universal Dependencies and we benefited from the UDPipe³ tool. Considering that all the datasets used consist of Twitter data, whenever possible, we used resources where this genre, or at least user-generated content of some kind was included as training data for parsing. More precisely, the model for English has been trained on the EWT treebank (Silveira et al., 2014), that for Spanish on both GSD-Spanish corpus (McDonald et al., 2013) and the ANCORA corpus (Taulé et al., 2008). Also the model for Catalan was trained on the ANCORA corpus, while that for French on the GSD-French corpus (McDonald et al., 2013). Finally, the model for Italian was trained on the POSTWITA-UD corpus (Sanguinetti et al., 2018), on the ISTD treebank (Simi et al., 2014) and on the TWITTIRÒ-UD corpus (Cignarella et al., 2019).

The precision in this phase of the work can be a bottleneck for what concerns the accuracy of the

²<https://github.com/mirkolai/MultilingualStanceDetection/tree/master/dataset>.

³<https://ufal.mff.cuni.cz/udpipe>.

experiments that we will describe in the following sections. In fact, the approach is entirely based on dependency syntax and the results strictly depend upon the quality of parsed data. The performance of UDPipe’s parsing is close to the state-of-the-art ones, therefore, we considered the annotation obtained automatically reasonably acceptable for the present study. However there always is margin for some error, we assumed precision and error were similarly distributed in each language setting.

2.2 Features and Models

Firstly, tweets were stripped from URLs and characters were normalized to lowercase. Later, thanks to the application of the UDPipe pipeline we were able to generate dependency-based syntactic trees for all the tweets taken into consideration in each language (e.g., Figure 1).

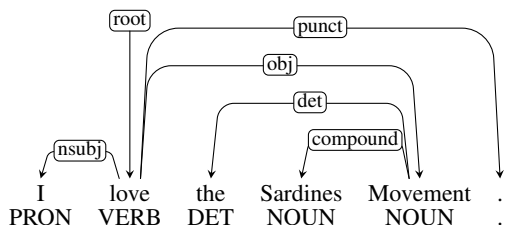


Figure 1: Example of a dependency tree in UD format.

On the basis of texts encoded in UD format, we engineered and tested the following features:

- ngrams,
chagrams;
- deprelneg,
deprel;
- relationformVERB,
relationformNOUN,
relationformADJ;
- Sidorovbigramsform,
Sidorovbigramsupostag,
Sidorovbigramsdeprel.

A detailed description for each feature is available in the Appendix and is inspired by our previous work (Cignarella et al., 2020a; Cignarella, 2021).

Having as primary goal the exploration of the features listed in the previous paragraph and testing their effectiveness in the task of SD, we fed them into a variety of models, including the following: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Multilayer Perceptron (MLP) and Multilingual BERT (M-BERT). The results obtained by combining all the features with

all the models listed above resulted in a very big amount of numbers, which most of the time were neither informative nor conclusive. Because of this we reported only the best scoring models in the section below.

3 Experiments and Results

We propose two different experimental settings. The first one aims at exploring the dependency-based features listed above paired with classical machine learning (ML) algorithms, in order to perform a feature selection and discover the best combination. In the second setting, we experiment with the Multilingual Bidirectional Encoder Representations from Transformers (M-BERT) and different additions of the features explored in the first setting.

3.1 Selection of best features

In order to identify the most relevant features, we tested different combinations of features and the models mentioned in Section 2.2 and we evaluated them according to the averaged macro F1-score.⁴

From the observation of Table 2 a vastly heterogeneous scenario emerges. There seems not to be any regular pattern among language scenarios, regarding the same features exploited for SD. On the contrary, the Multilayer Perceptron is proven to be the best performing classical ML algorithm across all languages, aside from the setting regarding the Constitutional Referendum in Italian. This has an explanation, that was already found out in precedent work (Lai et al., 2020) and a special clarification regarding the nature of the dataset is due. Indeed, the Italian dataset on the Constitutional Referendum seems to be particularly *sui generis* when compared with the other five. Within the dataset the exploitation of hashtags is wide and coherent in the whole corpus. For instance the hashtags #iovotosì (#Ivoteyes) and #ioivotono (#Ivoteno) have been exploited almost in each tweet that we took into consideration, and we believe that just their presence (as boolean value) already is a clear manifestation of stance. For this reason, only two features are already sufficient to reach an extremely high F1-score (0.967): ngrams and Sidorovbigramsupostag. The same reasoning applies to Support Vector Machines as they are sufficiently

⁴The average value obtained between the F1-score of the AGAINST class and the F1-score of the FAVOUR class as it was done in (Mohammad et al., 2016).

features	English Clinton	French Macron	Spanish Independencia	Catalan	Italian Referendum Sardines	
model	MLP	MLP	MLP	MLP	SVM	MLP
macro F1-score	.673	.596	.493	.497	.967	.651
ngrams			✓		✓	✓
chagrams	✓	✓	✓	✓		✓
deprel	✓		✓	✓		✓
deprelneg	✓	✓		✓		
relationformVERB	✓	✓	✓			
relationformNOUN		✓				
relationformADJ			✓			
Sidorovbigramsform		✓		✓		
Sidorovbigramsdeprel	✓			✓		✓
Sidorovbigramsupostag		✓	✓		✓	✓

Table 2: Features exploited in the best runs with classical ML algorithms in each language scenario.

language	target	best run (report and score)	SVM		M-BERT		
			+unigrams	—	+syntax	+best_feats	
English	H. Clinton	Zarrella and Marsh (2016)	.671	.570	.650	.562 (↓.088)	.636 (↓.014)
French	E. Macron	Lai et al. (2020)	.687	.526	.511	.511 (= .000)	.533 (↑.022)
Spanish Catalan	Independencia	Lai et al. (2017)	.489	.420	.467	.443 (↓.024)	.463 (↓.004)
		Lai et al. (2017)	.490	.468	.478	.462 (↓.016)	.476 (↓.002)
Italian	Referendum	Lai et al. (2020)	.971	.951	.959	.960 (↑.001)	.960 (↑.001)
	Sardines	Giorgioni et al. (2020)	.685	.578	.586	.599 (↑.013)	.563 (↓.023)

Table 3: Results obtained combining M-BERT and dependency-based syntactic features. Green values and arrows pointing up show an increment in performance with respect to results obtained by the bare architecture. Red values and arrows pointing down indicate a performance reduction, with respect to results obtained by the bare architecture. Orange values show no change.

good to perform textual classification where the only presence such a textual feature (a polarized hashtag) is so blatant in indicating stance.

From Table 2 it also emerges how in all the configurations used for achieving the best score at least one dependency-based syntactic feature was exploited and in particular those based on Sidorov et al.’s work, i.e., the last three rows of the table. This provides evidence for giving a partial answer to our research question (*Do features derived from morphology and syntax help automatic systems address the task of stance detection?*), since those are the features where the structure from root to branches of syntactic trees is encoded.

3.2 Syntactically-informed M-BERT

In the second setting, we performed experiments where, for each language scenario, we ran the straightforward M-BERT model. We also implemented the base architecture by adding the dependency-based syntactic features detailed in previous sections in two different ways, in order to have a clear-cut evidence on the actual contribution derived from dependency syntax to SD.

In Table 3 we report the results of the best system exploiting these datasets. Furthermore, we added the baseline results achieved with a SVM and a bag of words of unigrams, as it is the most common baseline proposed in most SD shared tasks. Each of the experiments with M-BERT has been performed 5 times with fixed hyper-parameters⁵ in order to take into account the differences of random initialization, and the average macro F1 score of such number of runs is reported.

Firstly, it is interesting to see how, the M-BERT base architecture never surpasses the results obtained with more complex architectures such as those proposed by the participants of shared tasks, confirming the complexity of the task.

Moreover, by having a look at the colorful right-hand side of Table 3, it can be seen how the addition of syntactic knowledge (M-BERT+syntax) determined a widely varied spectrum of outcomes. By the predominance of the colours orange and red (indicating stasis or loss in terms of performance), it is obvious to state that morphosyntactic information, taken alone and encoded into the M-BERT

⁵BatchSize = 8, LearningRate = $1e - 5$, EarlyStop = 5.

architecture does not provide strong nor consistent beneficial contribution to the task of SD. Not only the results obtained by the models *M-BERT+syntax* and *M-BERT+best_feats* obtain results lower than the state of the art approaches, but in most cases, they result in being also lower than the results obtained with the base architecture (M-BERT). Lastly, it is furthermore arguable that results show low (almost to none) statistical significance. In order to verify that, we applied the *t-test* with the Bonferroni correction and the outcomes have shown indeed not to be statistically significant. It might be worth it to explore new ways of encoding such features and integrating them into BERT, and also to perform new experiments with other BERT-based architectures that are language specific, rather than using the multilingual version (AIBERT_o for Italian (Polignano et al., 2019), BETO for Spanish (Cañete et al., 2020), CamemBERT for French (Martin et al., 2019), etc...).

4 Discussion and insights

The outcomes obtained in the investigation are slightly disappointing, but they do not come as a total surprise. When we were formulating the research question regarding SD, we had anticipated that there were no linguistic theories nor research work pointing towards the fact that morphosyntax might prove useful in this task. Furthermore, a clear explanation could be seen by observing how two simple sentences having opposite stance, present identical syntactic structure:

Ex.1 I love the Sardines Movement.

Ex.2 I hate the Sardines Movement.

we had already anticipated that taking morphology and syntax as only features to detect stance might indeed be calling a long shot.

With the experience matured with this research, we can state that – even if we are not obtaining the new state-of-the-art results – the outcomes lead in the direction of further investigation, pointing mainly towards a better understanding of features’ behaviour when stacked in a pre-trained language model such as M-BERT.

Finally, even though the results obtained with M-BERT turned out to be not statistically significant, this research was oriented in studying whether some features derived from morphology and syntax could help automatic systems to address the task

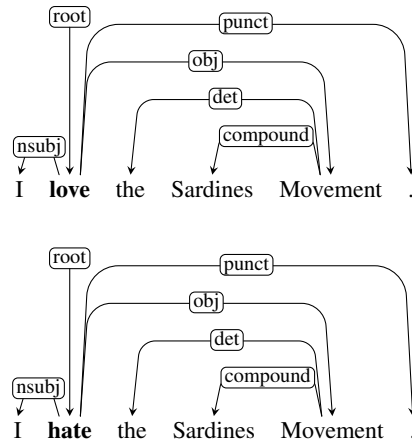


Figure 2: Dependency trees of Ex.1 and Ex.2.

of stance detection. It would be unfair not mentioning the fact that in the first experimental setting, that was mainly dedicated to the selection of the best features to be later fed as linguistic input into M-BERT, we actually obtained better results with respect to state-of-the-art models in four languages out of six and in the remaining two we obtained close results that are definitely comparable (see the macro F1-score of the best ML systems in Table 2 and compare it with the best results from shared tasks reported in Table 3).

5 Conclusion

The lesson learned from this work suggests that morphosyntactic cues combine well as features in classical machine learning algorithms, but they do not seem to provide an increment in terms of performance in the neural architecture of M-BERT in the case study of multilingual Stance Detection. Indeed, as shown in linguistics, the expression of one’s stance is frequently a phenomenon that seems to depend more often on semantics rather than on syntactic patterns or constructions.

Acknowledgements

This work has been partially funded by “Be Positive!” (under the 2019 “Google.org Impact Challenge on Safety” call) and also partially supported by the European project “STERHEOTYPES - Studying European Racial Hoaxes and stereOTYPES” funded by Compagnia di San Paolo and Volkswagen Stiftung under the “Challenges for Europe” call for Project (CUP: B99C20000640007). The work of the last author was partially funded by the Generalitat Valenciana under DeepPattern (PROMETEO/2019/121).

References

- Rodrigo Agerri, Roberto Centeno, María Espinosa, Joseba Fernandez de Landa, and Alvaro Rodrigo. 2021. VaxxStance @ IberLEF 2021: Overview of the Task on Going Beyond Text in Cross-Lingual Stance Detection. *Procesamiento del Lenguaje Natural*, 67:173–181.
- José Cañete, Gabriel Chaperon, Rodrigo Fuentes, and Jorge Pérez. 2020. Spanish Pre-Trained BERT Model and Evaluation Data. In *to appear in PMLADC at ICLR 2020*.
- Alessandra Teresa Cignarella. 2021. *Dependency Syntax in the Automatic Detection of Irony and Stance*. Ph.D. thesis, Università degli studi di Torino and Universitat Politècnica de València.
- Alessandra Teresa Cignarella, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, Farah Benamara, and Paolo Rosso. 2020a. Multilingual Irony Detection with Dependency Syntax and Neural Models. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*. ACL.
- Alessandra Teresa Cignarella, Cristina Bosco, and Paolo Rosso. 2019. Presenting TWITTIRO-UD: An Italian Twitter Treebank in Universal Dependencies. In *Proceedings of the 5th International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*. ACL.
- Alessandra Teresa Cignarella, Mirko Lai, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020b. 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)*.
- Bilal Ghanem, Alessandra Teresa Cignarella, Cristina Bosco, Paolo Rosso, and Francisco Manuel Rangel Pardo. 2019. 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.
- Simone Giorgioni, Marcello Politi, Samir Salman, Danilo Croce, and Roberto Basili. 2020. UNITOR@Sardistance2020: Combining Transformer-based architectures and Transfer Learning for robust Stance Detection. In *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. CEUR-WS.org.
- Hiroshi Kanayama and Ran Iwamoto. 2020. How Universal are Universal Dependencies? Exploiting Syntax for Multilingual Clause-level Sentiment Detection. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. ELRA.
- Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.
- Mirko Lai, Alessandra Teresa Cignarella, Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. Multilingual stance detection in social media political debates. *Computer Speech & Language*, 63(101075).
- Mirko Lai, Alessandra Teresa Cignarella, and Delia Irazú Hernández Farías. 2017. iTACOS at IberEval2017: Detecting Stance in Catalan and Spanish Tweets. In *Proceedings of the 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017)*. CEUR-WS.org.
- Lizhen Liu, Donghai Zhang, and Wei Song. 2018. Exploiting syntactic structures for humor recognition. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics (ACL).
- Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villamonte de la Clergerie, Djamé Seddah, and Benoît Sagot. 2019. CamemBERT: a Tasty French Language Model. *arXiv preprint arXiv:1911.03894*.
- Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, et al. 2013. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*. ACL.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*. ACL.
- Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):1–23.
- Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. ALBERTO: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets. In *Proceedings of the 6th Italian Conference on Computational Linguistics (CLiC-it 2019)*. CEUR-WS.org.
- Juan-Pablo Posadas-Duran, Grigori Sidorov, and Ildar Batyrshin. 2014. Complete syntactic n-grams as style markers for authorship attribution. In *Proceedings of the Mexican International Conference on Artificial Intelligence (MICA I 2014)*. Springer.
- Manuela Sanguinetti, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, Oronzo Antonelli, and Fabio Tamburini. 2018. PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*. ELRA.

- Grigori Sidorov. 2014. Should Syntactic N-grams Contain Names of Syntactic Relations? *International Journal of Computational Linguistics Applications*, 5(2):25–47.
- Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. 2012. Syntactic dependency-based n-grams as classification features. In *Proceedings of the Mexican International Conference on Artificial Intelligence (MICA I 2012)*. Springer.
- Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. 2013. Syntactic dependency-based n-grams: More evidence of usefulness in classification. In *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*. Springer.
- Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. 2014. Syntactic n-grams as Machine Learning Features for Natural Language Processing. *Expert Systems with Applications*, 41(3):853–860.
- Natalia Silveira, Timothy Dozat, Marie-Catherine De Marneffe, Samuel R Bowman, Miriam Connor, John Bauer, and Christopher D Manning. 2014. A gold standard dependency corpus for English. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. ELRA.
- Maria Simi, Cristina Bosco, and Simonetta Montemagni. 2014. Less is More? Towards a Reduced Inventory of Categories for Training a Parser for the Italian Stanford Dependencies. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. ELRA.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*. ACL.
- Mariona Taulé, Maria Antònia Martí, Francisco M. Rangel Pardo, Paolo Rosso, Cristina Bosco, and Viviana Patti. 2017. 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.
- Mariona Taulé, Maria Antònia Martí, and Marta Recasens. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*. ELRA.
- Mariona Taulé, Francisco M. Rangel Pardo, M. Antònia Martí, and Paolo Rosso. 2018. 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.
- Ruifeng Xu, Yu Zhou, Dongyin Wu, Lin Gui, Jiachen Du, and Yun Xue. 2016. Overview of NLPCC shared task 4: Stance detection in chinese microblogs. In *Natural Language Understanding and Intelligent Applications*. Springer.
- Guido Zarrella and Amy Marsh. 2016. MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*. ACL.

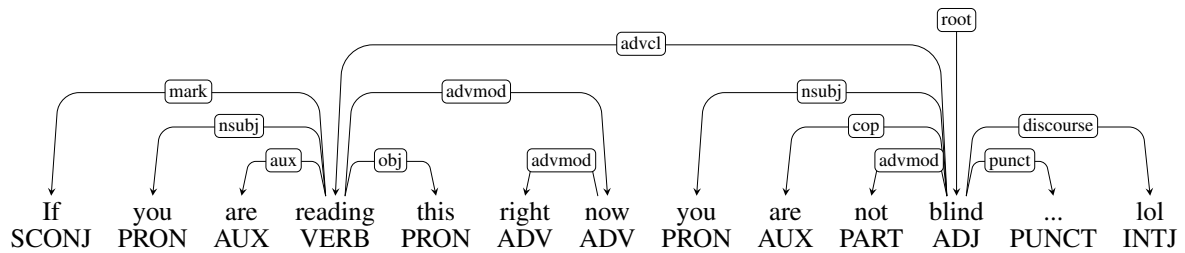


Figure 3: Dependency-based syntactic tree of an English tweet.

Appendix

The description of features as well as the content of the vectors for the syntactic features we developed, referring to the tweet in Figure 3, are as follows:

- n-grams: We extracted unigrams, bigrams and trigrams of tokens; e.g., [*If, you, are, reading, ..., If you, you are, are reading, ..., If you are, you are reading, are reading this, ...*];
- char-grams: We considered the sequence of char-grams in a range from 2 to 5 characters; eg [*If, fy, yo, ou, ..., Ifyou, fyoua, youar, ouare, uarer, ...*];
- deprelneg: We considered the presence of negation in the text, relying on the morphosyntactic cues present in the UD format. When a negation was present, we appended the correspondent dependency relation in the feature vector. For instance in Figure 3, we spot a negation in [*... are not blind ...*], the dependency relation of “not” is *advmod*, therefore, we append it in the feature vector;
- deprel: We built 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 nsubj aux obj advmod, nsubj aux obj advmod advmod, ..., advmod advmod nsubj cop advmod root punct, advmod nsubj cop advmod root punct discourse*];
- relationformVERB: 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 the first verb is “*reading*” and some of the tuples of tokens connected through this verb are, e.g., [*IfVERBthis, youVERBthis, areVERBthis, IfVERBnow, youVERBnow, ...*];
- relationformNOUN: We applied the same procedure of the feature above but considering nouns as starting points for collecting tuples;
- relationformADJ: in the same fashion of the two features above, we repeated the same procedure for adjectives too;
- Sidorovbigramsform: We created a bag of word-forms (tokens), considering the bigrams that can be collected following the syntactic tree structure (rather than the bigrams that can be collected reading the sentence from left to right).⁶ Such that: e.g., [*blind reading, blind you, blind are, blind not, reading if, reading you, ...*];
- Sidorovbigramsupostag: as the feature above, we created a bag of part-of-speech tags;
- Sidorovbigramsdeprel: as the two features above, we created a bag of words based on dependency relations (*deprels*).

⁶Please refer to (Sidorov et al., 2013) and (Sidorov, 2014) for more details on this regard.