

Debating Europe: A Multilingual Multi-Target Stance Classification Dataset of Online Debates

Valentin Barriere*, Alexandra Balahur*, Brian Ravenet†

*European Commission – DG-JRC, via Enrico Fermi 2749, 21027 Ispra (VA), Italy

†University Paris-Saclay, CNRS, LISN, CPU Team, 91400, Orsay, France
name.surname@ec.europa.eu, brian.ravenet@upsaclay.fr

Abstract

We present a new dataset of online debates in English, annotated with stance. The dataset was scraped from the “*Debating Europe*” platform, where users exchange opinions over different subjects related to the European Union. The dataset is composed of 2600 comments pertaining to 18 debates related to the “*European Green Deal*”, in a conversational setting. After presenting the dataset and the annotated sub-part, we pre-train a model for a multilingual stance classification over the X-stance dataset before fine-tuning it over our dataset, and vice-versa. The fine-tuned models are shown to improve stance classification performance on each of the datasets, even though they have different languages, topics and targets. Subsequently, we propose to enhance the performances over “*Debating Europe*” with an interaction-aware model, taking advantage of the online debate structure of the platform. We also propose a semi-supervised self-training method to take advantage of the imbalanced and unlabeled data from the whole website, leading to a final improvement of accuracy by 3.4% over a Vanilla XLM-R model.

Keywords: Stance Classification, Online Debates, Multilingual

1. Introduction

Stance detection and classification in online debates have been tackled by various approaches. Some of the first ones employed linguistics-based methods inside debates using pre-defined opposed targets such as “*iPhone vs BlackBerry*” (Somasundaran and Wiebe, 2009), classifying ideological debates (Somasundaran and Wiebe, 2010) and on social justice subjects such as “*Abortion*” or “*Gay Rights*”. They were followed by more complex probabilistic graphic systems (Sridhar et al., 2015), allowing to model the dynamics of the debate and the disagreements between speech turns, and finally deep neural methods (Augenstein et al., 2016; Allaway and McKeown, 2020), allowing efficient multi-target and zero-shot classification.

Recently, most of the work in this area focused on stance detection over tweets either in a non-interactive manner, like the SemEval-2016 task (Mohammad et al., 2016), or by including the interactions between the users (Barriere et al., 2018; Barriere, 2017) and applying stance detection over the whole thread (Gorrell et al., 2019). Building on seminal work in stance, the SemEval 2016 task was capable of targeting abstract concepts (e.g. “*Atheism*” or “*Abortion*”), as well as persons (e.g. “*Hillary Clinton*” or “*Donald Trump*”). On multilingual stance analysis over tweets, (Lai et al., 2020) present a model using mainly high-level linguistic features like stylistic, structural, affective or contextual knowledge, but no dense contextual vectors.

In (Vamvas and Sennrich, 2020), the authors propose the X-stance dataset, containing 67k comments over 150 political issues in 3 languages. Their approach was to reformulate the target in a natural question in order to easily train one multilingual multi-target model on the entire dataset. Similarly, in the *procon* dataset, contain-

ing 6,019 comments over 419 controversial issues, each target was also reformulated as a question (Hosseinia et al., 2020). However, none of these datasets contains interactional data.

The integration of the debate’s dynamics in the model can be done in many ways. It can be achieved using dialogic features (Abbott et al., 2011) or intrinsically in the shape of a graphical model (Walker et al., 2012; Sridhar et al., 2015), allowing to represent the dialogic structure of the debates which is important in term of agreements. Eventually, this integration was accomplished with transformer models like BERT (Prakash and Madabushi, 2020; Yu et al., 2020; Devlin et al., 2018). To the best of our knowledge, no work with transformers so far investigates the use of a context window, like us, for multi-target stance detection in debates. Self-training (ST) (Yarowsky, 1995) is interesting when annotation is scarce, but however rarely used for stance detection and even less with imbalanced data. A recent work is the one of (Glandt et al., 2021) that use Knowledge Distillation on COVID tweets. (Wei et al., 2021) propose an interesting self-training method for imbalanced images on CIFAR, but they assume the distributions of the unlabeled and labeled datasets are the same, which is not true in our case.

Motivations and Positioning The first motivation of this work relates to the lack of an appropriate multilingual multi-target stance-annotated debate dataset. We created such a corpus, together with the appropriate annotation schema and guidelines. It is composed of contemporary questions that can be debated in the Conference on the Future of Europe. The contributions of this paper are four-fold. Firstly, we propose a new dataset of annotated stance in online debates. Secondly,

¹<https://futureu.europa.eu/?locale=en>

Figure 1: Examples of comments from 3 debates of the Debating Europe Dataset

Can renewables ever replace fossil fuels 100%?	How to avoid an energy crisis ?	Should Europeans be encouraged to eat more sustainably?
... That would be the ultimate reality as fossil fuels are a non-renewable and antiquated energy resource. Carmelo A.	... Hydrogen! Germany is already studying and improving hydrogen as energy for cars. If cars can run... Vincente S.	... I agree that the key to influence consumers is through education. I think one major problem that... Alma
In foreseeable future and without subsidies, not. Pawet K.	Vicente, obtaining hydrogen from water is an endothermic reaction. I believe approximately -8 kCal/mole.... Jovan I.	What a great idea! Silvia P.
No , well not the current ones ,and please get rid of those horrible giant wind turbines Kevin	Well I am not an engineer, but once I saw a documentary about Hydrogen and they were showing... Vincente S.	Where's the freedom to chose whatever you want to eat, to wear, to do? Borislav V.
...

Label	% DE	Unit	μ_{com}	μ_{deb}	Σ
✗	100%	Comments	\emptyset	89.5	125,798
		Words	51.7	4,623	6,499,625
✓	2.0%	Comments	\emptyset	140	2,523
		Words	33.4	4,683	84,289

Table 1: Low-level statistics on the DE dataset, regarding there is label annotation or not. μ_{com}/μ_{deb} is the average mean of the respective units (comments or words) at the comment/debate-level.

we assess the quality of the data and annotation by showing that our dataset can be used to improve stance classification in non-English languages. Indeed, pre-training on English text stemming from Debating Europe (DE) allows us to reach better results on the multilingual X-stance dataset (Vamvas and Sennrich, 2020). Thirdly, we take advantage of the debate structure inside the learning model and analyze its impact on the performances. Finally, we show that self-training can be used on the unlabeled part of the dataset to enhance the model performances.

We differ from the existing works for three reasons. Firstly the dataset we are proposing allows to study stance in online debates in a multi-target and multilingual way. Secondly, we propose to use a context window in order to integrate the dynamics of the debate in a context-aware transformer model. Finally, we not only release an annotated dataset for one domain, but also a larger dataset of unlabeled data on other topics, and show how to enhance a multilingual stance classifier with a simple, yet efficient semi-supervised learning method for imbalanced and unlabeled datasets.

2. Datasets Overview

2.1. The Debating Europe dataset

We release the Debating Europe (DE) dataset which is composed of online debates annotated with stance annotations at the comment level.

2.1.1. Debating Europe and Extraction

The DE dataset is composed of debates scraped in September 2020 from the “Debating Europe” plat-

form². Most of the debates are related to questions such as “Should we have a European healthcare system?”, which can generally be reformulated as a yes/no question. Each debate is composed of a topic tag, a text paragraph with the context of the debate, as well as comments, either about the main context or about previous comments.

The dataset contains 125,798 comments for 1,406 debates. More statistics are shown in Table 1.

2.1.2. Annotation

Subset selection We annotated 18 debates from the whole dataset scraped from Debating Europe. The criteria chosen to select those debates are the number of comments associated to each debate and the relevance to one or more of the policy areas of the new “European Green Deal”³.

When needed, the debate question was reformulated into a closed question in order to make it compatible with our framework. We discarded the debates with less than 25 comments. More information about the debates and policy areas are available in the Appendix.

Annotation scheme The annotation scheme and corresponding guidelines aimed to capture citizens’ stance towards the debate question, at the comment-level. To achieve this, four labels were defined: *Yes*, *No*, *Neutral* and *Not answering*. For each comment, the annotation regarded whether the user replied to the answer and if so, whether if he/she was in favour or not, or neutral with respect to the original question. The questions of the annotated debates are shown in Appendix. The annotation has been done by one unique expert using the INCEpTION software (Klie et al., 2018).

Final annotations We obtained 2,523 labels over the 18 debates, with 4 classes: *Yes* (40.1%), *No* (19.4%), *Neutral* (11.2%) and *Not answering* (29.3%). We chose to add the last category in order to check if the commenter was interested in answering the debate question. In the following experiments we merged the *Neu-*

²<https://www.debatingeurope.eu/>

³<https://tinyurl.com/GreenDealEC>

	Intra-target			X-question			X-Topic			X-lingual
	DE	FR	Mean	DE	FR	Mean	DE	FR	Mean	IT
M-BERT (Vamvas2020)	76.8	76.6	76.6	68.5	68.4	68.4	68.9	70.9	69.9	70.2
XLM-R	76.3	78.0	77.1	71.5	72.9	72.2	71.2	73.7	72.4	73.0
XLM-R _{ft}	77.3	79.0	78.1	71.5	74.8	73.1	72.2	74.7	73.4	73.9

Table 2: Results over X-Stance dataset for a binary classification

tral and *Not answering* classes into a unique class in order to simplify the work (Mohammad et al., 2016; Küçük and Fazli, 2020). The validation using classical inter-annotator-agreement metrics was impossible with one unique expert annotations, hence we validated the dataset by showing its usefulness for cross-dataset, cross-topic and cross-lingual transfer learning in Subsection 3.1

More information about the general distribution of the words is available Table 1 and in the Appendix, Table 5.

2.2. X-stance: A Multilingual multi-target stance detection dataset

The X-stance (XS) dataset (Vamvas and Sennrich, 2020) contains 67,271 comments in French, German and Italian on more than 150 political issues (*targets*) retrieved from the Swiss application *Smartvote*. To tackle stance classification in this setting, the authors propose to integrate the target inside a natural question which can be seen as a debate’s title. This approach allows the model to learn across targets, to remain efficient in a zero-shot learning setting and to use the semantics information contained inside the pre-trained model (Yin et al., 2019). The 4 labels have been merged into 2 classes: *favor* and *against* the proposition, which can be seen as *yes* or *no* when the proposition is formulated as a question.

3. Experiments and Results

The 3 experiments below are complementary. The first experiment focuses on transfer learning across topics, targets and languages. The second one focuses on the interactive aspect of online debates. The last one highlights the value of the unlabeled DE dataset, with a self-training method handling unlabeled and imbalanced data.

3.1. Multilingual stance detection using transfer learning

It is known that when the source and target domains are dissimilar, standard transfer learning may fail and hurt the performance by conducting to a negative transfer (Rosenstein et al., 2005). Hence, showing the small DE dataset can improve the results on a bigger dataset via transfer learning across topics and language is a way to validate the annotations. The XS dataset, which is composed of multilingual comments answering to political debate questions from several topics, is the perfect candidate. We used a XLM-R (Conneau et al., 2020) as multilingual learning model, and call it XLM-R_{ft} when it has been already trained over one dataset.

3.2. Context-aware model

In order to model the dynamics aspect of a debate, we decided to use a context window to integrate an interactional context of variable size. We separated the different sentences using [SEP] tokens, rendering for a context window of size 2: [CLS] Debate Question [SEP] Sent n [SEP] Sent n-1 [SEP] Sent n-2 [SEP].

3.3. Data-augmentation with semi-supervised learning

As seen in Subsection 2.1, we annotated only a small part of the available DE dataset, leaving unlabeled a large amount of data that could potentially be useful to increase model performance. To maximise the potential of this unlabeled dataset, we propose to use a self-training method (Yarowsky, 1995). The general principle we follow is to leverage some of the model’s own prediction on unlabeled data by adding pseudo-examples in the training set. We compare two classical methods, using a threshold on the model’s class probability and taking the k predictions with the highest probability (resp. *thresh* and *k-best* in Table 3). When doing so, we keep aware of the downside of self-training such as the fact that the model is not able to correct its own mistakes and that errors are amplified (Ruder, 2019). Thus, if the unlabeled dataset is imbalanced, the classifier bias will be amplified by the pseudo-labels and the class-imbalance issue will be aggravated (Wei et al., 2021).

To mitigate this risk, we propose to combine both techniques, by adding a definite and balanced number of k_{max} examples chosen randomly amongst those which have a probability above the threshold, at each iteration of the SSL algorithm. Our technique makes no assumption on the label distribution of the unlabeled dataset and can thus help to prevent an overflowing of the training set with pseudo-examples from outer domains.

3.4. Methodological protocol

We followed the protocol of (Barriere and Balahur, 2020; Barriere and Jacquet, 2021) for the transformers’ learning phase, already used in the past for multilingual sentiment analysis and text classification. The pre-trained models that we used were made available online using the transformers library (Wolf et al., 2019). We used the Adam algorithm (Kingma and Ba, 2014) with early stopping for the optimization of the training loss, using a learning rate of $2e^{-6}$ for the first training of the model on a stance task, and $5e^{-7}$ when fine-tuning on another dataset for the transfer learning.

Unsupervised Method	Threshold	k_{max}	Balanced	Model	Prec.	Rec.	F1	Acc
\times	\times	\times	\times	XLM-R	68.6	69.3	68.9	70.1
				XLM-R _{ft}	70.7	69.9	70.2	72.1
thresh-0.99	0.99	\times	\times	XLM-R	68.6	69.8	69.1	70.7
				XLM-R _{ft}	68.9	69.6	69.0	70.9
k-best-2000	\times	2000	\times	XLM-R	67.5	68.3	67.8	69.3
				XLM-R _{ft}	70.4	69.9	69.8	71.9
k-best-600	\times	600	\times	XLM-R	69.4	68.5	68.0	69.5
				XLM-R _{ft}	72.5	70.3	71.1	73.3
our-2000	0.99	2000	\checkmark	XLM-R	69.5	69.4	69.4	71.3
				XLM-R _{ft}	70.5	69.9	69.3	71.7
our-600	0.99	600	\checkmark	XLM-R	70.9	71.6	71.1	72.7
				XLM-R _{ft}	71.5	71.5	71.4	73.5

Table 3: Results over the Debating Europe dataset for a 3-class classification using SSL

Ctxt	Prec	Rec.	F1	Acc
0	70.7	69.9	70.2	72.1
1	72.1	70.5	71.2	72.7
2	70.7	69.8	70.2	72.7

Table 4: Results over DE for different context windows. All the models were pre-trained over XS (XLM-R_{ft})

In contrast to (Vamvas and Sennrich, 2020), we do not perform any hyperparameter optimization on dev and use a shorter maximum sequence length (128 vs 512) to speed up training and evaluation.

We divided the DE dataset into 3 train/validation/test sets in a stratified way with a ratio of 75/5/20. To compare results, we proceeded the same partition as (Vamvas and Sennrich, 2020) for the XS dataset. For the SSL, we stopped at 5 iterations, used 0.99 for probability threshold, and 600 and 2000 as maximum number of examples added at each iteration when applicable.

3.5. Results

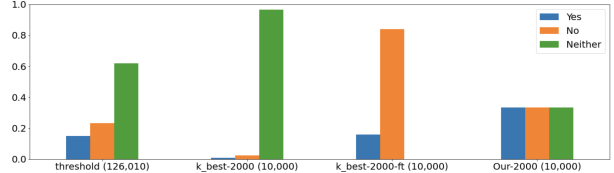
The 3 experiments are complementary. The first one gives an insight of the effect of a pre-training over a non-English multi-lingual dataset from another domain. The second one investigated the impact of the integration of the dialogic context inside the model, using context windows of variable sizes. The third experiment uses a self-training method applicable on a dataset of unlabeled and imbalanced data.

Cross-datasets This experiment gives an insight of the effect of a pre-training over a non-English multi-lingual dataset from another domain. As can be seen in Table 3 and 2, the transfer learning approach is efficient for both the datasets, even though they have different languages, topics and targets.

Impact of a context window This experiment investigated the impact of integrating dialogic context of variable size inside the model, using a context window. The results (Table 4) show that a context window can enhance the model and a context window of size 1 is optimal.

ST setting The results in Table 2 show that the ST setups were not all successful. To understand the causes

Figure 2: Distribution of the pseudo-labels



of this failure, Figure 2 shows the distribution (and amounts) of the pseudo-labels. Analysing the distribution, we can clearly observe the weaknesses of each method and draw a conclusion on why our method is working: it does not flood the gold labels with weak labels as pair with a balanced distribution.

The threshold method does not improve the performances of the model because of the small size of our dataset and the lack of model calibration. Too many pseudo-examples added at each iteration significantly degrade the performances of the model. The k-best method allows diminishing the number of examples added at every iteration and it performs well for the XLM-R_{ft}, as it has seen way more training examples and seems more robust.

4. Conclusion and Future work

In this work, we presented “*Debating Europe*” - a new dataset for stance detection and classification, composed of online debates and partly annotated for stance at the comment-level. This is as far as we know the first multi-target stance dataset in the literature. Although it has been annotated by one unique expert, we validated the quality of the annotation by showing the DE dataset is useful for transfer learning across languages and domains, and reaching a new state-of-the-art on the multi-lingual multi-target X-stance dataset. Additionally, we proposed and validated two methods to improve over the baseline results by integrating the interactional context inside a transformer models, and by utilising the imbalanced and unlabeled dataset with a home-made self-training algorithm that makes no assumption on the label distribution. The dataset and labels will be available online after publication. Future work includes extending DE dataset to further languages and domains, as well as testing the impact of annotation granularity.

5. Bibliographical References

- Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., and King, J. (2011). How can you say such things?!?: recognizing disagreement in informal political argument. *Proceedings of the Workshop on Languages in Social Media*, pages 2–11.
- Allaway, E. and McKeown, K. (2020). Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations.
- Augenstein, I., Rocktäschel, T., Vlachos, A., and Bontcheva, K. (2016). Stance detection with bidirectional conditional encoding. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 876–885.
- Barriere, V. and Balahur, A. (2020). Improving Sentiment Analysis over non-English Tweets using Multilingual Transformers and Automatic Translation for Data-Augmentation. In *COLING*.
- Barriere, V. and Jacquet, G. (2021). How does a pre-trained transformer integrate contextual keywords? Application to humanitarian computing. *Proceedings of the International ISCRAM Conference, 2021-May(May):766–771*.
- Barriere, V., Clavel, C., and Essid, S. (2018). Attitude Classification in Adjacency Pairs of a Human-Agent Interaction with Hidden Conditional Random Fields. In *ICASSP*.
- Barriere, V. (2017). Hybrid Models for Opinion Analysis in Speech Interactions. In *ICMI*, pages 647–651.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised Cross-Lingual Representation Learning at Scale. pages 31–38.
- Devlin, J., Chang, M.-w., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- Glandt, K., Khanal, S., Li, Y., Caragea, D., and Caragea, C. (2021). Stance Detection in COVID-19 Tweets. In *ACL-IJCNLP*, pages 1596–1611.
- Gorrell, G., Bontcheva, K., Derczynski, L., Kochkina, E., Liakata, M., and Zubiaga, A. (2019). RumourEval 2019: Determining rumour veracity and support for rumours. In *SemEval 2019*, pages 845–854.
- Hosseinia, M., Dragut, E., and Mukherjee, A. (2020). Stance Prediction for Contemporary Issues: Data and Experiments.
- Kingma, D. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, pages 1–13.
- Klie, J.-C., Bugert, M., Boullosa, B., de Castilho, R. E., and Gurevych, I. (2018). The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. *Proceedings of the International Conference on Computational Linguistics*, pages 5–9.
- Küçük, D. and Fazli, C. A. (2020). Stance detection: A survey. *ACM Computing Surveys*, 53(1).
- Lai, M., Cignarella, A. T., Hernández Farías, D. I., Bosco, C., Patti, V., and Rosso, P. (2020). Multilingual stance detection in social media political debates. *Computer Speech and Language*, 63.
- Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). A Dataset for Detecting Stance in Tweets.
- Prakash, A. and Madabushi, H. T. (2020). Incorporating Count-Based Features into Pre-Trained Models for Improved Stance Detection. *COLING*, pages 22–32.
- Rosenstein, M. T., Marx, Z., Kaelbling, L. P., and Dietterich, T. G. (2005). To transfer or not to transfer. *NIPS 2005 workshop on transfer learning*, 898:3.
- Ruder, S. (2019). Neural Transfer Learning for Natural Language Processing.
- Somasundaran, S. and Wiebe, J. (2009). Recognizing stances in online debates. *ACL-IJCNLP 2009*, pages 226–234.
- Somasundaran, S. and Wiebe, J. (2010). Recognizing Stances in Ideological On-Line Debates. In *NAACL Workshop*.
- Sridhar, D., Foulds, J., Huang, B., Getoor, L., and Walker, M. (2015). Joint Models of Disagreement and Stance in Online Debate. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 116–125.
- Vamvas, J. and Sennrich, R. (2020). X-stance: A Multilingual Multi-Target Dataset for Stance Detection. In *SwissText*.
- Walker, M. A., Anand, P., Abbott, R., and Grant, R. (2012). Stance classification using dialogic properties of persuasion. *NAACL HLT 2012 - 2012 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 592–596.
- Wei, C., Sohn, K., Mellina, C., Yuille, A., and Yang, F. (2021). CReST : A Class-Rebalancing Self-Training Framework for Imbalanced Semi-Supervised Learning. In *CVPR*.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., and Brew, J. (2019). HuggingFace’s Transformers: State-of-the-art Natural Language Processing.
- Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *ACL*, pages 189–196.
- Yin, W., Hay, J., and Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *EMNLP-IJCNLP 2019*, pages 3914–3923.
- Yu, J., Jiang, J., Khoo, L. M. S., Chieu, H. L., and Xia, R. (2020). Coupled Hierarchical Transformer for Stance-Aware Rumor Verification in Social Media Conversations. In *EMNLP*, pages 1392–1401.

Aggregation-level		Debate			Comment			All
Units	Label	μ	σ	med	μ	σ	med	Σ
Comments	All	140	99	101	1	0	1	2,523
	Yes	56	37	39	1	0	1	1,012
	No	29	39	14	1	0	1	489
	Neutral	18	18	11	1	0	1	282
	Not answering	41	23	35	1	0	1	740
Words	All	4,683	2,721	3,794	33	60	16	84,289
	Yes	1,933	1,221	1,772	34	74	13	34,790
	No	942	1,157	554	33	43	19	16,012
	Neutral	814	808	478	46	73	23	13,023
	Not answering	1,137	627	972	28	39	16	20,464

Table 5: Low-level statistics on the Debating Europe dataset. Here, μ represents the average mean, σ the standard deviation, med the median and Σ the sum.

Appendix

A. European Green Deal

We chose to select the debates that were falling under the scope of the European Green Deal European Commission’s priority.

The policy areas comprised in the European Green Deal are 9 and are the following: Biodiversity, From Farm to Fork, Sustainable agriculture, Clean Energy, Sustainable industry, Building and renovating, Sustainable mobility, Eliminating pollution and Climate action. More details are available online⁴

B. Questions of the annotated debates

The debates chosen for the annotation are the ones below: *Should we consume less energy?*, *Should we make the cities greener?*, *Can renewables ever replace fossil fuels 100%*, *Should we invest more in clean energies to avoid an energy crisis?*, *Should we cut CO2 emission and invest into clean energies?*, *Should we think about the real cost of the food we eat?*, *Should all cars be electric by 2025?*, *Does organic food really make a difference?*, *Should Europeans be encouraged to eat more sustainably?*, *Sustainable agriculture: With or without pesticides?*, *Should all EU countries abandon nuclear power?*, *Should we stop flying to help the environment?*, *Should plastic packaging be banned?*, *Should we all eat less meat?*, *Should we invest in cheap and clean energies?*, *Should we move towards a low-carbon economy or invest into clean energies?*, *Should the European Union ban plastic bags?* and *Should plastic water bottles be banned?*.

C. Debating Europe Dataset Statistics

More low-level statistics on the Debating Europe dataset are available in Table 5.

⁴https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en