

CoFE: A New Dataset of Intra-Multilingual Multi-target Stance Classification from an Online European Participatory Democracy Platform

Valentin Barriere, Guillaume Jacquet, Léo Hemamou

CENIA, European Commission's Publication Office, Sanofi R&D

AAACL-IJCNLP 22

What is CoFE? TLDR

- We propose a dataset of proposals and comments from the **CoFE** Participatory Democracy platform composed of 4k proposals and 20k comments
- **Intra-multilingual**: Automatic translation helps users to interact in their native languages
- **Multi-target**: 4k targets formulated in natural language
- It contains 7k comments self-annotated in stance by the users + 1.2k externally annotated comments in stance (3 classes)



Conference
on the **Future**
of **Europe**

What is CoFE? TLDR

- We propose a dataset of proposals and comments from the **CoFE** Participatory Democracy platform composed of 4k proposals and 20k comments
- **Intra-multilingual**: Automatic translation helps users to interact in their native languages
- **Multi-target**: 4k targets formulated in natural language
- It contains 7k comments self-annotated in stance by the users + 1.2k externally annotated comments in stance (3 classes)
- We propose a shared task @ CLEF 2023



Conference
on the **Future**
of **Europe**

- Participatory Democracy



- Participatory Democracy
- Large scale implies automation



- Participatory Democracy
- Large scale implies automation
- Europe has high diversity of languages, topics, propositions and arguments



What does a proposal look like?

Reforestation Europe



Eduardo Rafael Moreira Dos Santos

21/04/2021 00:31

21/04/2021 00:35 26 comments

Europe is becoming increasingly arid if it is not stopped, it will have disastrous consequences that could lead to the desertification of several southern countries. Planting trees is imperative and we must start this battle, as if a pandemic were being dealt with. Our planet is getting sick and climate changes are killing thousands of people every year. Planting millions of trees a year is possible. We must start now as fast as we can. Recover the Atlantic Mediterranean fauna and flora is one of the most urgent targets to achieve in next 2 years. 300 million trees across Europe , with special care in Portugal,Spain,Italy and Greece. The ecosystem will improve drastically ins a short period of time and the weather will get cooler. To improve our climate changes , we must plant trees, without them our world cannot have a life. We need urgent with effort from our world leaders to face this subject as the most urgent matter. The changes we all gonna achieved are so massive that we will see our planet healing and breath again. Everything is in our hands.We need to fight and save our planet 🌍

Please, let's make our old continent example and enrich our nations full of biodiversity.

168 Endorse

1025 | Follow

Reference: cofe-PROP-2021-04-834

Version number 2 (of 2) [see other versions](#)

[Share](#)

[Embed](#)

What does a proposal look like?

Reforestation Europe



Eduardo Rafael Moreira Dos Santos

21/04/2021 00:31

21/04/2021 00:35 26 comments

Europe is becoming increasingly arid if it is not stopped, it will have disastrous consequences that could lead to the desertification of several southern countries. Planting trees is imperative and we must start this battle, as if a pandemic were being dealt with. Our planet is getting sick and climate changes are killing thousands of people every year. Planting millions of trees a year is possible. We must start now as fast as we can. Recover the Atlantic Mediterranean fauna and flora is one of the most urgent targets to achieve in next 2 years. 300 million trees across Europe, with special care in Portugal, Spain, Italy and Greece. The ecosystem will improve drastically in a short period of time and the weather will get cooler. To improve our climate changes, we must plant trees, without them our world cannot have a life. We need urgent with effort from our world leaders to face this subject as the most urgent matter. The changes we all gonna achieved are so massive that we will see our planet healing and breath again. Everything is in our hands. We need to fight and save our planet 🌍

Please, let's make our old continent example and enrich our nations full of biodiversity.

168 Endorse

1025 | Follow

Reference: cofe-PROP-2021-04-834

Version number 2 (of 2) [see other versions](#)

[Share](#)

[Embed](#)

What does a proposal look like?

Reforesting Europe



Eduardo Rafael Moreira Dos Santos

21/04/2021 00:31

21/04/2021 00:35 26 comments

Europe is becoming increasingly arid if it is not stopped, it will have disastrous consequences that could lead to the desertification of several southern countries. Planting trees is imperative and we must start this battle, as if a pandemic were being dealt with. Our planet is getting sick and climate changes are killing thousands of people every year. Planting millions of trees a year is possible. We must start now as fast as we can. Recover the Atlantic Mediterranean fauna and flora is one of the most urgent targets to achieve in next 2 years. 300 million trees across Europe, with special care in Portugal, Spain, Italy and Greece. The ecosystem will improve drastically in a short period of time and the weather will get cooler. To improve our climate changes, we must plant trees, without them our world cannot have a life. We need urgent with effort from our world leaders to face this subject as the most urgent matter. The changes we all gonna achieved are so massive that we will see our planet healing and breath again. Everything is in our hands. We need to fight and save our planet 🌍

Please, let's make our old continent example and enrich our nations full of biodiversity.

168 Endorse

1025 | Follow

Reference: cofe-PROP-2021-04-834

Version number 2 (of 2) [see other versions](#)

[Share](#)

[Embed](#)

What does a comment thread look like?



Bernd Erbert 21/04/2021 12:10



In favor

Ich finde es wichtig das Wälder, insbesondere Mischwälder, Aufgeforstet werden

▲ 11 ▼ 0



Jan Jelínek 24/04/2021 09:45



U nás je ale běžnou praxí, že než dojde k sázení musí vzniknout rozsáhlá mýtina za použití těžké techniky. Tedy nejprve se zničí podrost, hrabanka a podhoubí a teprve po té přijde někdo se sazenicí. Ta následně uschne nebojí sežere přemnožená hladová zvěř, kterou myslivci v průběhu roku dokrmují, aby měli co střílet.

▲ 0 ▼ 0

What does a comment thread look like?



Bernd Erbert 21/04/2021 12:10



In favor

Ich finde es wichtig das Wälder, insbesondere Mischwälder, Aufgeforstet werden

▲ 11 ▼ 0



Jan Jelínek 24/04/2021 09:45



U nás je ale běžnou praxí, že než dojde k sázení musí vzniknout rozsáhlá mýtina za použití těžké techniky. Tedy nejprve se zničí podrost, hrabanka a podhoubí a teprve po té přijde někdo se sazenicí. Ta následně uschne nebojí sežere přemnožená hladová zvěř, kterou myslivci v průběhu roku dokrmují, aby měli co střílet.

▲ 0 ▼ 0

What does a comment thread look like?



Bernd Ebert

21/04/2021 12:10



In favor

Ich finde es wichtig das Wälder, insbesondere Mischwälder, Aufgeforstet werden

▲ 11 ▼ 0



Jan Jelínek

24/04/2021 09:45



U nás je ale běžnou praxí, že než dojde k sázení musí vzniknout rozsáhlá mýtina za použití těžké techniky. Tedy nejprve se zničí podrost, hrabanka a podhoubí a teprve po té přijde někdo se sazenicí. Ta následně uschne nebojí sežere přemnožená hladová zvěř, kterou myslivci v průběhu roku dokrmují, aby měli co střílet.

▲ 0 ▼ 0

Similar datasets

Datasets where the target is reformulated in a **natural question** in order to easily train one multilingual multi-target model on the entire dataset:

XStance [4]

Answers over one of 150 political issues in 3 languages by Swiss citizen

Procon [2]

Answers over one of 419 controversial issues in English from an online website

Debating Europe [1]

Comments in a debate over one of 18 hot topics in 2 languages by European citizen

- We introduce the concept of **intra-multilinguality**: comments and the propositions in the same discussion can be written in different languages
- **High variability** in targets
- Restricting a dataset to one language could induce nationality or cultural bias

⇒ Lack of an appropriate intra-multilingual multi-target stance-annotated debate dataset

Some statistics: 10 Designed Topics



Climate change and the environment



Health



A stronger economy, social justice and jobs



EU in the world



Other ideas



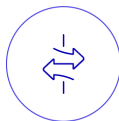
Values and rights, rule of law, security



Digital transformation



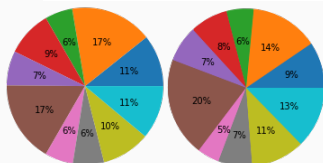
European democracy



Migration

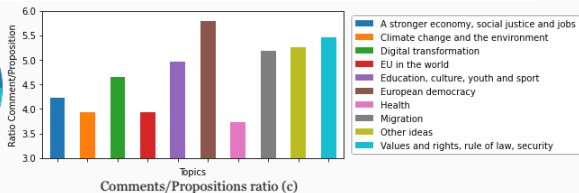


Education, culture, youth and sport



Propositions (a)

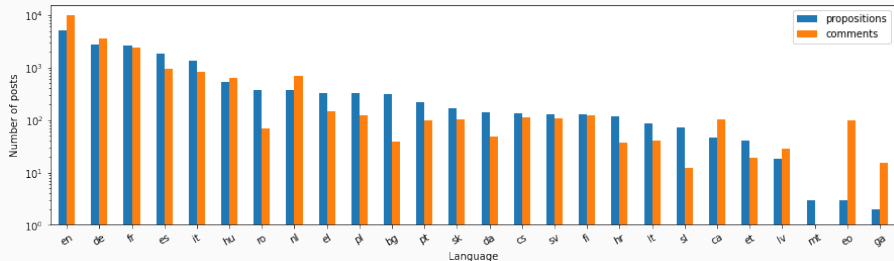
Comments (b)



Comments/Propositions ratio (c)

Some statistics: Variety in languages

26 languages: EU 24 languages + Esperanto and Catalan



- Very unbalanced distribution: English, German, French, Spanish and Italian are the main languages
- **Automatic translation** of every comments in any of the EU 24 languages allows the user to interact in different languages inside the same thread

Examples






Title	Topic	Proposal	Comment	Stance	url
Focus on Anti-Aging and Longevity research	Health	The EU has presented their green paper on ageing, and correctly named the aging...	The idea of prevention being better than a cure is nothing new or revolutionary. Rejuvenation...	Pro	
Set up a program for returnable food packaging...	Climate change and the environment	The European Union could set up a program for returnable food packaging made from...	Bringing our own packaging to stores could also be a very good option. People would be...	Pro	
Impose an IQ or arithmetic-logic test to immigrants	Migration	We should impose an IQ test or at least several cognitive tests making sure immigrants have...	On ne peut pas trier les migrants par un simple score sur les capacités cognitives. Certains fuient la guerre et vous...	Against	
Un Président de la Commission directement élu...	European democracy	Les élections, qu'elles soient présidentielles ou législatives, sont au coeur du processus...	I prefer sticking with a representative system and have the President of the...	Against	
Europa sí, pero no así	Values and rights, rule of law, security	En los últimos años, las naciones que forman parte de la UE han visto como su soberanía ha sido...	Zdecydowanie nie zgadzam się z pomysłem, aby interesy indywidualnych Państw miały...	Against	

Table 1: Examples of comments and proposals with the associated stance

3 Subdatasets

CF_S: Self-annotations

A subparts of the comments have been self-annotated by the commenter him/herself. Those annotations are binary: *In favor* or *Against*.

CF_U: Unlabeled

The subpart of the comments that does not have any annotation

CF_E: Externally annotated

We annotated 1200+ comments in a 3-class fashion

- We used the Inception annotation platform [3]
- 6 morphologically different languages: French, German, English, Greek, Italian and Hungarian
- We obtained a Krippendorff α of 0.69

3 Subdatasets

Dataset	XStance	DE	CF _S	CF _E	CF _U
Classes	2	3	2	3	∅
Languages	3	2	25	22	26
Targets	150	18	2,724	757	4,274
Comments	67,271	2,523	6,985	1,206	12,024
Debate	✗	✓	✓	✓	✓
Intra Mult.	✗	✗	✓	✓	✓

Table 2: Comparison with other annotated datasets

Baseline Experiments

We calculate some baselines over the **tri-class CF_E dataset**, using the similar available annotated resources: **X-Stance** and **CF_S** that are **binary annotated** and **Debating Europe** that is **ternary annotated**.

Model	DE	XS	CF_S	-	~	+	Acc.	m-F1
All - 1 training	✓	✓	✓	59.7	00.7	79.5	65.5	46.6
Cross-datasets	✓	✓	✗	54.3	30.5	73.9	59.6	52.9
Cross-debates	✓	✓	~	55.3	40.4	76.6	63.2	57.4
All - 2 trainings	✓	✓	✓	55.4	44.6	77.3	64.3	59.1

Table 3: F1, macro-F1 and Accuracy of the different baselines over the externally annotated dataset CF_E

Pre-training over the dataset that only have 2 labels before fine-tuning over the dataset that has 3 labels helps to obtain the best results.

- The CF_S and CF_U datasets will be available on demand
- CF_E will be used as a test set of one of the Touché Lab @ CLEF23: **Intra-Multilingual Multi-Target Stance Classification**



- CF_S will be used as data for the ValueEval task SemEval 2023: **Identification of Human Values behind Arguments**

SemEval

International Workshop on Semantic Evaluation

Conclusion

- New **stance dataset in online debate** on contemporary issues
- 20k **comments** and 4.2k **proposals**
- **Intra-multilingual** interactions between participants
- CF_S : 7k annotated comments in a **binary** way
- CF_E : 1.2k annotated comments in a **ternary** way
- **Baselines** proposed on the CF_E dataset without training over it
- Participate to our **shared task @ CLEF 2023** here

Thanks for listening!

Questions?



V. Barriere, A. Balahur, and B. Ravenet.

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

In *Proceedings of the First Workshop on Natural Language Processing for Political Sciences (PoliticalNLP)*, LREC, number June, pages 16–21, Marseille, France, 2022. European Language Resources Association.



M. Hosseinia, E. Dragut, and A. Mukherjee.

Stance Prediction for Contemporary Issues: Data and Experiments.

2020.



J.-C. Klie, M. Bugert, B. Boullosa, R. E. de Castilho, and I. Gurevych.

The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation.

Proceedings of the International Conference on Computational Linguistics, pages 5–9, 2018.



J. Vamvas and R. Sennrich.

X-stance: A Multilingual Multi-Target Dataset for Stance Detection.

In *SwissText*, 2020.