



LREC 2022 Workshop  
Language Resources and Evaluation Conference  
20-25 June 2022

**First Workshop on Natural Language Processing  
for Political Sciences  
(PoliticalNLP)**

**PROCEEDINGS**

Editors:

Haithem Afli (General Chair), Mehwish Alam, Houda Bouamor Cristina Blasi Casagran, Colleen Boland and Sahar Ghannay.

# Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences (PoliticalNLP)

Edited by:

Haithem Afli (General Chair), Mehwish Alam, Houda Bouamor, Cristina Blasi Casagran,  
Colleen Boland and Sahar Ghannay.

**ISBN: 979-10-95546-88-7**

**EAN: 9791095546912**

**For more information:**

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: [lrec@elda.org](mailto:lrec@elda.org)



© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License

## **Message from the General Chair**

The First Workshop on Natural Language Processing for Political Sciences (PoliticalNLP) took place on Friday, June 24, 2022, in Marseille, France, immediately after the 13th Edition of the Language Resources and Evaluation Conference (LREC 2022). The focus of our workshop was to explore the multifarious aspects of effective Natural Language Processing (NLP) techniques for socio-political data. The workshop provided a research platform dedicated to new methods and techniques for text processing of socio-political content and exploring the use of such methods in information extraction and analysis.

It was a venue for discussing the implementation of language technologies in the social and political sciences domain. Computational Social and Political scientists reported and discussed their NLP tools in comparison to their traditional coding approaches. Computational linguistics and machine learning practitioners and researchers investigated the challenges of real-world use cases in these domains.

Cristina Blasi Casagran from Universidad Autónoma de Barcelona gave the invited talk on "The Role of Emerging Predictive IT Tools in Effective Migration Governance".

We would like to thank the members of the Program Committee for their timely reviews. We also would like to thank the participants of the workshop for the interesting discussions around our PoliticalNLP topics, and our sponsors: the Horizon 2020 project ITFLOWS (Grant Agreement No. 882986) and the Science Foundation Ireland Research Centre ADAPT (Grant Agreement No.13/RC/2106).

Haithem Afli



## **Organizers**

Haithem Afli – ADAPT Centre, Munster Technological University (Ireland)  
Mehwish Alam – FIZ Karlsruhe - Leibniz Institute for Information Infrastructure (Germany)  
Colleen Boland – Universidad Autónoma de Barcelona (Spain)  
Houda Bouamor – Carnegie Mellon University (Qatar)  
Sahar Ghannay – Université Paris-Saclay, CNRS, LISN (France)  
Cristina Blasi Casagran – Universidad Autónoma de Barcelona (Spain)

## **Program Committee:**

Yiyi Chen, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Germany  
Adam Zebrowski, Microsoft, Saudi Arabia  
Bruno Andrade, Munster Technological University, Ireland  
Lenka Dražanová, European University Institute, Italy  
Georgios Stavropoulos, The Centre for Research and Technology, Greece  
Mohammed Hasanuzzaman, Munster Technological University, Ireland  
Amira Barhoumi, LIUM, Le Mans Université, France  
Colleen Boland, Universidad Autónoma de Barcelona, Spain  
Andrea Iana, University of Mannheim, Germany  
Nikolaos Gkevrekis, CERTH, Greece  
Praveen Joshi, Munster Technological University, Ireland  
Patrick Paroubek, Université Paris-Saclay, CNRS, LISN, France  
Ilias Iliopoulos, CERTH, Greece  
Zsolt Kardkovacs, Munster Technological University, Ireland  
Pintu Lohar, Dublin City University, Ireland  
Suman Adhya, Indian Association for the Cultivation of Science, India  
Valentin Barriere, Joint Research Center, Italy



## Table of Contents

<i>NewYeS: A Corpus of New Year’s Speeches with a Comparative Analysis</i> Anna Tramarin and Carlo Strapparava .....	1
<i>Correlating Political Party Names in Tweets, Newspapers and Election Results</i> Eric Sanders and Antal van den Bosch .....	8
<i>Debating Europe: A Multilingual Multi-Target Stance Classification Dataset of Online Debates</i> Valentin Barriere, Alexandra Balahur and Brian Ravenet .....	16
<i>An Unsupervised Approach to Discover Media Frames</i> Sha Lai, Yanru Jiang, Lei Guo, Margrit Betke, Prakash Ishwar and Derry Tanti Wijaya .....	22
<i>Electoral Agitation Dataset: The Use Case of the Polish Election</i> Mateusz Baran, Mateusz Andrzej Wójcik, Piotr Kolebski, Michał Bernaczyk, Krzysztof Rajda, Lukasz Augustyniak and Tomasz Kajdanowicz .....	32
<i>Enhancing Geocoding of Adjectival Toponyms With Heuristics</i> Breno Alef Dourado Sá, Ticiana Coelho da Silva and Jose Antonio Fernandes de Macedo .....	37
<i>Cause and Effect in Governmental Reports: Two Data Sets for Causality Detection in Swedish</i> Luise Dürlich, Sebastian Reimann, gustav finnveden, Joakim Nivre and Sara Stymne .....	46
<i>Does Twitter know your political views? POLiTweets dataset and semi-automatic method for political leaning discovery</i> Joanna Baran, Michał Kajstura, Maciej Ziolkowski and Krzysztof Rajda .....	56
<i>Political Communities on Twitter: Case Study of the 2022 French Presidential Election</i> Hadi Abdine, Yanzhu Guo, Virgile E. Rennard and Michalis Vazirgiannis .....	62
<i>What Does the Indian Parliament Discuss? An Exploratory Analysis of the Question Hour in the Lok Sabha</i> Suman Adhya and Debarshi Kumar Sanyal .....	72
<i>Don’t Burst Blindly: For a Better Use of Natural Language Processing to Fight Opinion Bubbles in News Recommendations</i> Evan Dufraisse, Céline Treuillier, Armelle Brun, Julien Tourille, Sylvain Castagnos and Adrian Popescu .....	79
<i>Creation of Polish Online News Corpus for Political Polarization Studies</i> Joanna E. Szwoch, Mateusz Staszko, Rafal Rzepka and Kenji Araki .....	86
<i>Annotation of expressive dimensions on a multimodal French corpus of political interviews</i> Jules Cauzinille, Marc Evrard, Nikita Kiselov and Albert Rilliard .....	91
<i>TransCasm: A Bilingual Corpus of Sarcastic Tweets</i> Desline Simon, Sheila Castilho, Pintu Lohar and Haithem Afli .....	98

# Workshop Program

**Friday, June 24, 2022**

**09:00–09:10** Opening and Welcome by Haithem Afi

**09:10–09:45** Keynote Speech by Cristina Blasi Casagran: The Role of Emerging Predictive IT Tools in Effective Migration Governance

**09:45–10:45** Oral session 1: chaired by Houda Bouamor

09:45–10:05 Debating Europe: A Multilingual Multi-Target Stance Classification Dataset of On-line Debates  
Valentin Barriere, Alexandra Balahur, and Brian Ravenet

10:05–10:25 Cause and Effect in Governmental Reports: Two Data Sets for Causality Detection in Swedish  
Luise Dürlich, Sebastian Reimann, gustav finnveden, Joakim Nivre, and Sara Stymne

10:25–10:45 An Unsupervised Approach to Discover Media Frames  
Sha Lai, Yanru Jiang, Lei Guo, Margrit Betke, Prakash Ishwar, and Derry Tanti Wijaya

**10:45–11:45** Poster session: chaired by Mehwish Alam

**11:45–12:45** Oral session 2: chaired by Sahar Ghannay

11:45–12:05 Electoral Agitation Dataset: The Use Case of the Polish Election  
Mateusz Baran, Mateusz Andrzej Wójcik, Piotr Kolebski, Michał Bernaczyk, Krzysztof Rajda, Lukasz Augustyniak and Tomasz Kajdanowicz

12:05–12:25 Does Twitter know your political views? POLiTweets dataset and semi-automatic method for political leaning discovery  
Joanna Baran, Michał Kajstura, Maciej Ziolkowski and Krzysztof Rajda

12:25–12:45 Political Communities on Twitter: Case Study of the 2022 French Presidential Election  
Hadi Abdine, Yanzhu Guo, Virgile E. Rennard and Michalis Vazirgiannis

**12:45–13:00** Closing remarks by Haithem Afi