



US Centre Summer Research Grant

Recipient name: Julia Leschke

Project title: Measuring Populism in Multi-Lingual Political Texts Using Hand-Coding and Machine Learning

Summary of project:

Globalization and de-industrialization have fundamentally restructured the societies of the US and Western Europe in two ways. Global market competition has rearranged the industrial and manufacturing sector by laying-off entire towns, deteriorating the working conditions of blue-collar workers, while creating a rapidly expanding precarious segment of gig economy workers. Simultaneously, very few globally acting players have increased their wealth significantly, leaving us with inequality levels of the 1920's. At the same time, globalization has also increased immigration into industrialized countries and given rise to culturally and socially liberal values competing with traditional norms. These economic, social and cultural tensions have polarized the electorate, so that parties such as Democrats and Republicans, have to cater to a very diverse and wide set of demands to win over voters. To succeed in elections, parties and politicians adapt different strategies of framing the current divisions in society – reaching from the extremist right-wing illiberal populism of Trump, over the socially liberal and unifying appeals of Biden, to the left-wing populist rhetoric of Sanders or Ocasio-Cortez. *But how exactly do political parties and actors in the US and Western Europe persuade voters to support their pluralist, radical left- or right-wing worldviews in times of entrenched political polarization? And who are these voters which find extremist, illiberal or anti-establishment appeals so enticing?* My PhD seeks to answer these questions by creating and analyzing an unprecedentedly rich and fine-grained dataset of political communication of more than 6.2 million speeches from the US Congress and multiple West European parliaments, along with almost 900 election manifestos covering the last 60 years. To measure different political positions, e.g. anti-elitist or pluralist appeals, I combine a well-tested coding scheme used to hand-label 150,000 text segments with state-of-the-art machine learning to automatically classify the discourse in millions of speeches. The US Centre Grant would enable me to complete this project by covering the full cost of the hand-labelling of the remaining set of 95,000 sentences in collaboration with native speaking research assistants. Linked with election data and voter surveys, this dataset would allow me to explain which exact segments of the US and West European electorates are most likely to endorse populist or pluralist actors from the left or right. This extremely rich data source and the resulting publications would significantly contribute to our knowledge of the motives of extremist, radical but also pluralist voters in a polarized and globalized Western world.

INTRODUCTION

Globalization and de-industrialization have fundamentally restructured the societies of the US and Western Europe in two ways. Global market competition has rearranged the industrial and manufacturing sector by laying-off entire towns, deteriorating the working conditions of blue-collar workers, while creating a rapidly expanding precarious segment of gig economy workers. Simultaneously, very few globally acting players have increased their wealth significantly, leaving us with inequality levels of the 1920's. At the same time, globalization has also increased immigration into industrialized countries and given rise to culturally and socially liberal values competing with traditional norms. These economic, social and cultural tensions have polarized the electorate, so that parties such as Democrats and Republicans, have to cater to a very diverse and wide set of demands to win over voters. To succeed in elections, parties and politicians adapt different strategies of framing the current divisions in society – reaching from the extremist right-wing illiberal populism of Trump, over the socially liberal and unifying appeals of Biden, to the left-wing populist rhetoric of Sanders or Ocasio-Cortez.

But how exactly do political parties and actors in the US and Western Europe persuade voters to support their pluralist, radical left- or right-wing worldviews in times of entrenched political polarization? And who are these voters which find extremist, illiberal or anti-establishment appeals so enticing?

PROJECT SUMMARY

My PhD seeks to answer these questions by creating and analyzing an unprecedentedly rich and fine-grained dataset of political communication of more than 6.2 million speeches from the US Congress and multiple West European parliaments, along with almost 900 election manifestos covering the last 60 years. To measure different political positions, I combine a well-tested coding scheme used to hand-label 150,000 text segments with state-of-the-art machine learning to automatically classify the discourse in millions of speeches. More precisely, I am interested to measure five kinds of political statements in text. The first three concepts are the constituent elements of classic populist discourse. The first one is anti-elitism, a position which attacks and accuses the political, cultural, and intellectual elite. Simultaneously, populists appeal to 'the people', which they portray as a unified actor, which is allegedly robbed of their self-autonomy and exploited by 'the elite'. The elite is portrayed as morally corrupt and inferior, whilst the group of 'the people' is portrayed as morally superior and virtuous. The third concept pertaining to populism is 'the will of the people'. Statements invoking 'the people's will' claim that the power should be returned to 'the people' and that politics should always follow the majority's will – even if constitutional principles are violated. The final two concepts whose salience I measure in text, represent the antithesis to populist discourse – pluralism and compromise-orientation. Pluralist statements contain the idea that societies consist of heterogeneous groups, who have different creeds, worldviews, and values. Instead of dividing these groups into normatively 'good' or 'bad' camps as is done in populist discourses, pluralism values the difference of worldviews and explicitly refrains from attaching a normative hierarchy to different lifestyles or values. Compromise-orientation as a concept demands that we need to forge compromises in political decision-making, which integrate minority interests and ideas. As such, pluralism and compromise-orientation constitute anti-populist positions. The coding scheme and coding manual was pre-tested and trialed multiple times before the start of the project with the help of several different native speaking coders and statistical tests. To make labelling, i.e. classifying single sentences taken from political party manifestos and political speeches, as reliable and valid as possible, I ask native speakers to code each of the five concepts independently on a large selection of carefully stratified samples of sentences.

Prior to the final project phase, I managed to collect approximately 150,000 text labels via the same method for a range of Central and Northern European countries. The final project phase was made possible through the generous US Phelan Centre grant, which gave me the opportunity to complete my ambitious project by covering

the full cost of the hand-labelling of the remaining set of 95,000 sentences in collaboration with native speaking research assistants during the past summer months.

SUMMARY OF DATA COLLECTION

DATE	PROJECT SCHEDULE AND TASKS
21 June – 30 June 2021	Recruitment and interviews of native speaking research assistants (one US American native speaker, one UK-English native speaker, one Austrian-German native speaker) to assist me with labelling extracts from political speeches. Online recruitment was done over mailing lists of Social Science Departments of European University mailing lists. All research assistants were students of a social science discipline.
1 July – 6 July 2021	Service contracts were issued to all native speakers.
8 July 2021	Online introductory session for all three coders on zoom was held, introducing each other, the project schedule, the project motivation, outcome, description of tasks and work schedule. Free online platform system (MEGA) is introduced which I designed to facilitate the sharing of raw and finished sheets of text, video material, code manuals and online chats.
9 July 2021	All three coders start with training session 1 of 5. Training is delivered via an online-training video (1 hour) produced by me to make sure each coder receives the exact training. Each training session is followed by an online coding test to validate the coding ability of the coders.
10 -12 July 2021	Results of the test coding, any problems and questions are discussed on zoom with all three coders.
13 -23 July 2021	First data collection phase starts on the concept of anti-elitism starts, one of the three central populist concepts. I regularly meet with each of the coders individually and as a group online to see how they are progressing; they can ask questions and flag problems. Meanwhile, I prepare stratified samples from each of the corpora for each country-context for the next category to code. First data collection phase ends.
26 July – 04 August 2021	Second data collection phase starts on the concept of pluralism, one of the two anti-populist concepts. Quality control of material produced so far. I prepare stratified samples from each of the corpora for each country-context for the next category to code. First data collection phase ends.
5 – 15 August 2021	Third data collection phase on the concept of compromise-orientation , the second concept to measure anti-populist positions.

	Quality control of material produced so far. I prepare stratified samples from each of the corpora for each country-context for the next category to code. First data collection phase ends.
16 – 26 August 2021	Fourth data collection phase on the concept of ‘the people’s will’, the second of the three constitutive concepts of populism. Quality control of material produced so far. I prepare stratified samples from each of the corpora for each country-context for the next category to code. First data collection phase ends.
27 August – 10 September 2021	Fifth data collection phase on the concept of ‘the people’, the third and final constitutive concept to populism. Quality control of material produced so far. I prepare stratified samples from each of the corpora for each country-context for the next category to code. First data collection phase ends.
11 – 18 September 2021	Final deadline for uploading all remaining coded data by coders. Buffer for possible extensions (used by two coders). Final quality control of coders’ material.
20 - 25 September 2021	Payment of all coders.

RESULTS AND CONCLUSION

Since the final data collection phase I have been able to descriptively plot some of the results and establish a data infrastructure to manage the codes and meta-information for each of the 95,000 sentences. The first descriptive results look promising, very varied, and interesting, and I will analyse them in the coming month for my dissertation. In hindsight, I am particularly happy about the fact, that the recruitment, online communication, and meetings with the three research assistants went so well, given that I only met all coders online and we all had to work remotely without ever meeting each other in real life. Overall, I am extremely happy with the course of the project, and I am grateful for and pleased about the native speakers’ dedicated work and their motivation throughout the training sessions and course of the project. Most of all, however, I am indebted to the generous US Phelan Centre grant, which made this final and crucial research phase possible. The Phelan US Centre grant allowed me to expand the data collection to a level at which I now have a sufficiently large dataset to test different machine learning methods for the automated classification of political speeches and political manifesto segments to then enlarge the dataset even more to gain a rich insight. Tracing populism and anti-populism across time, contexts and countries will allow us to identify specific conditions under which any of these polar political positions thrive and allow us to understand better the origins, accelerators, and consequences of populist or anti-populist rhetoric. Linked with election data and voter surveys, this completely novel, extremely rich, and fine-grained dataset will allow me to explain which exact segments of the US and West European electorates are most likely to endorse populist or pluralist actors from the left or right – and to do this in a variety of languages and country-contexts. I hope that this rich data source and the resulting publications will significantly contribute to our knowledge of the motives of extremist, radical but also pluralist voters in a polarized and globalized Western world.

I have chosen the option of using the entire sum of the US Phelan Centre Grant at once to make the data collection as time- and cost-efficient as possible. Recruiting, training, and supervising all three research assistant coders at the same time allowed me to discuss issues or questions which any one member of the group had, together with the others, which helped them learn from each other. At the same time, I had to only run the introductory session once, instead of three separate times. I could also be more efficient by occasionally holding supervision meetings with all three of them, instead of individually, and finishing the project within the summer months as opposed to stretching the project over autumn and perhaps even winter. At the same time meeting with all three research assistants at regular intervals, created much more of team spirit, which helped with all our work motivation and working morale. I got the feedback from all three coders that they enjoyed discussing problems with the other two coders and me, especially as they were all working remotely and online only.