Emotions, Populism and Social Media: A comparative analysis using Natural Language Processing (NLP) tools

The topic of emotions is not alien to political science. In international relations, scholars representing diverse theoretical approaches (e.g., Shimko 1992, Jervis 1976, Wendt 1992, Mercer 2010; Hutchison and Bleiker 2014; Hutchison 2016; Reus-Smit 2014; Koschut et al 2017) have argued that perceptions—presumably influenced by emotions—are a prominent factor for international relations for years. In comparative politics, as Bestvater and Monroe (2022, 2) put it "politics is often contentious and emotionally charged, so we frequently employ sentiment-laden language when we talk about it." The increased visibility of populism and the polarization of political debates worldwide have therefore contributed to the growing number of works on the role of sentiments, emotions, and stance detection across the field in recent years (Marcus 2000; Enyedi 2016; Moffitt 2016; Cox 2018; McCoy et al 2018; Abercrombie and Batista-Navarro 2020), some of which are interdisciplinary in nature. (Prior and van Hoef 2018)

Nevertheless, some important gaps in the literature still remain. For instance, although emotions are often linked with populism (Wahl-Jorgensen 2019; Clarke et al 2006; Hoggett and Thompson, 2012; Demertzis 2013; Rico et al 2017) the nature of their relationship is not fully explored. Similarly, few works exist at the intersection of politics and emotions in social media. (Sobaci and Karkin 2013; Hyvärinen and Beck 2014) Studies that use Natural Language Processing (NLP) methods and big data to analyze these topics are even fewer. (Widmann 2021) Rather than a lack of scholarly interest, this gap likely arises from several co-existing factors, including the overall ambiguity of emotions (Clarke et al 2006, 3) and the challenges to its observation or measurement (Clément and Sangar 2018). Taking emotions to the front seat also goes against the deep-rooted rationality assumption for actors in political science. (Lynggaard 2019, 1208) Measuring populism across time in comparative studies similarly remains a tricky subject for researchers. (Hawkins and Kaltwasser 2018; Hawkins et al 2019; Norris 2020; Jankowski and Huber 2022) Another complicating fact is that many of the available methods to extract and analyze big data on these topics originate from STEM fields, such as Computer Science, which have had limited contact with political science.

Yet, the intersection between emotions and populism in social media demands attention for several reasons. First, recent research suggests that emotions can significantly influence what is perceived by humans and their subsequent behavior. (Brosch et al 2013) As such, it invalidates any dismissive attitude towards them as "irrational" and thus, unworthy of reflection for political inferences. Second, both populism and social media are expected to continue to influence politics in the forthcoming years. (Levy 2018; Duncombe 2019; Wahl-Jorgensen 2019; Grossman et al 2020) In a world where news and social interactions increasingly take place online, including social media, it is hardly surprising. However, other factors also stand out. For instance, in many countries where traditional media (e.g., radio, television, newspapers) are heavily monitored by the political regime, social media have emerged as relatively free and prominent platforms to receive unfiltered news and reflect the political views, debates, and emotions of users. As a result, they have become quite popular: For instance, there are roughly 69 million social media users in Turkey since 2022. Turkey is also globally ranked at the top seventh place among Twitter users, which approximately translates into 16 million users. (Datareportal 2022) Although these figures are not identical to

individual users, and not every user is politically active or engaged online, they are significantly large numbers to warrant further analysis. Yet, it has become impossible to analyze the rapidly accumulating big data and assess their implications in politics without the aid of computational techniques that use artificial intelligence.

My research proposes to address this lacuna: It aims to comparatively analyze the emotions of populist and non-populist political parties in Turkey on social media (Twitter) and the parliamentary corpora online (extracted from the Turkish parliament's official website) using NLP (Natural Language Processing) techniques. It also aims to map the emotions of tweets posted in Turkish on Twitter on the political parties represented in the Turkish parliament. (Data extraction from Twitter is subject to restrictions, so the actual cutoff dates can vary) Once the data collection and organization are complete, the next step involves comparing them with similar data available for the relevant parliamentary and social media corpora of western democracies. (e.g., Abercrombie and Batista-Navarro 2019; Hawkins et al 2019; Card et al 2022)

Based on the available literature, this project is possibly the first one of its kind at the intersection of political science and computational methods that draws from both Turkish social media and parliamentary corpus: Although some works exist on the speeches by political leaders and MPs in the Turkish parliament (Güngör 2014; Devran and Özcan 2016; Elçi 2019), on party manifestoes (Çarkoğlu and Erol 2022) and on social media (Demirci 2014; Ayata et al 2017; Bulut and Yörük 2017; İmik Tanyıldızı and Ateş 2018; Tocoglu et al 2019; Şen and Altın 2020; Tunçer 2020; Baloğlu 2021), none of them focuses on all of these factors simultaneously. Furthermore, many of these works use small datasets, and rely almost exclusively on qualitative methods for their analyses. Similarly, those studies that use computational methods and utilize bigger datasets often adopt a narrow technical focus that are of little interest or direct relevance to an average political scientist working on these topics. Moreover, Turkey stands out as a compelling case to study the relationship between populism and emotions due to its past experiences with different regime types, populist political parties in and out of power and a lively social media.

In addition to these factors, I am also personally motivated to pursue this line of study due to my research background and academic past: As a political scientist, I have worked on democratization, political regimes, and the impact of perceptions on politics and security (on both domestic and international levels) for decades. Many of my works studied Turkey from a comparative perspective. Aside from my publications and research experience in international project networks and consortia on related topics, I have also taught extensively on these subjects and research methods for many years. In recent years, I have also taken online courses to learn computational methods to analyze big data on politics. Furthermore, I recently co-authored two cross-disciplinary works that focus on the parliamentary corpus and public stance in Turkish tweets towards various countries using NLP methods, which can be taken as precursors to this project.

The starting point of this research is similar to a recent study by Widmann (2021), which compares the emotionality of populist and non populist political parties in German speaking countries in Europe. In Widmann (2021), the goal is to find out whether there is a difference between these actors when it comes to emotional communication, or not. My project also starts off with a similar question: Do populist parties and politicians display more emotions

than their non-populist counterparts in general in Turkey? Related to this general research question are further tentative questions for reflection and analysis: i) Are populist party supporters more responsive to emotional language than non-populist party supporters in social media? ii) Does incumbency influence the overall display of emotions for political parties? iii) What is the impact of regime types (democratic/hybrid/non-democratic) on the variety and the intensity of emotions expressed in social media and the parliamentary corpus?

Notwithstanding its similar origins, my project also diverges from Widmann (2021) in the following aspects: First, while focusing on the emotions of populist parties, Widmann (2021) and other works on the parliamentary corpora (Abercrombie and Batista-Navarro 2019) do not explore the possible impact of the regime type (e.g., democratic, hybrid, non-democratic) as an independent variable. Yet, the political environment of different regime types can influence how emotions are expressed in parliamentary or social media settings. In democracies, individuals and non-governing political parties can express a wider range of emotions with diverse levels of intensity, as opposed to non-democratic regimes where public expressions are closely monitored and even censored. (Altindag et al 2021) However, it does not mean the absence of any public emotions in nondemocratic settings altogether, either: For instance, in authoritarian countries like China, the goal of the existing regime seems not to completely ban the expression of emotions in social media, but to ensure that they do not lead to any organized mass mobilization. (King et al 2013) Since the relevant literature on the topic almost exclusively discusses democracies and has little to say on nondemocracies (including hybrid regimes) any findings of this project can therefore significantly contribute to the existing debate in the literature.

A second missing element in Widmann (2021) and the rest of the literature is the incumbency effect on populism and emotions. Although a recent study (Dai and Kustov 2022) on U.S. politicians argues that incumbency does not influence the populist rhetoric, it does not offer any findings on emotions, either. Yet, the impact of incumbency can vary in different regime settings due to its potential distributional effects. Similarly, assuming that the ruling political parties in nondemocratic regimes would display little emotional variation at face value can be misleading: Instead of remaining constant, emotions utilized by incumbent populist parties in authoritarian countries can also fluctuate in time to suit their perceived needs. This project can help to understand the overall dynamics of this process by mapping the topography of the emotions of political elites, and the public represented in social media in Turkey.

While exploring these themes, my research aims to adopt a flexible approach towards the big data it collects and the related theoretical framework. This decision arises from the nature of working with text as data and big data, which often requires going beyond standard deductive methods when needed. (Hegelich 2016) Although classic structured research using deductive approach is seriously helpful to outline a clearly defined theoretical framework, it can also limit and mislead when the same framework was not informed by big data originally. (Grimmer et al 2022, 38-39)

However, adopting of a flexible theoretical framework does not mean discarding it altogether in favor of random data collection and displaying those findings as they are. Instead, it refers to the possibility of developing new theories while working with big data (Grimmer et al 2022, 41) Although predicting future events and behavior remain an important goal of this project, the immensity of accessible (and ever increasing) raw information and data also

prioritize the need to draw a theory informed initial descriptive map of the existing terrain and draw inferences from it where needed. (Grimmer et al 2022, 31)

My proposed project can also make a modest methodological contribution to the existing literature by employing NLP tools. Widmann (2021) focuses on German speaking countries and German texts and expands on a sentiment dictionary that was built earlier to track positive and negative sentiments in German politics. For Turkey, although there is the TREMO dataset for Turkish (Toçoğlu and Alpkoçak, 2018) for emotions, along with other datasets, they do not particularly target Turkish politics and cannot be used as is as a trainer for the political corpora targeted here. Similar to Widmann (2021), one of the goals of this project then is to tailor a dataset that is suitable to study Turkish political texts. The resulting dataset can then be used by other researchers using NLP to study Turkish politics.

Natural Language Processing (NLP) refers to a combination of computational techniques which enable machines to "read" texts, or "understand" pictures or sounds akin to human use of languages. When doing text analysis using NLP, words are treated as data. Various NLP techniques enable the machines to process the words in a text as a whole, instead of trying to understand their individual meanings. This process can utilize tools from a variety of fields, ranging from linguistics to statistics. Machines are then trained on corpora related to the researched topics. Since the computational techniques are mostly industry driven, the existing research and relevant corpora using NLP has primarily involved the online consumer behavior and marketing. However, due to the increasing impact of the internet on politics, the vast (and ever increasing) amounts of online text data and its misuse (e.g., the Cambridge Analytica incident in recent past), efforts to apply these techniques to political issues have increased.

NLP frameworks like BERT are crucial to any text analysis using big data, since they can be specifically tailored to the corpus that a researcher wants to analyze. Although commercial software packages (e.g., Nvivo, Maxqda or Atlas.ti) exist to conduct qualitative content analysis for texts, their analytical power can be limited, since they cannot be fully tailored to meet the demands of a specific corpus. Furthermore, they are not free, and their dependence on coder's input during the keyword selection phase of the research can increase the overall bias risk of any findings. (King et al 2017)

However, NLP has certain limits, too. For this project, using ML methods would not explain why people or political parties gravitate towards populist (or emotional) speeches in the first place, or what makes them succeed. (Dai and Kustov 2022, 398) The accuracy of NLP assisted analyses also need improvement. Another criticism involves distinguishing between stance detection and general sentiment of examined texts, as ignoring it can lead to misleading results. (Bestvater and Monroe 2022, 2) This criticism, however, needs further verification, as it neglects the potential impact of regime type behind this distinction. If the political environment where the examined corpora are subject to monitoring and restrictions (including auto censure), NLP findings on sentiments may be the closest or only available and consistent estimations of stance detection. Therefore, although text analysis can reveal significant amounts of information concerning social interactions, a deeper understanding of its relevance and meaning require some knowledge of the political background it generated from. (Grimmer et al 2022, 24)

There are several compelling reasons to conduct this project in the USA. First, computational social science methods, including the NLP techniques proposed to be used in this project are

almost unrepresented in Turkey for political science: Except for a few research centers located in private universities in Istanbul, no public universities or centers exist that focus on this field and conduct research using its methods to my knowledge. Although computer science departments do employ these methods, the topics of their research are frequently outside the existing theories, pressing issues and debates in political science.

Second, data extraction, cleaning, organization, and its overall quality in computational social sciences takes a considerable amount of time, expertise, developed laboratory settings and funding. Therefore, using readily available datasets in the leading US research centers on computational social sciences is crucial to conducting research that is also capable of keeping up to date with the current political trends and obtaining timely results. At the same time, for researchers who want to work on political issues using big data, remaining unaware of the latest technological developments and discussions in this fast-growing field is quite risky: Although I have taken some online courses on computational methods and conducted some research using them, I want to get a firmer grasp of the latest techniques and discussions that are related to this fast-growing field. Although an ever-growing number of online courses are available for NLP and ML, they also have some serious drawbacks: They are mostly developed by researchers with a computer science background, which means that they do not always address the background and needs of researchers in social sciences, including political science. Therefore, it makes the overall learning process more challenging and less fruitful. Moreover, gaining methodological proficiency in computational methods like NLP is not enough to conducting research on political science topics on its own: The optimal use of these methods also requires having a firm knowledge of the leading theories and getting actively engaged in the ongoing relevant discussions in political science. From this angle, the USA is the ideal country for me to pursue my research as it offers some of the top-ranking research centers at well-known universities that also lead the developments in computational social science methods. Moreover, the political science departments, and policy institutes of the same universities also lead some of the most prominent contemporary discussions in the topics that my research plans to focus on.

In addition to my personal career development, such an experience would be also crucial for my students in Turkey. Compared to STEM fields, social sciences in general and political science specifically lags in adopting new techniques for big data analysis. Political science scholars and students in developing countries like Turkey are under increased risk. Almost three decades ago, as a young student my Fulbright grant allowed me to pursue my graduate studies in one of the most outstanding U.S. universities in political science and pass all that knowledge and analytical skills onto my students in Turkey later. Receiving the Fulbright grant as a senior scholar will enable me once again to catch up with the latest techniques, debates, and developments relevant to computational social sciences, particularly NLP and political science in USA and enable me to transfer all that knowledge to my students in Turkey after my return.