### UNIVERSITÄT POTSDAM

MASTER'S THESIS

# Transformer-Based Analysis of Climate Discourse and Green Party Evolution in German Party Manifestos

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in

**Cognitive Systems** Faculty of Human Sciences

Submitted by: Juliane HANEL Supervisors: Prof. Dr. Manfred Stede Dr. Pola Lehmann

BERLIN, 02.12.2022

## Abstract

Climate change is arguably one of the largest existential threats to the future of human society. However, it is not only a physical phenomenon, but also poses substantial social, political, and communicative challenges. Mitigating climate change will require cooperation on every level of society, and gaps in communication preclude such unified, effective action. This thesis intersects computational linguistics and political science, investigating the evolution of German policymakers' public messaging on this issue. Over two experiments, the manifestos of three parties - the two mainstream parties, the CDU and SPD, as well as the Green Party, Die Grünen – are examined from 1980 to 2021, with a focus on the evolution of the Greens. A transformerbased dynamic topic modeling approach is implemented to model the parties' communication. The results show a convergence of Green and mainstream discourse over this period. A so-called *greening* of the discourse is observed, with climate-related issues becoming an integral part of all election programs by 2009. Accordingly, while the Greens have remained true to their core values, their messaging has veered somewhat from eco-radicalism toward moderate pragmatism. The first experiment demonstrates that the SPD and CDU initially showed little interest in climate change, and continue to primarily address its economic aspects. Meanwhile, the Greens have increasingly focused on economic considerations. The subsequent experiment explores the usage of compound words created with the word Klima (climate). These *climate compounds*, viewed as expressive framing devices, can mark key turning points in climate discourse, and would identify three distinct periods: one of little attention (1980 - 2009), one of description (2009 – 2013), and one of calling for climate action (2013 – ?). The insights presented in this thesis help model and understand policymakers' portrayals of climate change – an essential prerequisite to communicating about and mitigating the issue – and are thus a valuable addition to interdisciplinary climate change research.

## Abstract

Der Klimawandel ist längst nicht mehr nur ein physikalisches Phänomen, sondern auch eine soziale, politische und kommunikative Herausforderung. Ausgehend von der Schnittstelle zwischen Computerlinguistik und Politikwissenschaft werden in dieser Arbeit zwei Studien vorgestellt, die die Perspektive der politischen Entscheidungsträger auf das Thema untersuchen und die Entwicklung des Klimawandeldiskurses in den deutschen Parteiprogrammen der beiden großen Parteien, CDU und SPD, sowie der Partei Die Grünen analysieren. Ein transformerbasierter Ansatz zur dynamischen Themenmodellierung wird implementiert, um den Diskurs in den Wahlprogrammen von 1980 bis 2021 zu untersuchen. Die Ergebnisse zeigen, dass die Volksparteien nach anfänglich geringem Interesse vornehmlich die wirtschaftlichen Aspekte des Klimawandels adressieren. Auch der Diskurs der Grünen entwickelt sich zunehmend zu wirtschaftlichen Überlegungen, nachdem umweltbezogene Themen zunächst die Priorität der Partei waren. Die zweite Studie konzentriert sich auf Klimakomposita, zusammengesetzte Wörter um das Wort Klima (z.B. Klimawandel, Klimakrise). Betrachtet man diese Wörter als ausdrucksstarke Framing-Elemente, so lassen sich drei Segmente in der Entwicklung des Klimadiskurses identifizieren: (i) geringe Aufmerksamkeit (1980 – 2009), (ii) Beschreibung der Situation (2009 – 2013) und (iii) Aufruf zum Klimaschutz (2013 - ?). In beiden Studien wird ein besonderes Augenmerk auf die Entwicklung der Grünen Partei gelegt, die ihren Grundwerten treu bleibt, aber ihren Diskurs kontinuierlich vom anfänglichen Ökoradikalismus zu einem moderaten Pragmatismus verschiebt. Die Ergebnisse beider Studien zeigen außerdem, dass Klimadiskurs ab 2009 ein fester Bestandteil aller Wahlprogramme ist. Die in dieser Arbeit vorgestellten Erkenntnisse tragen dazu bei, die Darstellung des Klimawandels durch politische Entscheidungsträger zu modellieren und zu verstehen - eine wesentliche Voraussetzung für die Kommunikation und Eindämmung des Problems.

# Contents

1	Introduction		1			
2	Background Concepts					
	2.1	Climate Discourse in Politics	4			
	2.2	Political Science	5			
		2.2.1 Why Manifestos?	5			
		2.2.2 <i>Die Grünen</i> in Context	6			
		2.2.3 Green Politics in Germany	8			
	2.3	Natural Language Processing	9			
		2.3.1 NLP in Political Science	10			
		2.3.2 Topic Modeling	11			
		2.3.3 Climate Compounds as Framing Devices	14			
3	Data	I .	17			
	3.1	Manifesto Corpus	17			
	3.2	Manifestos and NLP	18			
	3.3	Subcorpus	20			
4	Methodologies 23					
	4.1	Topic Modeling	23			
		4.1.1 BERTopic	24			
		4.1.2 Data Preprocessing	26			
		4.1.3 Parameters	27			
	4.2	Climate Compounds as Framing Devices	29			
5	Res	ılts	33			
	5.1	Topic Modeling	33			
	5.2	Climate Compounds as Framing Devices	35			
6	Disc	ussion	41			
	6.1	Topic Modeling	41			

	6.2 6.3	Climate Compounds as Framing Devices	43 46
7	Con	clusions & Future Research	49
Bibliography			52
Α	Figu A.1 A.2	r <b>es</b> Figures for Topic Modeling	<b>60</b> 60 61

v

# **List of Figures**

2.1	Election posters of <i>Die Grünen</i> 1980 and 2021	7
3.1 3.2	Length of German election manifestos in words (all parties) . Word clouds generated from party manifestos 1980-2021 (all	21
	parties)	22
4.1	Topic model selection process	28
5.1	Dynamic Topic Modeling visualization for environment-related	
	topics (all parties)	36
5.2	Frequency of climate compounds in German election mani-	
	festos (all parties)	37
5.3	Evolution of climate compound usage by category (all parties)	37
5.4	TOP 3 most frequent compounds (all parties)	38
5.5	Frequency of <i>Klimakrise</i> (all parties)	39
5.6	First mentions of climate compounds (all parties)	40
A.1	Hierarchical clustering of topics (Die Grünen)	60
A.2	Hierarchical clustering of topics (CDU)	60
A.3	Hierarchical clustering of topics (SPD)	61
A.4	Frequency of <i>Klimaschutz</i> (all parties)	61
A.5	Frequency of <i>Klimawandel</i> (all parties)	62
A.6	Evolution of climate compound usage by category (Die Grünen)	62
A.7	Evolution of climate compound usage by category (CDU)	63
A.8	Evolution of climate compound usage by category (SPD)	63

# List of Tables

4.1 Climate compound dictionary excerpt before and after n		
	harmonization	29
4.2	Example of climate compounds with categories	31
5.1	Global TOP 10 topics (all parties)	33
5.2	TOP 3 most frequent climate compounds (all parties)	38
5.3	TOP 5 most frequent compounds (all parties)	39

# **List of Abbreviations**

API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
CC	Climate Change
CNN	Convolutional Neural Network
UMAP	Uniform Manifold Approximation and Projection
CSS	Computational Social Science
DTM	Dynamic Topic Modeling
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
PCA	Principal Component Analysis
t-SNE	t-distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection
Party Names	
ĊDŮ	Christlich Demokratische Union
SPD	Sozialdemokratische Partei Deutschlands
FDP	Freie Demokratische Partei
Die Grünen	Bündnis 90/Die Grünen
AfD	Alternative für Deutschland

### Chapter 1

# Introduction

From *climate change* to *climate crisis*, many terms have been employed to characterize one of the biggest challenges humanity has ever faced. Indeed, finding the right words to describe such an abstract, yet existential threat poses a key challenge in itself. While scientific consensus is all but unanimous that humans play a substantial causative role in climate change, the evidence and implications of this are often lost in translation for laypeople (Penz, 2022). It is the responsibility of democratic policymakers to transform the imminent threats of climate change into actionable policy, which requires effective communication of these dangers – not only within their party and parliamentary structures, but also outwardly to the general public. Modeling and understanding the political discourse surrounding climate change is key to understanding how narratives around the matter evolve and how public opinion is shaped.

Relegated to the political agendas of niche parties before the onset of the 21<sup>st</sup> century, environmental issues and the imminent threats posed by climate change have since entered public discourse and mass media, and consequently mainstream politics (Gärtner and Schoen, 2021). Green parties paved the way for this increasing political awareness. After having emerged as mostly fringe, single-issue parties in the 1970s, Green parties have since achieved considerable election success in a number of Western societies. The German Green Party *Bündnis 90/Die Grünen* can be considered a prototypical case of such success in global green politics (Grant and Tilley, 2019). In the 2021 German federal election, it received a historically high 14.5% of votes and subsequently formed a coalition government with the SPD, a social democratic *Volkspartei* (i.e., an established mainstream party), and the FDP, a liberal party.

Die Grünen's success seems to have grown in tandem with public concern

about climate change and the shortcomings of mainstream parties in adequately addressing it. The nature of climate change makes communication about it particularly challenging: not only is it an issue with complex causes and ramifications, but it also cannot be solved by means of traditional policymaking. The 40 years since the formation of the Greens have also seen substantial developments in scientific discoveries, public perception, and framing of environmental issues. Framing, in simple terms, is the way information is presented, be it in everyday conversations, by media, or by politicians (Stecula and Merkley, 2019). A classic example can be the decision to speak of a glass being half-full or half-empty. Similarly, the two terms climate change and climate catastrophe refer to the same concept but aim to evoke different feelings in the intended audience (see Li and Su (2018) for a review of framing in climate change communications). Investigating the framing of issues is part of discourse analysis, in which language is examined in its social, cultural, and/or historical context. Assuming the gestalt of meaning in language is greater than the sum of the discrete words it contains, discourse analysis strives to analyze and reveal the underlying meaning expressed in natural language (Pan and Kosicki, 1993).

In this thesis, established Natural Language Processing (NLP) methods are employed to investigate these dimensions of meaning in the sphere of German politics – with a focus on the development and role of the Green Party – by analyzing the evolution of climate change discourse in German party manifestos. Climate-related topics and their longitudinal development are examined using a transformer-based dynamic topic modeling approach. Furthermore, the emerging vocabulary around climate change and its development are explored by leveraging the discursive cluster of *climate compounds*. These lexical creations are joint nouns containing the German word *Klima* (climate), which can serve as framing devices for climate discourse. Similarly, they could provide key insights into the linguistic processes involved in climate change discourse.

The interdisciplinary approach of this thesis extends existing political science research with current methods of computational linguistics and NLP to investigate how climate change discourse in German election programs evolved over time. To this end, changes in climate-related topics and framing of climate change are identified and analyzed. In addition, it is examined how different actors (i.e., the established parties and the Greens) talk about climate change. Special attention is paid to the role and development of the Green Party.

In order to explore these issues, the thesis first introduces relevant background concepts in Chapter 2, starting with a general overview of (partylevel) climate discourse in politics, Green Party history, and green politics in Germany. Chapter 2 continues with a brief explanation of NLP applications in political science and the techniques used in this study, alongside a presentation of relevant interdisciplinary research. Chapter 3 then introduces the Manifesto Corpus – the primary data source utilized in this study – and provides context of previous research conducted using this corpus, as well as insights into the subcorpus created for analysis. Subsequently, Chapter 4 offers a detailed overview of the methodologies and techniques implemented in the study, followed by the presentation of the experimental results in Chapter 5. These results are then discussed and interpreted in Chapter 6. Lastly, Chapter 7 provides concluding thoughts and proposals for future research.

### Chapter 2

## **Background Concepts**

In this chapter, background concepts relevant to the work will be introduced. First, a general overview of climate change discourse in politics is given, which is then placed in the context of German green politics, emphasizing the importance of discourse research at the party level. Afterwards, Natural Language Processing tools for climate change discourse analysis relevant to this work will be presented, briefly explaining different topic modeling approaches suitable for framing analysis and elaborating on the role of *climate compounds* as discursive markers.

### 2.1 Climate Discourse in Politics

As discussed by Stede and Patz (2021), the term *discourse* is not a precise one, and its interpretation depends vastly on the context it is used in. In the field of environmental politics, Hajer and Versteeg (2005) define discourse as "an ensemble of ideas, concepts, and categories through which meaning is given to social and physical phenomena". The notion that language ascribes meaning to phenomena negates the existence of a single *true* reality and instead implies the co-existence of multiple realities which gain and lose importance depending on societal context. One device for shaping societal context is *Framing*, the communication of an issue based on a narrative, with the intention of invoking a certain way of looking at said issue (Nisbet, 2009). Frames are an unavoidable concept in human communication and do not imply a misrepresentation of facts, rather a (strategic) redistribution of importance concerning an issue. An example of environment-related framing is the nuclear power discourse in the U.S.: in its early years, the concept of a *nuclear future* was promoted with slogans such as the famous *Atoms for peace* by former

U.S. president Eisenhower. Nuclear power was framed as a futuristic technology providing the means for social progress and economic growth. These frames coexisted with the (much quieter) voices of people opposed to nuclear power, who called for "public accountability" and alternative ways to energy independence (Nisbet, 2009). The Three Mile Island partial nuclear meltdown that coincided with the release of the nuclear disaster thriller movie The China Syndrome 12 days prior changed the framing of nuclear power in public discourse immensely from opportunity to threat to nature and people (Anshelm and Hultman, 2015). The previously important positive aspects of nuclear power lost relevance and the dramatic events were rarely contextualized in terms of previous safety records, leaving the discourse centered around a generally unfavorable framing of all things nuclear (Friedman, Gorney, and Egolf, 1992). Climate change discourse as it exists today is still in a perceptual divide with opposing narratives, a divide between growth and sustainability; freedom and responsibility. The role of politicians in climate change discourse is not only to address target audiences and perpetuate narratives, but also to adapt to frames prevalent in media and society, which are in ever-changing interdependence. As Hajer and Versteeg (2005) phrase it: "[l]anguage has the capacity to make politics". Discourse analysis is the study of languagein-use, acknowledging its ability to profoundly shape reality, and provides a means of analyzing the role of language in politics, thus also revealing insights to society, its norms and conventions (Hajer and Versteeg, 2005).

### 2.2 Political Science

In this section, relevant background information from the political science domain will be provided. First, the role of election manifestos in discourse research is explained, followed by a brief introduction of the German Green Party's history and an overview of green politics in Germany.

#### 2.2.1 Why Manifestos?

Farstad (2018) discusses the lack of party focus in comparative climate policy research, stating that analyzing political parties is key to understanding climate discourse in politics. Party competition has a large impact on government policymaking and thus also on a country's climate change mitigation response. Parties play an important part in the aforementioned interdependence, as they aim to bridge the gap between abstract politics and society. Stede and Patz (2021) argue that the climate change debate is unfolding among three different actor communities: the general public, the policymaking communities and the scientific communities. Furthermore, they state that "[e]ach community uses different genres, registers, and terminologies to communicate with each other and with other communities about CC. These communities shape individual and collective ideas, frames, and, ultimately, the behavior that is consequential for the future evolution of anthropogenic climate change." (Stede and Patz, 2021). Looking at the political discourse in a democratic country such as Germany, manifestos can be considered a unilateral means of communication between two of the aforementioned actor communities: the *policymaking community* is communicating their policy plans to the general public. Election manifestos are the key elements of a party's election campaign and represent its views, goals, and pledges for government participation. What is important here is that the election programs represent the opinion of the party as a whole (or at least of its democratically elected common denominator), its identity, so to speak.

#### 2.2.2 Die Grünen in Context

Green parties were seen as single-issue parties for the longest time. Only having emerged in the second half of the 20<sup>th</sup> century, they constitute a relatively new phenomenon in the political landscape, with the international Green network, the Global Greens, founded as recently as 2001. The German Green Party unconventionally started off as an institutionalization of an array of social movements, rather than a formally established party in the West of divided Germany (Milder and Jarausch, 2015). These social movements were composed of citizens who were strongly committed to social and ecological issues such as environmental protection and the rejection of nuclear energy and war, but who did not feel seen by the incumbent government. In addition, there was a clear anti-war stance and a generally pacifist attitude among the supporters. The movement's self-image was not that of a traditional party – quite the contrary: Petra Kelly, one of the founding members of *Die Grünen*, called it an *anti-party party*. Non-traditional bottom-up

hierarchies and the rejection of the idea of a single party leader reflected this (Kwidziński, 2020). The patchwork nature of the party made it heterogeneous from the very beginning. Radical activists were paired with idealistic ecological socialists, who were in turn trying to accommodate the ecolibertarians. Despite these inherent differences, the Greens have defined their core values, their common agenda, in the shape of four pillars: *ecological wisdom*, grassroots democracy, social justice, and nonviolence (Carter, 2013).



FIGURE 2.1: Election posters of Die Grünen 1980 and 2021

In 2022, four decades after the Greens' foundation, this common agenda has transformed – a development that is also reflected by the party's election posters. The Greens' 1980 election poster (Figure 2.1a) focuses on the aforementioned *four pillars*. Entitled *Mut zum politischen Frühling* (Courage for a political spring) and adorned with painted flowers, it introduces the newly emerged party as a transformative force which is *ökologisch, sozial, basisdemokratisch, gewaltfrei* (ecological, social, grassroots, non-violent). The 2021 poster (Figure 2.1b), on the other hand, conveys a different message. Displaying the slogan *Wirtschaft und Klima ohne Krise* (economy and climate without crisis), along with a portrait of Annalena Baerbock, the party's candidate for chancellor in the 2021 federal election, the poster gives an impression of the party as a competent force with government potential, even in challenging times. While the *climate crisis* is taken into account, the pragmatic goal of a stable economy is stated first, addressing a more tangible problem and with it a broader audience.

As of 2022, the Greens have been coalition partners in 11 of the 16 German state governments and currently constitute the country's government in a coalition with the SPD and the FDP. Their election success shows that the Greens have gained the position of a valuable coalition partner in the current German party system. While this is likely to secure the party an ongoing role in the German political landscape, a general openness to connectivity into all directions can come with compromises and changes to original values (Blühdorn, 2009). In her extensive report of the Green Party, Sloat (2020) investigated this perceived transformation of the Greens "from the left-wing fringe to the moderate middle". According to the report, "Germany needs the Greens to act as a moderating force during this era of polarization; but in its attempt to appeal to as many voters as possible, the party may be defined by what others want, fail to achieve its core socio-ecological objectives, and eventually fall out of favor.". Kwidziński (2020) takes a similar stance and states that the Green Party has undergone a "departure from fundamentalism in favor of pragmatism, an ideological transformation towards a political center, the associated increase in coalition capabilities and an increase in political relevance".

#### 2.2.3 Green Politics in Germany

The political landscape in Germany is shaped through a gradual transformation rather than a radical change, shaped by the emergence of new parties and their competition in the German six-party-system<sup>1</sup> (Anderson, 2015). Since the formation of (West) Germany in 1949, its federal system has been dominated by the two mainstream parties SPD and CDU and the liberal party FDP, who acted as a pivotal party and took up a role as coalition partner for either of the other two parties (Saalfeld, 2002). The German reunification in 1990 did not change this and the government has consistently been constituted of at least one of these two *Volksparteien* (major parties), if not both of the parties together as a *grand coalition*.

<sup>&</sup>lt;sup>1</sup>CDU, SPD, FDP, Die Grünen, Die Linke, AfD

Framing, as discussed in Section 2.1, creates the boundaries around an issue and allows certain actors to claim ownership of it (Koteyko, Thelwall, and Nerlich, 2010). In the context of German politics, the ownership of the green issue has historically belonged to the Green Party (Spoon, Hobolt, and Vries, 2014). The party has vastly contributed to politicizing the green issue in Germany, as it campaigned for Germany to abandon nuclear energy as early as 1983 and, in a government coalition with the SPD, introduced the nuclear phase-out in 2000 (Sander, 2016). However, with the rise of public awareness and concern related to climate change, and the growing electoral success of the Greens, the two Volksparteien were forced to adapt and enter the hitherto uncharted territory of green politics, causing the Greens to lose the sole ownership of their fundamental green policy (Spoon, Hobolt, and Vries (2014); Green-Pedersen and Mortensen (2010)). In 1983, the Greens first entered the German parliament, and eventually co-ruled Germany alongside with the SPD from 1998 until 2005, after which the Grand Coalition came into power. In the following 16 years without government participation of the Green Party, it became increasingly difficult for the other parties to keep the green issue down on the agenda. In 2011, after the Fukushima nuclear disaster, the government around chancellor Angela Merkel announced the shutdown of all German nuclear power plants by December 2022, revoking the lifetime extension decided just one year prior and, in a way, returning to the green energy vision introduced by the Red-Green government (Wittneben, 2012).

As green issues have made it to the top of the policy agendas, both environmental ambitions among mainstream parties and the radical position of the Green Party seem to have changed. Analyzing the evolution of climate change discourse in German politics can shed light on the arising programmatic challenges and contribute to identifying effective ways to communicate climate change in the political sphere.

### 2.3 Natural Language Processing

In this section, relevant background information from Natural Language Processing will be presented. First, an overview of the role of NLP in political science will be described. Afterwards, background on the NLP techniques and concepts relevant for the analysis in this thesis will be provided, starting with an explanation of topic modeling and moving on to an introduction of *climate compounds*.

#### 2.3.1 NLP in Political Science

The amount of available text data is constantly increasing, and the field of machine learning has made rapid progress in recent decades. More importantly, the novel computational approaches to empirical research, an integral part of social science, have become more accessible to the scientific community, as the boundaries between computer science and other disciplines have increasingly blurred. Tutorials and tools for a plethora of computational research applications became freely available in the digital age, lowering the threshold of entering the emerging field (Lazer et al., 2009). Buyalskaya, Gallo, and Camerer (2021) talk about this new era of social science as its golden age – one that calls for interdisciplinary research and collaboration instead of mere passive borrowing from other fields. Nowadays, there are often "teams of computationally literate social scientists and socially literate computer scientists" (Lazer et al., 2009). Increasingly, data-driven machine learning approaches have been employed to collect and analyze data, and facilitate quantitative research, leading to the emergence of *computational social* science (CSS) (Lazer et al., 2009).

In political science, as one of the social sciences, research is often concerned with text-as-data. Subjects of interest to scholars, such as elections, campaigning, party competition, and public policy, can be studied through analyzing text. This involves drawing on various sources that can be analyzed, such as party platforms, press releases, speeches, and social media such as Twitter. Given the abundance of political textual data in the digital age, it is virtually impossible for scholars to manually read and analyze all relevant texts. Here lies the promise of computational social science: it allows established methods to be applied on a much larger scale and at a fraction of the cost (both financial and time), which "can make possible the previously impossible in political science" (Grimmer and Stewart, 2013). In the past, for example, researchers have redistributed resource-intensive tasks like manual coding to expert human coders, who were to manually read all the text in a corpus and extract the desired information. Today, this task can be delegated to Natural Language Processing and be performed semi-automatically on large corpora using supervised text classification techniques (Marathe and Toyama, 2018). Other applications of NLP methods for handling text-as-data in political science include (but are not limited to) performing ideological scaling (Slapin and Proksch, 2008), the examination of priorities expressed by political actors (Grimmer and Stewart, 2013), and analyzing the policy positions and sentiment in parliamentary debates (see Abercrombie and Batista-Navarro (2020) for a systematic literature review).

Adding the prefix computational to social science brings opportunities and challenges alike. Grimmer and Stewart (2013) remind the reader that automated text analysis methods do not fully relieve them of the menial necessity of at least some kind of *human understanding* of a corpus. In the end, the most important decisions are up to the researcher to make. They have to collect and clean the data, choose an appropriate model for analysis, interpret its output and validate their findings. In summary, Grimmer and Stewart (2013) say: "[r]ather than replace humans, computers can *amplify* human abilities.". They also state, however, that treating CSS as the interdisciplinary endeavor it is and continuing to improve and validate statistical text models has the potential to revolutionize political science research. Bridging the gap between two disciplines is always a challenge, and this case is no different. When thinking of the common stereotypes for computer science and social science, it is not so intuitive to imagine an overlap – after all, computer science does not seem too concerned with people. However, the interdisciplinary approach CSS takes shows how valuable it can be to step out of the scientific comfort zone, and how two different disciplines can complement rather than contradict each other. Computer scientists can benefit from the expertise that political scientists bring to the table, which includes decades of rigorous manual coding and an extensive knowledge of theoretical frameworks for analyzing political texts. Political scientists, in turn, can benefit from the programming skills and conceptual knowledge of computer scientists and learn to perform automated content analysis at scale using sophisticated machine learning algorithms (Wilkerson and Casas, 2017).

#### 2.3.2 Topic Modeling

Probabilistic topic modeling is a statistical technique in text mining which allows for hidden data discovery. It is used for extracting the main themes in unstructured documents by clustering words that frequently occur together. Topic models do not rely on prior human annotation of the textual data, which allows for convenient and unsupervised organization of documents at a large scale (Blei, Ng, and Jordan, 2003). The algorithm usually employed for topic modeling is Latent Dirichlet Allocation (LDA), which was introduced by Blei, Ng, and Jordan (2003). For this Bayesian topic modeling approach, it is assumed that a set of documents can be modeled by a set of topics, which is defined by words that belong to it with a certain probability. LDA uses the bag-of-words representation of a document, thus leaving word order out of consideration.

Since topics simply consist of words that frequently occur together, it is not a given that they have an inherent meaning or describe an actual discussed topic. An example of this are function words (e.g., articles, prepositions), which could form a *topic* in the sense of the statistical co-occurrence definition but do not describe the corpus in any meaningful way. Topics may also not be clearly distinct because frequent words may occur in multiple topics, making a clear differentiation challenging to impossible (Ramage et al., 2009). The final labeling of the extracted topics is a manual (and consequently subjective) task that requires thorough investigation of the clustered words with the highest probability scores per topic.

Additional steps to ensure intratopic semantic validity can include the manual analysis of a sample of documents containing a large share of the investigated topic to determine if a topic *makes sense* (Maier et al., 2018). Despite its popularity, there is no state-of-the-art (automated) strategy for ensuring the validity and reliability of topic modeling results. Computational approaches for topic model validation and evaluation include diagnostic metrics such as topic diversity (Dieng, Ruiz, and Blei, 2020) and topic coherence (Bouma, 2009), and may serve as a proxy for human evaluation.

#### **Dynamic Topic Modeling**

Topic models such as LDA are designed for categorical data and thus assume independence of documents. A shortcoming of the established LDA approach is its lack of consideration for the interdependencies of documents. The static nature of the basic topic modeling approach does not allow for modeling of temporally-organized documents, such as newspapers or election programs. For some applications, assuming independence of documents will thus omit crucial information. In order to overcome this drawback, Blei and Lafferty (2006) introduced Dynamic Topic Modeling (DTM). DTM augments classical topic modeling and allows for the analysis of the temporal evolution of fixed topics by assuming the data is divided by time slice (e.g., year) rather than independent. This implies that global topics can be linked and examined in a temporal-spatial manner, even if their representation differs vastly across the examined time frame (Blei and Lafferty, 2006).

For example, documents assigned the topic *consumer electronics* investigated throughout the past two decades might for example contain words like *VCR* or *Disc man* in the 1990s – words that are extremely unlikely to appear in the very same topic 20 years later. Despite these differences, the documents might nonetheless have fundamental words, just like *computer* or *television*, in common.

In the past, DTM has been successfully used for political discourse analysis. An example for this is the study published by Müller-Hansen et al. (2021), who analyzed 70 years of German parliamentary debates on coal. Their analysis revealed that economic topics have consistently played a dominating role in the German parliamentary debates on coal. In recent years, however, the relative importance of economic coal-related debates has declined. Environmental topics like climate protection and the general discussion of energy transition occurred more frequently and became linked to the coal debate. The topic modeling results suggest that a shift in the framing of coal has taken place and moved the debates away from *coal as a main economic driver* and in the direction of *coal as an environmental problem*. Applying DTM to mediums of political communication can, according to the authors, lead to nuanced insights into temporal developments of the political discourse related to climate change and the energy transition.

#### BERTopic

Grootendorst (2022) introduced the neural topic modeling technique called BERTopic, which leverages *Bidirectional Encoder Representations from Transformers* (BERT). The BERT language model was released in 2019 by Devlin et al. (2019) and "has become a ubiquitous baseline in NLP experiments" in little over a year by 2020 (Rogers, Kovaleva, and Rumshisky, 2020), still obtaining state-of-the-art results in various NLP applications as of 2022. BERT models, unlike bag-of-words models, rely on an embedding approach. This means they account for the context of words in a sentence. This is important, as language is often ambiguous. Context can for example help resolve polysemy and clarify if *mouse* in the sentence *A mouse ate my shoe*. refers to the small rodent or its silicon relative, the computer pointing device. Embedding models make it possible to spatially locate words in a vector space because the more similar two words (or sentences, documents) are, the closer to each other they can be found in a vector space (Kusner et al., 2015). Using these powerful embedding models, BERTopic pursues a novel approach to topic modeling. Recent studies have found it to work excellently with pre-trained embeddings, frequently outperforming other topic modeling algorithms (Egger and Yu, 2022). BERTopic offers a variety of topic modeling extensions to researchers, such as multi-language support and automated topic reduction. Furthermore, it provides access to variations such as dynamic topic modeling (Grootendorst, 2022).

#### 2.3.3 Climate Compounds as Framing Devices

Nerlich, Evans, and Koteyko (2011) stress the importance of the linguistic analysis of climate change, stating that "[n]ew words and new concepts provide people with new ways of experiencing themselves and their world, with new ways of being and new ways of knowing". Humankind has long lacked the words to properly describe climate change and the looming threat posed by it, as knowledge and awareness in the public were limited until the second half of the 20<sup>th</sup> century (Nerlich and Koteyko, 2009). Understanding the evolving vocabulary that aims to make sense of this complex issue is fundamental to interdisciplinary research on climate change mitigation and scientific communication with the general public. The novel collective linguistic response to this matter can give insights into the framing of climate change (Nerlich and Koteyko, 2009). One discursive cluster of the lexical changes accompanying this response are compounds in climate change communication. Compounding words is considered one of the most productive and efficient word formation processes in the English (and German) language (Booij, 2011). Their nature is simple: compounds can be defined as "words that are composed of two (or more) bases, roots, or stems" (Lieber, 2010).

*Jellyfish*, for example, is a compound noun that describes a marine animal with a gelatinous consistency and simply combines the two nouns *jelly* and *fish*, creating an apt description of the animal. Existing work on corpus linguistic-based lexical compound analysis in climate change discourse has

focused on English-language data, mostly in the public sphere (see Nerlich and Koteyko (2009); Nerlich, Koteyko, and Brown (2010)).

Koteyko, Thelwall, and Nerlich (2010) conducted a study on carbon compounds and their role in web-based climate change discourse and framing in the public sphere. Carbon compounds are lexical combinations around the word carbon, such as carbon tax or carbon emissions. Koteyko, Thelwall, and Nerlich (2010) claim carbon compounds to be examples of an emergent shared carbon lexicon which can be used to index, track and study the existing framing of the climate issue, as the compounds represent a condensed representation of collective creativity. This lexicon could, according to the researchers, be leveraged to improve science communication and strive towards a common climate language. A case study on the media coverage around the carbon compound low carbon diet was conducted by Nerlich, Evans, and Koteyko (2011). The study found that the term first occurred and gained prominence in U.S. newspapers in 2006, even though it has been around since the early 2000s. The researchers attribute this to the increased pressure exerted by scientific evidence around anthropogenic climate change and the raised awareness in society. The influence of Al Gore's climate campaign, particularly the impact of the 2006 documentary An inconvenient truth, is cited as an important factor. Their findings follow up on the results of Koteyko, Thelwall, and Nerlich (2010) and indicate that the emergence of new lexical expressions in the form of carbon compounds is a way of making climate change amenable to understanding and acting. Furthermore, they allow for a simplified framing of a complex issue, making them important communicative devices in climate change discourse.

As described in Section 2.2.1, manifestos are a means of communication between the policymakers and the public. Looking at the presented research, it can be assumed that the evolution of climate change framing in public discourse can be analyzed using climate compounds as framing devices. As the scientific topic of climate change has entered the social discourse, logically not only citizens but also policymakers need new terminology that is tailored to the specific audience (Hajer and Versteeg, 2005). Here is where the concept of *framing* comes into play. Naturally, each party has its own story line regarding climate change, placing more weight on certain aspects of the matter or looking for responsible entities in different places – they frame the issue of climate change in different ways. This does not mean that there is an inherently *right* or *wrong* way of talking about the topic. As Nisbet (2009) puts it: "[t]here is no such thing as unframed information".

Transferring the concept of climate compounds to German is simple, as the language allows for easy compounding. Here, the equivalent are *Klimakomposita* (climate compounds), words which are build around the word *Klima*. The presented studies suggest that climate compounds, and thereby probably also *Klimakomposita*, can serve as markers of the evolution of framing in climate change discourse, and a longitudinal study of their occurrence in election manifestos can potentially reveal discursive changes in the German political landscape.

## **Chapter 3**

# Data

In this chapter, the *Manifesto Corpus*, a large human-annotated and standardized corpus of election manifestos, will be introduced. In Section 3.2, a brief overview of related research utilizing the Manifesto Corpus will be presented. Finally, the subcorpus created for this study will be described in Section 3.3.

### 3.1 Manifesto Corpus

Section 2.2.1 elaborates on the importance of election manifestos in partylevel discourse research. In order to study climate change discourse and its evolution in German party manifestos, this study will leverage the *Manifesto Corpus* provided by the Manifesto Project. The project was formerly called *Manifesto Research Group* (MRG) in 1979-1989, then *Comparative Manifesto Project* (CMP) from 1989 until 2009 and is now operating under the name *Manifesto Research on Political Representation* (MARPOR). For simplicity reasons, this thesis will refer to it with the name *Manifesto Project*. The Manifesto Corpus is an open-access, annotated text corpus containing electoral programs from more than 60 free, democratic countries since 1945 (Lehmann et al., 2022a). It is based on the work of the Manifesto Project since its beginning in 1979.

As of November 2022, the corpus contains more than 1.6 million humancoded digital quasi-sentences (i.e., (sub-)units of sentences containing exactly one policy statement (Merz, Regel, and Lewandowski, 2016)) from nearly 5,000 different manifestos published by more than 1,200 parties<sup>1</sup>. With these numbers, the corpus ranges among the largest manually annotated, international and multilingual text corpora in social and political science. The goal of the project is to make electoral programs freely available for research in

<sup>&</sup>lt;sup>1</sup>https://manifesto-project.wzb.eu/

a digitized and annotated form (Merz, Regel, and Lewandowski, 2016). Access to the coded election programs in the Manifesto corpus is free and allows for the download of subsets of the Manifesto Corpus based on user-defined criteria, such as specific codes, election years, or language. Options for accessing the data are an online web application, the R package manifestoR (Lewandowski, Merz, and Regel, 2020), the manifestata package for STATA and an application programming interface (API), depending on the user's needs.

Human coders native to the respective countries annotate the quasi-sentences based on a coding scheme containing 76 codes, which capture substantial policy issues. To ensure reliability and consistent use of the coding scheme in different countries and across different elections, the Manifesto Project coders undergo a standardized training process and receive expert feedback during multiple stages of the coding process (Lehmann et al., 2022b). Like all human-coded and crowd-sourced data, the Manifesto corpus cannot claim infallibility. Researchers have examined and identified methodological problems in its approach, with validity, reliability and bias of the human coding being the main concern (see: Mikhaylov, Laver, and Benoit (2012)). Despite the coders being *experts*, i.e. people with a social science background living in the manifesto country, and the extensive review process, it is impossible to avoid measurement errors, as manual annotation inherently comes with a degree of subjectivity. While these problems should be taken into account when performing analyses, it is important to keep in mind that this is one of the standard pitfalls of treating bodies of text as data. For good scientific practice and reproducibility of results, a critical look at the data is essential.

#### 3.2 Manifestos and NLP

As mentioned in Section 2.3.1, computational social science opened up new research directions for political scientists. In light of the recent *text-as-data* wave in political science, a considerable amount of methodological research on policy positions was conducted using the Manifesto Corpus<sup>2</sup>. Since the corpus provides labeled text data, most methodological studies leverage supervised machine learning techniques, for example, for performing classification tasks. However, due to the nature of the study presented in this thesis,

<sup>&</sup>lt;sup>2</sup>see: https://manifesto-project.wzb.eu/publications/all

existing research concerned with unsupervised and/or transformer-based machine learning approaches to manifesto content analysis is of particular interest.

Chatsiou (2020) built a sentence-level political discourse classifier by training a convolutional neural network on a subset of annotated English-language manifestos. The 7-domain CNN classifier was successfully applied on political manifestos and transferred to other political texts, in this case the COVID-19 Press Briefings Corpus, where political discourse classification also yielded good results. Enhancing the CNN classifier with out-of-the-box BERT embeddings improved the classification performance significantly<sup>3</sup>. Although the approach taken in this study is supervised in nature, it is of interest in that it demonstrates the feasibility of automatic discourse analysis with BERT without the need to adapt the model to the political domain.

Zirn and Stuckenschmidt (2014) analyzed and compared German election manifestos using topic modeling, more precisely an extension of Latent Dirichlet Allocation that allocates documents to a set of predefined topics. They found the approach to be viable for the reliable comparison of individual topics and predefined policy areas across documents. The findings suggest that topic modeling techniques can be successfully applied to the specific subcategory of the political domain, allowing for analysis of (German) election manifestos.

Another study conducted on German-language data was published by Olbrich and Banisch (2021), who analyzed German party manifestos from 1949 to 2017 by employing LDA. The study found topic modeling to be a viable unsupervised approach to identify political spaces, a concept which spatially locates a party, augmenting the traditional left-right axis and allowing for the creation of distinct party profiles. This approach also yielded good results for a newly emerged populist party, the *Alternative für Deutschland*. The results suggest that German party discourse and positions can be captured meaningfully by performing topic modeling on party manifestos.

The review of existing methodological publications based on the Manifesto Corpus reveals that to date, there is no research focusing on the analysis of climate change discourse in German party manifestos and little research using unsupervised NLP approaches. Furthermore, researchers have so far not employed climate compounds as discursive markers for investigating

<sup>&</sup>lt;sup>3</sup>F1-score of 48.62 and 64.58 for CNN+Word2Vec and CNN+BERT, respectively

the evolution of climate change framing in election manifestos. The study presented in this thesis thus constitutes a novel approach to the investigation of the linguistic and discursive evolution of climate change discourse in German politics.

### 3.3 Subcorpus

For the purpose of studying the evolution of climate change discourse and Green Party evolution in German election manifestos, a subcorpus of the Manifesto Corpus was created. This subcorpus contained the election manifestos of the three investigated parties, Die Grünen, CDU and SPD, from 1980 until 2021. This corresponds to twelve elections, each with three different election programs, for a total of 36 manifestos. The election manifestos used for analysis were attained from the Manifesto Corpus using the R companion package manifestoR<sup>4</sup>. Since German election programs prior to 1998 were only available in digitized, but not in annotated form at the time of analysis, only textual data was considered in this paper, leaving the work with annotated documents to future studies.

To present the characteristics of the manifestos, a descriptive analysis was performed. Figure 3.1 visualizes the manifesto length for each party per year. Manifesto length in this case refers to the amount of words contained in an election platform.

The analysis shows that the election programs of all parties became longer over time. From 1987 onward, the longest programs were always published by the Greens, often more than twice as long as the programs of the CDU and SPD. The 2013 manifesto of *Die Grünen* is the longest one in the subcorpus, consisting of nearly 80,000 words.

Figure 3.2 represents the contents of the manifestos graphically using word clouds. The figures provide a brief and intuitive overview of the most relevant topics for all manifestos published by a party. For the analysis, only (proper) nouns were considered, both as single words and two-word combinations. The size of a word in this visualization is proportional to its frequency in the manifestos. While word clouds do not provide a comprehensive overview of the data, as frequency of a word does not have to correlate

<sup>&</sup>lt;sup>4</sup>https://github.com/ManifestoProject/manifestoR/



Length of Manifestos (No. of Words)

FIGURE 3.1: Length of German election manifestos in words (all parties)

with importance of a word, the visual cues help get a first impression of textual data and consequently the similarities and differences between election programs. For example, the Greens' party manifestos (Figure 3.2a) seem to contain frequent mentions of environmental issues and references to the own party (e.g., *grün, umwelt, erneuerbare energie*), as well as social issues (e.g., *behinderung, grundsicherung, gesellschaft*). In the CDU manifestos (Figure 3.2b), on the other hand, frequent mentions of Germany and its fundamental values can be found (e.g., *deutschland, frieden, freiheit, einheit*). The SPD manifestos (Figure 3.2c) deal, among other things, with civil society and the rights of workers in the country (e.g., *arbeitnehmerin, gewerkschaft, vereinbarkeit familie*).



(C) SPD manifestos

FIGURE 3.2: Word clouds generated from party manifestos 1980-2021 (all parties)

### Chapter 4

# Methodologies

All analyses, unless specified otherwise, were conducted using Python 3.9. For visualization, the interactive open source graphing library plotly.py<sup>1</sup> was used. For the transformer-based neural topic modeling approach, the BERTopic implementation (v0.12.0) provided by Grootendorst (2022) was used. The implementation is freely available on GitHub<sup>2</sup> and the Python Package Index PyPi<sup>3</sup>. For visualization of the topic modeling results, a modified version of LDAvis<sup>4</sup> implementation provided by Grootendorst was used. For accessibility and reproducibility reasons, all code was written and documented in Jupyter notebooks (Kluyver et al., 2016). Computation was performed on Colab<sup>5</sup>, a Jupyter notebook service which provides free-of-charge cloud access to high-performance CPUs and GPUs.

In the following chapter, the methodological background for the techniques used in this study will be provided. In Section 4.1, neural topic modeling and the BERTopic approach to this concept will be explained. Afterwards, the text preprocessing and the process of optimal model selection in the study will be presented. In Section 4.2, the methodological approach to climate compound collection and analysis will be provided.

### 4.1 Topic Modeling

In the following, the methodology behind the neural topic modeling approach BERTopic will be explained. Afterwards, the process of parameter

<sup>&</sup>lt;sup>1</sup>https://github.com/plotly/plotly.py

<sup>&</sup>lt;sup>2</sup>https://github.com/MaartenGr/BERTopic

<sup>&</sup>lt;sup>3</sup>https://pypi.org/project/bertopic/

<sup>&</sup>lt;sup>4</sup>https://github.com/cpsievert/LDAvis

<sup>&</sup>lt;sup>5</sup>https://colab.research.google.com/

and model selection specific to the study of climate discourse in German party manifestos will be presented.

#### 4.1.1 BERTopic

The topic modeling approach adopted in this work is neural topic modeling, more specifically BERTopic. BERTopic is modular and allows for fine-grained parameter tuning of the model. Additionally, it grants free choice of the required sub-models for sentence embedding, dimensionality reduction and clustering. The BERTopic approach to neural topic modeling can be roughly divided into three steps:

- 1. **Embedding**: Transform document to high-dimensional vector embedding representation using pre-trained language model
- 2. **Clustering**: Reduce embedding dimensionality, then create semantically similar clusters of the embeddings (i.e., topics)
- 3. **Topic Representation**: Extract topic representations (i.e., words in a topic) from clusters using class-based TF-IDF approach

**Embedding.** The embedding step leverages Sentence-BERT (SBERT) which allows for creating dense vector representations of sentences. SBERT is a modification of the original BERT network introduced by Devlin et al. (2019), which is fine-tuned for semantic similarity and clustering tasks, significantly lowering computation time (Reimers and Gurevych, 2019).

**Clustering.** After the sentences are preprocessed and transformed into vector representations, clustering has to be performed in order to detect which sentences are semantically similar and potentially belong to the same *topic* (i.e., cluster). As with any work in high-dimensional spaces, the *curse of dimensionality* is haunting the embedding vector space, meaning the data becomes so sparse that concepts like *spatial proximity* between data points lose their meaning. This makes it difficult (or even impossible) to apply the distance metrics traditionally used for clustering (Aggarwal, Hinneburg, and Keim, 2001). One way to solve this problem is to reduce the dimensionality of the vector space, preserving the meaningful dimensions (i.e., attributes) of the data, while discarding dimensions that are merely noise (Ding et al.,

2002). The default method employed by BERTopic is *Uniform Manifold Approximation and Projection* (UMAP) as it preserves more of the data features and global structure than other common methods such as PCA or t-SNE (McInnes et al., 2018).

Once the dimensionality of the vector space is reduced, the data can be clustered. The clustering step is performed using HDBSCAN, a hierarchical clustering-based version of DBSCAN (Density-based spatial clustering of applications with noise). This soft clustering approach recognizes outliers and models them as noise instead of assigning them to a cluster that is *close enough*, thus improving the final topic coherence (Campello, Moulavi, and Sander, 2013). Furthermore, due to the method's hierarchical nature, clusters are arranged logically in a tree-like structure, allowing for an easier understanding and merging of topics.

**Topic Representation.** Based on the newly created clusters of documents, the topic representations are then calculated using c-TF-IDF. This class-based (or: cluster-based) version of the TF-IDF is not so different from its wellknown relative. It is simply adjusted to function for clusters (i.e., topics) rather than individual documents by considering the differences of documents across clusters. In addition, it allows for a ranking of the words in a topic representation by importance, making it easier to understand the essence of a topic. Computing this metric also makes it possible to merge uncommon topics with topics that have a similar c-TF-IDF representation but are more common, providing an option for efficient topic reduction. The basic BERTopic algorithm lays the foundation for the Dynamic Topic Modeling and is technically concluded after this last step. However, in order to acquire meaningful and consistent results, the described process has to be conducted repeatedly. The suitable choice of hyperparameters for each submodel differs vastly across use cases. In the absence of a rule of thumb, researchers must resort to a practice that essentially amounts to *trial and error* in determining the best hyperparameters for their BERTopic model.

**Dynamic Topic Modeling.** The topics and representations created at this point of the process represent a global summary of the documents. For a sequential topic representation (i.e., *What did a topic look like at time step t or t-1*?), Dynamic Topic Modeling has to be performed. After the best model and

with it the global topic representations are determined, the *local* representations of a topic for each year can be created. The arrangement of documents by time slice is a prerequisite for DTM. Therefore, each sentence is assigned a timestamp corresponding to the year in which the respective manifesto was published. For the DTM step of BERTopic, no further clustering or other computationally expensive procedures are needed, as a topic representation for a year is calculated by multiplication of the topic's term frequency for said year with its previously calculated global IDF values.

#### 4.1.2 Data Preprocessing

Mapping the textual data to a vector space is a necessary prerequisite for modeling it with BERTopic. The first preprocessing step in this study was splitting the manifestos into subunits, which could be either sentences or paragraphs. Separating the manifestos into single sentences instead of paragraphs yielded more coherent topic modeling results according to human assessment and computation of the C<sub>v</sub> coherence score. In the study, the manifestos were thus separated into subunits using sentence-level splitting, performing sentence boundary detection using spaCy with the de\_core\_news\_sm pipeline for German.

As demonstrated by Denny and Spirling (2018), there is no *right* way to preprocess textual data for unsupervised machine learning, as the requirements vary across applications. For this application, the literature review and guidelines provided by Maier et al. (2018) were used as reference for the preprocessing pipeline, starting off by cleaning the data of any unwanted characters. In practice, this meant that escape sequences starting with a backslash (e.g., \n for inserting a newline), numbers and non-essential punctuation marks were removed, and the text was transformed to lowercase characters. Afterwards, analogously to Müller-Hansen et al. (2021), all words were lemmatized (i.e., reduced to their dictionary form) using spaCy in order to avoid the processing of inflected forms and thus limit redundancy in the topics (e.g., *election* and *elections* in the same topic). Lemmatizing (rather than stemming) was the term unification method of choice here, since a lemma is usually easier to interpret. In initial trials, lemmatizing has also yielded, according to human assessment, better results than stemming (i.e., reducing words to their stem).

The aforementioned preprocessing guide by Maier et al. (2018) also suggested stop word removal at this stage of the process. However, this was only performed later on in the study - something that might seem counterintuitive to anyone who has worked with textual data before. This is owed to the context-sensitive nature of transformer-based models. It is advised to preserve as much context in the data as possible, as the model leverages these contextual features. Therefore, stop words were not removed from the documents themselves but were only filtered out from the final topic representations (i.e., words that represent a topic). In order to achieve this, a CountVectorizer instance with the standard nltk stop-word list for the German language<sup>6</sup> was initialized. This stop word list was chosen since it is modular and can be expanded as needed, in case further stop words are recognized later on in the process. In this case of working with German election manifestos, an inspection of topic representations in preliminary topic modeling trials revealed domain-specific stop words which were then added to the stop word list, such as party and politican names and variations thereof.

#### 4.1.3 Parameters

The process of optimal model selection undergone in this study is visualized in Figure 4.1.3. It was repeated until the topic model yielded the best (i.e., most coherent) results. Hyperparameters for each submodel (UMAP, HDB-SCAN, BERTopic) were adjusted iteratively. The model performance was evaluated after each round of model fitting via manual qualitative inspection in addition to the computation of the  $C_v$  coherence score. Manual inspection in this case meant reviewing the TOP 15 words of the TOP 15 topics and a visual inspection of the topic model diagrams.

Syed and Spruit (2017) examined the  $C_v$  score as a proxy for topic quality and found scores between 0.5 and 0.6, (with 0.59 being the highest score) to be correlated with topics that were deemed coherent by human assessors. Using the results as orientation, models with a  $C_v$  score below 0.5 were discarded in this study, and models with a score above 0.5 were manually assessed. Once the best model was determined, it was selected and the topic names were assigned to the TOP 10 topics.

<sup>&</sup>lt;sup>6</sup>nltk.corpus.stopwords.words(''german'')



FIGURE 4.1: Topic model selection process

**Word Embeddings.** Since the manifestos are written in German, a multilingual sentence transformer model distiluse-base-multilingual-cased-v1<sup>7</sup> was used. This model can be used out-of-the-box for more than 50 languages and maps the sentences to a 512 dimensional dense vector space.

**Dimensionality Reduction.** UMAP was employed for reducing the dimensionality of the sentence embeddings. The adjusted parameters were: number of neighboring points, the minimum allowed distance of data points and the desired dimensionality of the results (n\_neighbors=3, min\_dist=0.08 n\_components=3).

Limiting n\_neighbors and lowering the value for the min\_dist between data points ensured the preservation of detailed local structures.

**Clustering.** For clustering the UMAP results, HDBSCAN was used. The adjusted parameters were: minimum cluster size and minimum samples (min\_cluster\_size = 50, min\_samples = 40). Keeping min\_samples value slightly lower than the min\_cluster\_size allowed for the modeling of interpretable and stable clusters with reduced noise inclusion.

<sup>&</sup>lt;sup>7</sup>https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v1
**BERTopic.** The adjusted parameters for the BERTopic instance were: the number of extracted words per topic, the topic diversity and the minimum topic size (top\_n\_words = 15, diversity = 0.2, min\_topic\_size = 10). The top\_n\_words and min\_topic\_size parameters ensured that a representative amount of coherent topics was created and the creation of a few big or many small clusters was avoided. The choice of the diversity parameter ensured the topic representations to be sufficiently diverse while still preserving coherence.

#### 4.2 Climate Compounds as Framing Devices

As mentioned in Section 2.3.3, existing research in the field of corpus linguisticbased lexical compound analysis in climate change discourse was mostly conducted on English-language texts from the public sphere. In this study, research paradigms were modified and applied to German-language election manifestos in order to analyze framing in climate change discourse in the German policymaking community. For this analysis, all election manifestos published by the *Die Grünen*, SPD and CDU since 1980 were considered.

The initial approach used for extracting climate compounds was based on partial string matching. Terms created by compounding in German consist of one word without spaces, not multiple words with spaces in between, like in the English language (e.g., *Klimawandel* and *climate change*). For simplicity reasons, it was assumed that every word longer than seven characters containing the substring *klima* (climate) is a potential climate compound. This heuristic allowed for a simple yet effective detection and collection of carbon compounds.

Before	After
Klimarisikoversicherungen	Klimarisikoversicherung
Klimaschutz-Sofortprogramme	Klimaschutz-Sofortprogramm
Klimaverträglichkeitsprüfung	Klimaverträglichkeitsprüfung
Klima-Investition <b>en</b>	Klima-Investition
Klimaanpassungsgesetz <b>es</b>	Klimaanpassungsgesetz

TABLE 4.1: Climate compound dictionary excerpt before and after manual harmonization

As the compounds were extracted from natural language text, the collection contained inflected and not inflected words alike. Declinations of words (e.g.,

*Klimaziele* vs. *Klimazielen*) were detected sometimes, at other times the words were truncated in a way that made them incomprehensible (e.g., *Klimaproblematik* vs. *Klimaproblemat*). This posed a problem requiring manual attention. A computational linguist's method of choice would usually be the automatic reduction of words, i.e. stemming or lemmatization. Unfortunately, the commonly used tools (e.g. Snowball Stemmer, Hanover Tree Tagger), did not perform consistently on the presumably newly emerged climate compounds. Harmonization of the words to their lemma was thus performed semi-automatically, by combining human validation and a dictionary-based approximate string matching heuristics. First, all extracted climate compounds, regardless of year, party, and inflection, were added to a list, which was manually harmonized word for word by a human assessor (see Table 4.1). After this *climate compound dictionary* was created and duplicates were removed, the compound collection process was repeated for each year and party in order to collect harmonized versions of the words.

LISTING 1: Approximate String Matching in Climate Compounds

1	for new_word in found_compounds:
2	# case 1: already in dictionary
3	if new_word is in compound_dictionary:
4	return new_word
5	for dict_word in compound_dictionary
6	#case 2: low edit distance with dictionary word
7	if edit_distance(new_word, dict_word) <= 2:
8	return dict_word
9	<pre>#case 3: no match found; manual inspection</pre>
.0	return "no match found"

The collected compounds for each party and year were iteratively compared to the dictionary entries using approximate string matching. The string metric used was the *edit distance*, which expresses how many operations (i.e., replacing/inserting/deleting letters) would be necessary to transform one string into another one. The lower the edit distance between two words, the more similar they are. When comparing new, possibly inflected, climate compounds and the non-inflected ones in the dictionary, matches with low edit distance (in this case: distance  $\leq$  2) yielded the manually harmonized (i.e., non-inflected version) of the climate compound (see Algorithm 1).

Word	Category
Klimaerhitzung	description
Klimaemissionsziel	measure
Klimaeffizienzreform	measure

TABLE 4.2: Example of climate compounds with categories

After the climate compounds were extracted from the manifestos, they were separated into two categories: *Descriptions* and *Measures* (see Table 4.2). Descriptive compounds simply *describe* the situation or other facts surrounding climate change and do not imply any propositions, such as *Klimaproblem* (climate problem). Compounds categorized as *measures* refer to specific propositions regarding climate change and its mitigation, such as *Klimaschutzabkommen* (climate protection treaty). The words were categorized independently by three human annotators, and the final annotation was determined by majority vote.

```
{"Klimaschutzprogramm": {
1
             "occurence_by_party": {
2
                  "Die Grünen": 3,
3
                  "CDU": 0,
4
                  "SPD": 2
5
              },
6
              "total_occurence": 5,
7
              "first_mentioned": [
8
                  ["SPD", "1994"
9
                  ]],
10
              "category": "actions"
11
         }}
12
```

LISTING 2: Example of JSON entry in climate compound dictionary

Finally, the climate compounds along with collected meta-information were stored in JSON files for every year and every party (see Listing 2 for an example dictionary entry). Using the climate compound dictionary, analyses were performed using text and data mining methods, including computation of the:

- absolute number of compounds
- most frequent compounds
- longitudinal evolution of the TOP 3 most frequent compounds
- first mention of a compound
- longitudinal evolution of category distribution

### **Chapter 5**

# Results

In this chapter, the results of the analysis will be presented. First, the results of the dynamic topic modeling study will be described. Afterwards, the results of the climate compound analysis in German election manifestos will be shown.

### 5.1 Topic Modeling

In the following, the topic modeling results will be presented. This includes the presentation of global topic representations, the dynamic topic modeling results, and a closer examination of environment-related topics.

It is recommended to examine the topic cluster representations (*intertopic dis-tance maps*) and the dynamic topic modeling results interactively. Interactive visualizations are provided in the supplementary material and on GitHub<sup>1</sup>.

Die Grünen	CDU	SPD
Jugendbildung	Politik	Jugendbildung
Politik	EU	Arbeitnehmer
Gleichberechtigung	Familie	Energiewende
EU	Rente	Familie
Familie	Energiewende	Steuern
Energiewende	Jugend	Rente
Verkehrswende	Steuern	Regierung
Einwanderung	Wirtschaft	Innovation
Klimaschutz	Militär	Politik
Landwirtschaft	Digitalisierung	Innovation

TABLE 5.1: Global TOP 10 topics (all parties)

<sup>&</sup>lt;sup>1</sup>https://github.com/julesha/plots\_thesis/tree/main/Topic\_Modeling

The manually labeled topic modeling results for the global TOP 10 topics for each party are presented in Table 5.1. Topics which appeared in the results across all three parties are: *Politics, Family,* and *Energy Transition*.

**Die Grünen.** In the cluster representations of the topics, three distinct clusters were identified (Figure A.1). One cluster consisted of four environmentrelated topics (Energy Transition, Mobility Transition, Climate Protection, Agriculture), the second one comprised three society-related topics (Youth education, Equality, Family) and the last cluster contained three politicsrelated topics (Politics, EU, Immigration). Among the TOP 10 topics computed for the Greens, there were three climate distinct change-related ones, namely *Energy Transition, Mobility Transition* and *Climate Protection*. After 2005, these topics gained relevance in Green Party manifestos, and the topic *Energy Transition* became the most prevalent one of the three (see Figure 5.1a).

**CDU.** The cluster representations of topics generated from CDU manifestos revealed three distinct clusters (Figure A.2). One cluster contained four society-related topics (Family, Pensions, Youth, Military). The second cluster comprised four economy-related topics (Taxes, Digitalization, Market Economy, Energy Transition). The last cluster contained two politics-related topics (EU, Politics). No distinct climate change topic was found. However, the *Energy Transition* topic had mentions of climate-related issues. Dynamic topic modeling revealed that the *Energy Transition* topic has been mentioned infrequently by the CDU until 2009, except for a peak in 1987. The topic gained importance in 2009 and reached its highest frequency in 2021 (see Figure 5.1b).

**SPD.** Two clusters of topics were found in the results for the SPD (Figure A.3). One cluster contained topics concerned with societal issues (Family, Youth Education, Pension, Work), the other one comprised topics with political and economic interest (Politics, Taxes, Government, Innovation, Digitalization, Energy Transition). Similar to the CDU topic modeling results, among the TOP 10 topics extracted for the SPD manifestos, no specific climate change topic was found. However, the *Energy Transition* topic did, like in the case of the CDU, include mentions of general environmental policies. As visualized by the DTM results, the topic has consistently been on the party's

agenda, albeit less frequently so before 2009, with exception of a peak in 1990. The topic was discussed most frequently in 2013 (see Figure 5.1c).

#### 5.2 Climate Compounds as Framing Devices

An analysis of climate compounds in manifestos was performed to determine the level of change in salience in climate discourse, assuming climate compounds to be discursive markers.

First, the general longitudinal development of climate compound usage will be presented, followed by the presentation of party-level insights. Lastly, a closer look at the results related to the Greens' evolution with respect to climate compound usage will be taken.

**General Development.** The evolution of climate compound frequency (aggregated for all parties) is visualized in Figure 5.2. It revealed an upward trend in climate compound usage, with the Green Party consistently using more compounds than the CDU and the SPD. Visually, the graph can be divided into two sections: 1980-2005 and 2009-2021, the first one containing very few and the second one the majority of the compound mentions. Until 2009, climate compounds remained virtually unmentioned by the mainstream parties, while the Greens displayed infrequent but consistent usage starting in 1990. In 2009, the overall frequency increased five-fold compared to 2005, from under 30 to nearly 150 observations. From this point, all parties used climate compounds in their manifestos. In 2021, the most compounds were used (total: 205) and the highest frequency was observed for all parties (Greens: 167; CDU: 68; SPD: 33).

Figure 5.3 visualizes the development of climate compound usage by category for all parties (see Section 4.2 for more information on the categories). Both *Descriptions* and *Measures*, showed a similar development until the year 1998 and consistently ranged below a frequency of 20. From 2002 on, the curves diverged, with descriptive words becoming more frequent. In 2009, a more than five-fold increase in frequency can be observed in the *Descriptions* category, while the *Measures* category merely doubled in frequency. After this, descriptive compounds lost importance until 2017, while measurerelated ones were used more often. In 2021, words of both categories marked



FIGURE 5.1: Dynamic Topic Modeling visualization for environment-related topics (all parties)



#### Frequency of Climate Compounds in Manifestos

FIGURE 5.2: Frequency of climate compounds in German election manifestos (all parties)



Climate Compounds in Manifestos by Category

FIGURE 5.3: Evolution of climate compound usage by category (all parties)

peak frequencies and were used equally often. A more detailed overview of compound usage by category for every party is offered in Appendix A.2.

**Most frequent compounds.** Table 5.2 presents the most frequent climate compounds, considering all parties and years. The most frequent climate compound was *Klimaschutz* (climate protection), occurring more than twice as often as the second most frequent word, *Klimawandel* (climate change). The third most frequent word was *Klimakrise* (climate crisis).

Word	Frequency	Category
Klimaschutz	234	Measure
Klimawandel	92	Description
Klimakrise	44	Description

TABLE 5.2: TOP 3 most frequent climate compounds (all parties)



Top 3 Most Frequent Climate Compounds in Manifestos

FIGURE 5.4: TOP 3 most frequent compounds (all parties)

Figure 5.4 visualizes the evolution of these three most frequent climate compounds from 1980 until 2021. The word *Klimaschutz* first appeared in 1994 and was the only one of the three words mentioned before 2005. From 2002 on, all parties consistently utilized the word. *Klimawandel* was first mentioned in 2005 and showed a similar development to *Klimaschutz* through

28
28
7
6
l 5
]

TABLE 5.3: TOP 5 most frequent compounds (all parties)

2021, albeit with fewer mentions. Curves for both words peaked in 2009, showed a recess in 2013 and 2017 and gained more relevance again in 2021. The graph for *Klimakrise* looks different. The word first appeared in 2009 with a total of two mentions, but quickly gained prominence in the discourse, being used more frequently than the other descriptive word *Klimawandel* in 2017. Among the analyzed words, *Klimakrise* was the only one that shows a steady upward trend in usage. Looking at the distribution of mentions across parties, it becomes evident that the word *Klimakrise* was used exclusively by the Green Party (see Figure 5.5). Visualizations for the frequency of *Klimaschutz* and *Klimawandel* by party can be found in Appendix A.2.



Frequency of "Klimakrise" in Manifestos

FIGURE 5.5: Frequency of Klimakrise (all parties)

Table 5.3 presents the five most frequent words for each party. The Greens used a wide array of different climate compounds, with an even distribution in frequency in the TOP 5 words. The CDU and the SPD predominantly used their three most frequent climate compounds, which happen to be the same for both parties.



#### First Mentions of Climate Compounds in Manifestos by Party

FIGURE 5.6: First mentions of climate compounds (all parties)

Figure 5.6 visualizes the amount of times a party first introduced a previously unseen climate compound to the discourse. The years 1980 until 2005 showed low frequencies of first mentions of new words, changing in 2009 as the overall frequency increased five-fold. Furthermore, peaks in 2009 and 2021 can be observed, with a decline during the two elections in between. The largest number of first mentions is shown by the Greens, who also displayed the largest difference in numbers between two consecutive election programs, introducing seven times more new compounds in 2009 than in 2005. Except in the year 1983, the Green Party introduced new climate compounds in every manifesto, for a total of 96 new words. The CDU was responsible for the introduction of 24 new words, and the SPD was the first party to mention 8 of the climate compounds.

### Chapter 6

# Discussion

This chapter discusses the results presented in Chapter 5. Section 6.1 details the results of the dynamic topic modeling study. Afterwards, the evolution of climate compounds in climate change discourse is interpreted in Section 6.2. Finally, Section 6.3 explores the presented results as they relate to the evolution of the Green Party.

#### 6.1 Topic Modeling

The topic modeling results show pronounced differences in climate change discourse between the Greens and the two mainstream parties, with the discussion of the climate in Green Party manifestos being distinct and multifaceted. Of the TOP 10 topics found, three are directly climate-related, and further mentions of green policies can be found throughout other topics. This division of the Greens' climate-related topics into subtopics with high frequency suggests an elaborate and fine-grained discussion of the issue. Also, the Greens' contribution to the German climate discourse is often stronglyworded, as evidenced by emotional terms such as *catastrophe* or *crisis*, which are frequently found in topic representations. The Greens' primary climaterelated focus areas are the energy transition, the mobility transition, and climate protection. Interestingly, in the DTM, Energy Transition saw the steepest increase in frequency among these climate-related topics. While the topic started off as clearly climate-adjacent in the 1980s - with a focus on producing and saving energy while preserving the environment – by 2022 it bore a closer resemblance to economics, efficiency, and renewables talk. Seeming to bridge economic and environmental issues, the *Energy Transition* topic may be indicative of a change in climate discourse in the Green Party. This is also supported by the thematic shift and increase in the frequency of topics after 2005, the year in which the Greens departed from government.

Looking for clear markers of the evolution of climate discourse in the mainstream parties poses a challenge. While a well-defined subset of environment-related topics can be found for the Green Party, both SPD and CDU lack a distinct climate topic. Instead, results also show the *Energy Transition* topic functions as a kind of *catch-all* category for both parties, containing sentences related to not only energy politics, but also environment- and sustainability-related issues. Despite the topic occurring in all parties, the visualization of the topic clusters further demonstrates the juxtaposition between the mainstream party and Green Party discourse surrounding energy and the climate. While in the case of the Green Party, the Energy Transition topic is clustered with three other environment-adjacent topics, for both the CDU and SPD, in contrast, its nearest neighbors are related to the economy and innovation.

The topic does not occur frequently in both mainstream parties' discourse until 1998. An exception can be observed in the DTM results for the CDU, for whom a peak in topic frequency occured in 1987, with the topic representation containing words such as *Gesundheitsgefahren* (health risks) and *Übergangstechnologie* (transitional technology). Considering the manifesto's publishing date, which was only one year following the 1986 catastrophic Chernobyl nuclear accident, the topic representations could represent a shortlived response to the disaster and the question of the safety of nuclear energy. A similar development can be observed for the SPD, for whom a temporary peak in topic frequency occurred in 1987 and 1990, with topic representations containing various mentions of nuclear power-related issues. Here, the Chernobyl disaster seems to have had a slightly longer-lasting effect on the discourse.

Initially characterized by terms related to the general discussion around nuclear power and its associated environmental and safety concerns, the *Energy Transition* topic representations for the mainstream parties shifted in 1998 toward renewable energies such as solar and wind power, as well as energy efficiency in general. At this time, the economic aspects of the energy transition seem to have been prioritized over its role in climate protection. This changed again in the early 2000s, as terms such as *environment* and *climate protection* reappeared in the topic representations, perhaps due in part to the

SPD-Green government's emphasis on the green aspect of the energy transition. This period also saw a radical shift toward renewables through the adoption of key environment-related legislation, such as the *Renewable Energy Sources Act* in 2000 and the *Nuclear Phase-out Act* in 2002. This shift is consistent with the findings of Müller-Hansen et al. (2021), who saw similar developments in the coal discourse, in which the economic advantages lost importance as climate concerns gained prominence. These findings are in line with Weingart, Engels, and Pansegrau (2000), who already found at the turn of the century that mainstream parties have not prioritized climate-related issues, instead reinterpreting them and addressing their sub-problems that are easier to grasp and solve, such as sustainable development and the energy transition. This molds the *future catastrophe* narrative of climate change, which poses an unknown risk of unknown scope, into a problem that fits into typical political discourse, thus allowing for concrete actions to preclude it.

### 6.2 Climate Compounds as Framing Devices

The results presented in Section 5.2 reveal several interesting developments in climate compound usage in German party manifestos. This section adopts a top-down approach to discussing these results, starting with an exploration of the general developments, followed by a more fine-grained party-level interpretation.

**General.** From a macro perspective, the most evident development in climate compound usage is the sharp increase in their frequency in all parties' manifestos after 2005, indicating an uptick in the prominence of climate change discourse during the legislative period between 2005 and 2009. Before 2009, climate compound use was rare, with their presence mostly limited to Green Party manifestos. However, from 2009 to 2021, a substantial shift occurred, with all parties incorporating the compounds into their vocabularies and the absolute frequency increasing more than threefold. While an increase in the frequency of climate compounds does not irrefutably demonstrate a re-framing of climate change, it suggests that climate discourse has become more important in political discourse in Germany, and this ascent in mainstream politics has come with the emergence of a specific vocabulary. This potentially provides evidence for Koteyko, Thelwall, and Nerlich (2010)'s concept of an emergent *carbon lexicon*, as new concepts and turning points in discourse seem to foster linguistic creativity. The evolution of climate compound usage in manifestos seems to support this theory.

Nerlich, Evans, and Koteyko (2011) also observed an increase in climate compound usage in other domains, with more elaborate words such as *low carbon diet* gaining prominence in 2006. This phenomenon is attributed to the raised (global) awareness of climate change in society due to climate campaigns such as the release of Al Gore's *An Inconvenient Truth* (2006), a documentary film about the dangers of global warming, and one of the first climate change-related pieces of mass media to gain mainstream success. The increase in climate compound usage in German manifestos also correlated with the first commitment period of the Kyoto Protocol, an international climate treaty signed by 192 countries. After being adopted in 1997, the Kyoto Protocol came into force in 2005 with the objective of limiting and reducing worldwide greenhouse gas emissions in order to mitigate anthropogenic climate change (Almer and Winkler, 2017).

Grundmann and Krishnamurthy (2010) showed similar results: analyzing a total of almost 60,000 German news articles from 1980 until 2007 from the Nexis news media archive, they found that more than 50% of all analyzed articles on climate change were published in 2007. The authors also attribute this, among other factors, to the release of *An Inconvenient Truth*, the publishing of the Fourth Assessment report of the IPCC and the award of the Nobel Peace Prize to Al Gore together with the IPCC in the same year. Similar results were published by Schmidt, Ivanova, and Schäfer (2013), who found that media attention to climate change has increased rapidly around 2006/early 2007. As discussed in Section 2.1, politicians must adapt to frames prevalent in media and society, with the shaping of climate change discourse resulting from an interplay of media, civil society, and politics. Fittingly, this described sudden surge in media attention in 2007 corresponded to the sharp increase in climate compound frequency in manifestos before the 2009 German federal elections.

In addition to this first increase in frequency in 2009, a second increase in climate compound usage was observed starting in 2013. The first shift in climate discourse is marked by an increase in climate compounds in the *Descriptions* category in 2009. The second discursive shift comes in the form of a steep increase in compounds describing *Measures* related to climate change

(mitigation) in 2013. Thinking of these shifts as *key junctures*, or transitive periods, of climate change discourse, its evolution could be divided into three segments: a period of little to no attention (1980 – 2009), a period of description, or *talking about the issue* (2009 – 2013) and a period of calling for climate action (2013 – ?). In the 2021 manifestos, compounds of both categories get mentioned equally often again for the first time in more than 20 years. This could mark a new period of climate discourse, potentially uniting description and action.

**Party-Level.** Key differences in climate change discourse in election programs can be observed not only in broad longitudinal terms, but also at the level between parties. Generally, the Green Party's discourse is characterized by a more diverse *compound vocabulary*, meaning they employ a large assortment of compounds, as opposed to frequently repeating a few words. Three of the TOP 5 words used by the party are synonyms of climate change, with a varying sense of urgency: it is framed as climate *change*, *crisis*, and *catastrophe*, suggesting a range of narratives in the discourse. In contrast, the CDU and SPD have a much smaller compound vocabulary, primarily referring to climate change and climate protection, with additional mentions of a *Klimaschutzziel* (climate protection goal). This observation is again consistent with Weingart, Engels, and Pansegrau (2000) that mainstream parties rarely address the climate issue as one big problem, instead breaking it down into solvable sub-problems and thus creating reachable *goals*.

Even with mainstream parties' comparatively simple vocabulary, the upward trend of climate compound usage has given rise to an evolving lexicon, which may provide insights into the framing of climate change in political discourse. A starting point for this is the discussion of the most frequent climate compounds. Indeed, the three overall most frequent words – *Klimaschutz* (climate protection), *Klimawandel* (climate change), and *Klimakrise* (climate crisis) – have key semantic differences between them. The most frequent word, *Klimaschutz*, generally relates to measures to counteract climate change and its associated consequences. Additionally, it is the only one of the three that occurred in election programs before the beginning of the 21<sup>st</sup> century.

While the second and third most frequent terms, Klimawandel and Klimakrise,

both describe the same concept of climate change, they have disparate connotations. *Klimawandel* is an objective description of climate change that makes no statement about the extent or implications of the phenomenon. Meanwhile, *Klimakrise* adds a sense of urgency and an implicit call for action, as it is no longer characterized as simply a *change* in the climate, but rather a *crisis* that must be addressed. The component *crisis* frames climate change as an imminent threat and stresses the gravity of its effects, as well as its implications for the future of humanity. The emotionally charged term *Klimakrise* has quickly established itself as a key component of climate discourse in politics: after not being mentioned in any election programs before 2009, by 2017 – just two elections later – its frequency surpassed that of its neutral counterpart *Klimawandel*.

#### 6.3 Green Party Evolution

The topic modeling results reveal some of the underlying structures of party level discourse in election manifestos. Based on the evaluation of only relatively few documents of three parties, no general conclusion can be made, and it is wiser to speak about *trends* rather than to make statements about specific *party profiles*. However, distinct differences in discourse across parties can be observed, starting off with the apparent commonalities and the three shared TOP 10 topics (Politics, Energy Transition, Family). While these topics seem like essential building blocks of established parties, three of them (EU, Energy Transition, Family) only became relevant in Green Party manifestos in the late 1990s – around the time the Greens were part of a government coalition with the social democrats for the first time. Especially after this joint government ended in 2005, these issues were addressed much more frequently in Green Party programs, suggesting that government participation brought the party closer to *mainstream* discourse.

Out of the four common topics, the *Politics* one has been part of Green manifestos since the party's establishment. While the topic was initially summarized by critique of weapons of mass destruction and nuclear power and words like *Widerstand* (resistance), the topic representations contained more concrete, policy-oriented concepts such as *Gesellschaftsvertrag* (Articles of association) and *Bürgerversicherung* (single-payer health care) after the end of

the Red-Green coalition in 2005. During this time, the Greens faced previously unknown challenges. The depression that followed government participation, fueled by an uncertain standing in the German party system and poor election results, required what Blühdorn (2009) called strategic repositioning or reinventing green politics for the contemporary era. This shift in the very general *Politics* topic is a plausible manifestation of this repositioning – it can be interpreted as an example of the rearrangement of the party's priorities to facilitate government participation. In a way, the former grassroots party grew up and exchanged its past fundamentalism for pragmatism, in an attempt to secure its position in mainstream politics. The Greens coined the term *radical realism* to describe this position, which moves closer to the political center (Blühdorn, 2009). This radical realism has opened doors for new policies, but also kept space for the Greens' principal message, the environment. Despite the ideological and strategic evolution the party underwent, the Greens have consistently led the climate discourse in the German political landscape.

Looking at the results of the climate compound analysis in the Green Party's manifestos, more interesting developments are revealed. As discussed in Section 2.2.3, framing allows actors to claim (and keep) ownership of an issue. Keeping this in mind, the Greens' usage of climate compounds as framing devices plays an important role in the discourse evolution. The party introduced the most new climate compounds into the discourse, with some of them also being adopted by other parties (e.g., *Klimaabkommen, Klimaschutzgesetz*) and some others being exclusive to the party's vocabulary (e.g., *Klimakrise, Klimakiller*).

Focusing on the words unique to the Greens, the term *Klimakrise* (climate crisis) deserves special consideration. Despite being first used as late as 2009, it is the overall second most frequent climate compound used by the party, more often than *climate change*. Unlike the latter, it is emotionally charged and describes the concept of climate change as an imminent threat, urgently calling for action. The party almost ceased using the more neutral word *Klimawandel* (climate change) by 2017, while its drastic synonym gained frequency rapidly, suggesting that a change in vocabulary has taken place. This re-framing of the climate issue from a mere *change* to an outright *crisis* is radical in comparison to the wording of the mainstream parties, which refrain from using emotionally charged vocabulary. It is, however, not so radical

when seen in context with the zeitgeist of the 21<sup>st</sup> century. Politics is a reflection of the state of contemporary society. Major news outlets, such as the British daily newspaper The Guardian, have changed their language conventions in 2019, no longer talking about global warming or climate change, but instead referring to the phenomenon as a climate crisis or emergency (Carrington, 2019). Similar developments were observed in other domains: the Oxford University Press declared *climate emergency* the Word of the Year 2019, noting an increase in its use of 10,796%, compared with the previous year. The word *climate crisis* was also in the Word of the Year shortlist and saw a 2,510% increase in usage (Oxford University Press, 2019). The trend observed in the public, in science and in media seems to be picked up by the Green Party but not by the mainstream parties. It seems that, despite converging to the political center and mainstream politics, the politicization of climate change and the moralistic aspect of addressing the problem is still the responsibility of the Green Party.

### Chapter 7

## **Conclusions & Future Research**

This thesis examined the evolution of climate change discourse in German party manifestos, with a focus on the development of the Green Party. To bridge the gap between computational linguistics and political science, stateof-the-art neural topic modeling methods were employed to conduct a longitudinal discourse analysis on German-language election programs. To this end, climate compounds proved to be effective markers for the evolution of climate change framing in German party discourse. The Manifesto Corpus has proven to be a valuable resource in this context for party-level climate discourse analysis.

Using this corpus, German parties' evolving positions and engagement with climate-related topics were quantified and analyzed, allowing for the exploration of topical shifts and changes in climate compound usage. Generally, the findings suggest that climate discourse has gained importance in the German political landscape, indicating a *greening* of the German party system.

While the established parties tend to focus on *sustainable development*, which primarily addresses the economic aspects of climate change and its mitigation, the Greens have historically focused on environmental aspects of climate change. However, as discourse on sustainability has become increasingly mainstream, the Greens have similarly seen a mainstreaming of their topics, shifting away from niche issues and toward the CDU and SPD's economic considerations of environmentalism.

These results suggest that over the last 40 years, the Greens have gradually departed from their radical grassroots past, becoming firmly integrated into the German multi-party system. While the Greens have remained true to their core values, adamantly addressing the imminent threat posed by the climate crisis, their position seems to have shifted from eco-radicalism toward

a more agreeable pragmatism, prioritizing the desire to govern and keeping coalition options to all sides open. Although the contents of discourse differ, similar patterns in attention shifts can be observed among all parties. The climate-aware zeitgeist in society and the media seems to impact climate change discourse, regardless of a party's position on a scale of popularity or left-to-right ideology.

Despite the promising results presented in this thesis, it should be kept in mind that they do not represent a full picture of climate change discourse in German political parties, but merely approximate the longitudinal development based on selected criteria. While the results suggest a successful application of NLP techniques for addressing the problem of discourse analysis in the German political sphere, they also highlight some aspects that could be addressed in future research.

Various approaches can be taken here. Concerning the data, it could be of interest to take manifestos of all German parties into consideration. In particular, the rise of right-wing populism and the associated anti-climate sentiment introduced by the *Alternative für Deutschland* could yield interesting insights to the evolving discourse in the German parliament. In addition to considering more parties, research could also include texts from different sources. As manifestos are only published every four years, they leave large explanatory gaps. Including other textual data, such as transcribed parliamentary speeches or coalition agreements, in the analyses could help understand the process and pace of discourse shifts, such as the large shift in issue attention between 2005 and 2009 reported in this study.

The topic modeling task could be augmented by employing different methods, such as guided topic modeling, if specific aspects of the climate discourse are to be investigated. This way, topics could be created around a set of predefined words. Establishing a strategic set of seed words could provide insight into specific aspects of the discourse, such as the set *coal*, *sustainability*, *solar* for the examination of energy transition discourse. In addition, performing dynamic topic modeling on a subcorpus of only environment-related sentences could contribute to understanding facets of climate change discourse. Utilizing the Manifesto Corpus codes for sustainable development and environmental protection (416.1, 416.2 and 501), it would be possible to extract such sentences from election manifestos. However, since sentences with these codes only make up a fraction of a manifesto, it would be necessary to either include multiple countries in the analysis or manually code additional documents for a country. If a sufficiently large subcorpus were to emerge, the climate discourse could be modeled in detail.

As for the climate compound research, a more fine-grained analysis of the words could be considered. In this study, the compounds were classified in two categories, *Descriptions* and *Actions*. However, the categories could be altered or augmented based on the researcher's needs, for example by adding the sentiment valence of the words. As the *climate compound dictionary* stands for a novel approach to climate discourse analysis and is modular in nature, it potentially represents a valuable resource for researchers. Adding more meta-information could increase its usefulness. For example, it might be of interest to further investigate the origin and common usage of words. Indication of whether a word has a dictionary entry or in which domain it is predominantly used may provide valuable insights for discourse and framing analysis.

This outlook once again emphasizes that interdisciplinary collaboration among researchers is critical to addressing the new challenges arising in this era of uncertainty. Combining a computational linguistic approach with a political science investigation, this work represents a potential starting point and new perspective for researchers at a time when understanding climate discourse is of greater importance than ever before. The insights presented in this thesis help to model and understand policymakers' portrayals of climate change – an essential prerequisite for communicating and mitigating the issue.

As of November 2022, new developments in global communications are already emerging as war and recession dominate the discourse and divert focus from environmental issues. The development of the German climate discourse could also take an interesting turn in the near future. In the face of the 2022 energy crisis, the Greens have moved further away from their roots and are arguing for a postponement of the nuclear phase-out beyond the legally mandated date. This development could impact the public perception of the Green Party and thus the climate discourse, and is therefore of great interest to researchers.

# Bibliography

- Abercrombie, Gavin and Riza Batista-Navarro (2020). "Sentiment and position-taking analysis of parliamentary debates: a systematic literature review". In: *Journal of Computational Social Science* 3.1, pp. 245–270. DOI: 10.1007/s42001-019-00060-w.
- Aggarwal, Charu C., Alexander Hinneburg, and Daniel A. Keim (2001).
  "On the Surprising Behavior of Distance Metrics in High Dimensional Space". In: *Database Theory ICDT 2001*. Ed. by Gerhard Goos et al. Vol. 1973. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 420–434. DOI: 10.1007/3-540-44503-X\_27.
- Almer, Christian and Ralph Winkler (2017). "Analyzing the effectiveness of international environmental policies: The case of the Kyoto Protocol". In: *Journal of Environmental Economics and Management* 82, pp. 125–151. DOI: 10.1016/j.jeem.2016.11.003.
- Anderson, Karen M. (2015). "The politics of incremental change: institutional change in old-age pensions and health care in Germany". In: *Journal for Labour Market Research* 48.2, pp. 113–131. DOI: 10.1007/s12651-015-0180-y.
- Anshelm, Jonas and Martin Hultman (2015). *Discourses of global climate change: apocalyptic framing and political antagonisms*. Routledge studies in environmental communication and media. Abingdon, Oxon ; New York, NY: Routledge. DOI: 10.4324/9781315769998.
- Blei, David M. and John D. Lafferty (2006). "Dynamic topic models". In: Proceedings of the 23rd international conference on Machine learning - ICML '06. Pittsburgh, Pennsylvania: ACM Press, pp. 113–120. DOI: 10.1145/ 1143844.1143859.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent dirichlet allocation". In: *The Journal of Machine Learning Research* 3, pp. 993–1022.

- Blühdorn, Ingolfur (2009). "Reinventing Green Politics: On the Strategic Repositioning of the German Green Party". In: *German Politics* 18.1, pp. 36–54. DOI: 10.1080/09644000802649189.
- Booij, Geert E. (2011). "Compounding and Construction Morphology". In: *The Oxford Handbook of Compounding*. Oxford University Press. DOI: 10. 1093/oxfordhb/9780199695720.013.0010.
- Bouma, Gerlof (2009). "Normalized (Pointwise) Mutual Information in Collocation Extraction". In: *From Form to Meaning: Processing Texts Automatically, Proceedings of the Biennial GSCL Conference 2009,* pp. 31–40.
- Buyalskaya, Anastasia, Marcos Gallo, and Colin F. Camerer (2021). "The golden age of social science". In: *Proceedings of the National Academy of Sciences* 118.5. DOI: 10.1073/pnas.2002923118.
- Campello, Ricardo J. G. B., Davoud Moulavi, and Joerg Sander (2013).
   "Density-Based Clustering Based on Hierarchical Density Estimates". In: *Advances in Knowledge Discovery and Data Mining*. Ed. by Jian Pei et al. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 160–172. DOI: 10.1007/978-3-642-37456-2\_14.
- Carrington, Damian (2019). Why the Guardian is changing the language it uses about the environment. https://www.theguardian.com/environment/ 2019/may/17/why-the-guardian-is-changing-the-language-ituses-about-the-environment. Accessed: 2022-09-23.
- Carter, Neil (2013). "Greening the mainstream: party politics and the environment". In: *Environmental Politics* 22.1, pp. 73–94. DOI: 10.1080/09644016. 2013.755391.
- Chatsiou, Kakia (2020). "Text Classification of Manifestos and COVID-19 Press Briefings using BERT and Convolutional Neural Networks". In: DOI: 10.48550/arXiv.2010.10267. arXiv: 2010.10267.
- Denny, Matthew J. and Arthur Spirling (2018). "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It". In: *Political Analysis* 26.2, pp. 168–189. DOI: 10.1017/pan.2017.
  44.
- Devlin, Jacob et al. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational*

*Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).* Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

- Dieng, Adji B., Francisco J. R. Ruiz, and David M. Blei (2020). "Topic Modeling in Embedding Spaces". In: *Transactions of the Association for Computational Linguistics* 8, pp. 439–453. DOI: 10.1162/tacl\_a\_00325.
- Ding, C. et al. (2002). "Adaptive dimension reduction for clustering high dimensional data". In: 2002 IEEE International Conference on Data Mining, 2002. Proceedings. Maebashi City, Japan: IEEE Comput. Soc, pp. 147–154. DOI: 10.1109/ICDM.2002.1183897.
- Egger, Roman and Joanne Yu (2022). "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts". In: *Frontiers in Sociology* 7, p. 886498. DOI: 10.3389/fsoc.2022.886498.
- Farstad, Fay M (2018). "What explains variation in parties' climate change salience?" In: *Party Politics* 24.6. Publisher: SAGE Publications Ltd, pp. 698–707. DOI: 10.1177/1354068817693473.
- Friedman, Sharon M., Carole M. Gorney, and Brenda P. Egolf (1992). "Chernobyl coverage: how the US media treated the nuclear industry". In: *Public Understanding of Science* 1.3, pp. 305–323. DOI: 10.1088/0963-6625/1/ 3/005.
- Grant, Zack P. and James Tilley (2019). "Fertile soil: explaining variation in the success of Green parties". In: *West European Politics* 42.3, pp. 495–516. DOI: 10.1080/01402382.2018.1521673.
- Green-Pedersen, Christoffer and Peter B. Mortensen (2010). "Who sets the agenda and who responds to it in the Danish parliament? A new model of issue competition and agenda-setting". In: *European Journal of Political Research* 49.2, pp. 257–281. DOI: 10.1111/j.1475-6765.2009.01897.x.
- Grimmer, Justin and Brandon M. Stewart (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". In: *Political Analysis* 21.3. Publisher: Cambridge University Press, pp. 267– 297. DOI: 10.1093/pan/mps028.
- Grootendorst, Maarten (2022). "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". In: *arXiv:2203.05794* [*cs*]. arXiv: 2203.05794. DOI: 10.48550/arXiv.2203.05794.

- Grundmann, Reiner and Ramesh Krishnamurthy (2010). "The discourse of climate change: a corpus-based approach". In: *Critical Approaches to Discourse Analysis across Disciplines* 4(2), pp. 125–146.
- Gärtner, Lea and Harald Schoen (2021). "Experiencing climate change: revisiting the role of local weather in affecting climate change awareness and related policy preferences". In: *Climatic Change* 167.3, p. 31. DOI: 10.1007/ s10584-021-03176-z.
- Hajer, Maarten and Wytske Versteeg (2005). "A decade of discourse analysis of environmental politics: Achievements, challenges, perspectives".
  In: *Journal of Environmental Policy & Planning* 7.3, pp. 175–184. DOI: 10. 1080/15239080500339646.
- Kluyver, Thomas et al. (2016). "Jupyter Notebooks a publishing format for reproducible computational workflows". In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by F. Loizides and B. Schmidt. IOS Press, pp. 87–90. DOI: 10.3233/978-1-61499-649-1-87.
- Koteyko, Nelya, Mike Thelwall, and Brigitte Nerlich (2010). "From Carbon Markets to Carbon Morality: Creative Compounds as Framing Devices in Online Discourses on Climate Change Mitigation". In: *Science Communication* 32.1, pp. 25–54. DOI: 10.1177/1075547009340421.
- Kusner, Matt et al. (2015). "From Word Embeddings To Document Distances". In: Proceedings of the 32nd International Conference on Machine Learning. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, pp. 957–966.
- Kwidziński, Emil (2020). "German Green Party: the evolution of political agenda". In: *Journal of Geography, Politics and Society* 10.2. Number: 2, pp. 45–51. DOI: 10.26881/jpgs.2020.2.06.
- Lazer, David et al. (2009). "Computational Social Science". In: Science 323.5915, pp. 721–723. DOI: 10.1126/science.1167742.
- Lehmann, Pola et al. (2022a). *Manifesto Project Dataset*. Version Number: 2022a Type: dataset. DOI: 10.25522/MANIFESTO.MPDS.2022A.
- Lehmann, Pola et al. (2022b). *The Manifesto Project Dataset Codebook. Manifesto Project (MRG / CMP / MARPOR)*. Version 2022a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB). Berlin.

- Lewandowski, Jirka, Nicolas Merz, and Sven Regel (2020). *manifestoR: Access and Process Data and Documents of the Manifesto Project*. R package version 1.4.0.
- Li, Nan and Leona Yi-Fan Su (2018). "Message Framing and Climate Change Communication: A Meta-Analytical Review". In: *Journal of Applied Communications* 102.3. DOI: 10.4148/1051-0834.2189.
- Lieber, Rochelle (2010). *Introducing morphology*. Cambridge: Cambridge University Press. DOI: 10.1017/CB09780511808845.
- Maier, Daniel et al. (2018). "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology". In: *Communication Methods and Measures* 12.2-3, pp. 93–118. DOI: 10.1080/19312458. 2018.1430754.
- Marathe, Megh and Kentaro Toyama (2018). "Semi-Automated Coding for Qualitative Research: A User-Centered Inquiry and Initial Prototypes". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, pp. 1–12. DOI: 10.1145/3173574. 3173922.
- McInnes, Leland et al. (2018). "UMAP: Uniform Manifold Approximation and Projection". In: *Journal of Open Source Software* 3.29, p. 861. DOI: 10. 21105/joss.00861.
- Merz, Nicolas, Sven Regel, and Jirka Lewandowski (2016). "The Manifesto Corpus: a new resource for research on political parties and quantitative text analysis". In: *Research and Politics* 3.2, pp. 1–8. DOI: 10.1177/ 2053168016643346.
- Mikhaylov, Slava, Michael Laver, and Kenneth R. Benoit (2012). "Coder Reliability and Misclassification in the Human Coding of Party Manifestos". In: *Political Analysis* 20.1, pp. 78–91. DOI: 10.1093/pan/mpr047.
- Milder, Stephen and Konrad H. Jarausch (2015). "Renewing Democracy: The Rise of Green Politics in West Germany". In: *German Politics and Society* 33.4, pp. 3–24. DOI: 10.3167/gps.2015.330402.
- Müller-Hansen, Finn et al. (2021). "Who cares about coal? Analyzing 70 years of German parliamentary debates on coal with dynamic topic modeling".
  In: *Energy Research & Social Science* 72, 101869. DOI: 10.1016/j.erss. 2020.101869.

- Nerlich, Brigitte, Vyvyan Evans, and Nelya Koteyko (2011). "*Low carbon diet* : Reducing the complexities of climate change to human scale". In: *Language and Cognition* 3.1, pp. 45–82. DOI: 10.1515/langcog.2011.003.
- Nerlich, Brigitte and Nelya Koteyko (2009). "Compounds, creativity and complexity in climate change communication: The case of 'carbon indulgences'". In: *Global Environmental Change* 19.3, pp. 345–353. DOI: 10.1016/ j.gloenvcha.2009.03.001.
- Nerlich, Brigitte, Nelya Koteyko, and Brian Brown (2010). "Theory and language of climate change communication". In: *WIREs Climate Change* 1.1, pp. 97–110. DOI: 10.1002/wcc.2.
- Nisbet, Matthew C. (2009). "Communicating Climate Change: Why Frames Matter for Public Engagement". In: *Environment: Science and Policy for Sustainable Development* 51.2, pp. 12–23. DOI: 10.3200/ENVT.51.2.12-23.
- Olbrich, Eckehard and Sven Banisch (2021). "The Rise of Populism and the Reconfiguration of the German Political Space". In: *Frontiers in Big Data* 4, p. 731349. DOI: 10.3389/fdata.2021.731349.
- Oxford University Press (2019). Oxford Word of the Year 2019 | Oxford Languages. https://languages.oup.com/word-of-the-year/2019/. Accessed: 2022-09-19.
- Pan, Zhongdang and Gerald Kosicki (1993). "Framing analysis: An approach to news discourse". In: *Political Communication* 10.1, pp. 55–75. DOI: 10. 1080/10584609.1993.9962963.
- Penz, Hermine (2022). "Communicating climate change: how (not) to touch a cord with people and promote action". In: *Text & Talk* 42.4, pp. 571–590. DOI: 10.1515/text-2020-0081.
- Ramage, Daniel et al. (2009). Topic Modeling for the Social Sciences. In: Workshop on Applications for Topic Models, NIPS. https://nlp.stanford.edu/pubs/ tmt-nips09.pdf. Accessed: 2022-09-21.
- Reimers, Nils and Iryna Gurevych (2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, pp. 3982–3992. DOI: 10.18653/v1/D19-1410.

- Rogers, Anna, Olga Kovaleva, and Anna Rumshisky (2020). "A Primer in BERTology: What We Know About How BERT Works". In: *Transactions of the Association for Computational Linguistics* 8, pp. 842–866. DOI: 10.1162/ tacl\_a\_00349.
- Saalfeld, Thomas (2002). "The German Party System: Continuity and Change". In: *German Politics* 11.3, pp. 99–130. DOI: 10.1080/714001303.
- Sander, Hendrik (2016). Die Klimagerechtigkeits-Bewegung in Deutschland. Rosa-Luxemburg-Stiftung. URN: https://nbn-resolving.de/urn:nbn: de:kobv:109-1-9008546.
- Schmidt, Andreas, Ana Ivanova, and Mike S. Schäfer (2013). "Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries". In: *Global Environmental Change* 23.5, pp. 1233–1248. DOI: 10.1016/j.gloenvcha.2013.07.020.
- Slapin, Jonathan B. and Sven-Oliver Proksch (2008). "A Scaling Model for Estimating Time-Series Party Positions from Texts". In: *American Journal* of Political Science 52.3, pp. 705–722. DOI: 10.1111/j.1540-5907.2008. 00338.x.
- Sloat, Amanda (2020). Germany's new centrists? The evolution, political prospects, and foreign policy of Germany's Green Party. https://www. brookings.edu/research/germanys-new-centrists-the-evolutionpolitical-prospects-and-foreign-policy-of-germanys-greenparty/. Accessed: 2022-06-17.
- Spoon, Jae-Jae, Sara B. Hobolt, and Catherine E. de Vries (2014). "Going green: Explaining issue competition on the environment: Going green: Explaining issue competition on the environment". In: *European Journal* of Political Research 53.2, pp. 363–380. DOI: 10.1111/1475-6765.12032.
- Stecula, Dominik A. and Eric Merkley (2019). "Framing Climate Change: Economics, Ideology, and Uncertainty in American News Media Content From 1988 to 2014". In: *Frontiers in Communication* 4. DOI: 10.3389/fcomm. 2019.00006.
- Stede, Manfred and Ronny Patz (2021). "The Climate Change Debate and Natural Language Processing". In: *Proceedings of the 1st Workshop on NLP for Positive Impact*. Online: Association for Computational Linguistics, pp. 8–18. DOI: 10.18653/v1/2021.nlp4posimpact-1.2.

- Syed, Shaheen and Marco Spruit (2017). "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation". In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Tokyo, Japan: IEEE, pp. 165–174. DOI: 10.1109/DSAA.2017.61.
- Weingart, Peter, Anita Engels, and Petra Pansegrau (2000). "Risks of communication: discourses on climate change in science, politics, and the mass media". In: *Public Understanding of Science* 9.3, pp. 261–283. DOI: 10.1088/ 0963-6625/9/3/304.
- Wilkerson, John and Andreu Casas (2017). "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges". In: Annual Review of Political Science 20.1, pp. 529–544. DOI: 10.1146 / annurev polisci-052615-025542.
- Wittneben, Bettina B.F. (2012). "The impact of the Fukushima nuclear accident on European energy policy". In: *Environmental Science & Policy* 15.1, pp. 1–3. DOI: 10.1016/j.envsci.2011.09.002.
- Zirn, Cäcilia and Heiner Stuckenschmidt (2014). "Multidimensional topic analysis in political texts". In: *Data & Knowledge Engineering* 90, pp. 38– 53. DOI: 10.1016/j.datak.2013.07.003.

# Appendix A

# Figures

### A.1 Figures for Topic Modeling







FIGURE A.2: Hierarchical clustering of topics (CDU)



FIGURE A.3: Hierarchical clustering of topics (SPD)

### A.2 Figures for Climate Compounds as Framing Devices



FIGURE A.4: Frequency of Klimaschutz (all parties)



Frequency of "Klimawandel" in Manifestos

FIGURE A.5: Frequency of Klimawandel (all parties)



Climate Compounds by Category for Die Grünen

FIGURE A.6: Evolution of climate compound usage by category (Die Grünen)



FIGURE A.7: Evolution of climate compound usage by category (CDU)





FIGURE A.8: Evolution of climate compound usage by category (SPD)

# Eidesstattliche Erklärung

Hiermit versichere ich, Juliane HANEL, dass ich die Masterarbeit mit dem Titel "Transformer-Based Analysis of Climate Discourse and Green Party Evolution in German Party Manifestos" selbständig verfasst habe und dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt wurden. Diese Aussage trifft auch für alle Implementierungen und Dokumentationen im Rahmen dieses Projektes zu.

Berlin, den 02.12.2022