Exploring the application of a sentiment classifiers based on Bidirectional Encoder Representations from Transformers (BERT) in a political context.

Student details

Name: N.E.F. Tettero
Student number: u399961

Word count: 8168

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee

Supervisor: dr. S. Kempeneer
Second reader: dr. E Vanmassenhove

Tilburg University
School of Humanities & Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands

January 23, 2023

Acknowledgements

While writing this thesis, several people have guided me and pointed me in the right direction at times I lost track of the end goal. To them I want to express my gratitude. First of all, I would like to thank my supervisor Shirley Kempeneer for her support and guidance throughout this project. Also, thank you, Eva Vanmassenhove to be my second reader and providing some valuable tips to finish this project. Last, but definitely not least, I would like to thank Frederik Heylen of Datamarinier for teaching me a ton of new practical data science skills that have been applied for writing this thesis.

Table of Contents

Glossary

Concept	Full Name	Description
NLP	Natural Language Processing	
BERT	Bidirectional Encoder Representation from Transformers	An open source machine learning framework for NLP.
MLM	Masked Language Modelling	A way to perform word predictions, used for pre-training BERT.
NSP	Next Sentence Prediction	Prediction of the sentence after the current sentence, used for pre-training BERT.
BERTje	-	A Dutch pre-trained BERT model.
RobBERT	-	A Dutch pre-trained RoBERTa based model.
MP	Member of Parliament	-
Tidyverse	-	A collection of R packages designed for data science.
Flempar	-	R package built for querying the web API of the Flemish Parliament.
get_work()	-	Function of the Flempar R packages for retrieving documents from the database of the Flemish Parliament.

get_mp()	-	Function of the Flempar R packages for retrieving information on MPs.
IAA	Inter-Annotator Agreement	Lorum ipsum

Exploring the application of a sentiment classifiers based on Bidirectional Encoder Representations from Transformers (BERT) in a political context.

N.E.F. Tettero

Abstract

Open government data provides an immense amount of new data sources. Debate transcripts are an example of open government data that, due to their size, are very hard to fathom manually. Natural language processing (NLP) could provide a solution to gain new insights. This study aims to examine the oral questions and interpellations related to climate in the plenary sessions of the Flemish Parliament. The Flemish Parliament's data is accessed through an API which is queried using the Flempar R package, specifically created to query this API easily. The goals of this study are: 1) to test the functionalities of the Flempar package and assess whether the data that can be retrieved from the Flemish Parliament's API can be used to conduct valuable analysis, 2) to test which finetuned BERT-based sentiment analysis model from Hugging Face performs best on political transcripts, and 3) to evaluate the outcomes of the sentiment analysis by looking at how the sentiment of climate related oral questions and interpellations has changed within the Flemish Parliament over the last two decades. The study concludes that: 1) using the Flempar package it is possible to collect Flemish Parliament data that, after some pre-processing steps is suitable to perform valuable analyses on, 2) of the three sentiment analysis models evaluated, a multilingual BERT algorithm finetuned on product reviews performs best, and 3) the sentiment has become more positive over time. Differences are observed between parties. These differences can partly be explained by whether parties are part of the coalition or the opposition.

Keywords: sentiment analysis, parliamentary debates, Flempar

Data Source and Ethics

Work on this thesis did not involve collecting data from human participants or animals. I acknowledge that I do not have any legal claim to the data used in this study and that the original owner of the data and code used in this thesis, retains ownership of the data and code during and after the completion of this thesis. All Python and R codes that are used in this thesis are publicly available and can be obtained from the following repository: https://github.com/ntettero/thesis. The repository also includes a Dockerfile that includes all software and packages necessary for the data collection and analysis of this thesis.

Introduction

Technological advancements have made it possible to collect an ever-increasing amount of data and put this data to use to ease our lives and help us take decisions. To this, governments are no exception. Governments increasingly rely on data for decision-making and creating a new policy. Although the government, in essence, works for its citizens, initiatives to open up this data started less than two decades ago and research within the field of open government data only began to rise around 2011 (Attard et al., 2015). Opening government data is done for three main reasons; *transparency, releasing social and commercial value*, and *participatory governance* (*Open Government Data - OECD*, n.d.). In 2019, the Open Data Directive entered into force. With this directive, the European Union aims to increase government transparency across all member states (*European Legislation on Open Data*, 2019).

The Flemish Parliament's database holds all data debates, legislative proposals, and related documents. This data is accessible through an API. However, collecting the data through the API is no simple task. To ease this process, the company Datamarinier has built an R package called Flempar[1], an interface to the API of the Flemish Parliament. This study employs this novel package to collect a large amount of parliamentary data.

As the ability to access and collect parliamentary data does not necessarily increase transparency, further analysis is necessary to determine to what extent the data is valuable and can be used to achieve more transparency. More specifically, this study will use sentiment analysis to gain insight into how the tone of the parliamentary debates has developed over time.

[1] An R package designed to query the Flemish Parliaments API. All information regarding Flempar functionalities are described in the following blogpost: https://www.Flempar.be/.

Analysing the sentiment over time could help to assess the political polarization that has been rising in Europe for the last three decades (Casal Bértoa & Rama, 2021).

Over the last two decades, European countries have faced a variety of crises. From the ongoing climate crisis, the economic crisis in 2008, and COVID-19 to, most recently, the Russo-Ukrainian War and the subsequent energy crisis. Crises result in strong opposing opinions. Although democracy is centred on institutionally accommodating and strongly opposing social and political coalitions, literature has long commented on the indispensable role of the basic social consensus (Vachudova, 2019, p. 690). The findings of Dunlap et al. (2016) show a lack of social consensus and an increase in partisan polarization on climate change.

Examining the sentiment towards all these crises reaches beyond a single study. This study will focus on the climate crisis because this is an ongoing and progressing crisis, and debates concerning this crisis reach back beyond a decade. By performing sentiment analysis on transcripts of the oral questions and interpellations in the plenary sessions within the Flemish Parliament over the last 20 years, this study aims to detect changes in the sentiment towards climate change in the political arena. Based on previous findings, it is expected that the sentiment towards climate changes differs between political parties, but overall, the sentiment has become more negative over time (Dunlap et al., 2016).

Initially, sentiment analysis has been applied to web-scraped texts, such as tweets and product reviews. However, the implementation reaches beyond social media and sentiment analysis can be applied to any text, including the increasing amount of publicly available government data. As a result, sentiment analysis of parliamentary debates has attracted attention from researchers with backgrounds in computer science and political and social sciences (Abercrombie & Batiste-Navarro, 2020b). While both areas could benefit from each other's expertise, their focus is different, and the crossover is limited. Social and political scientists with proper context knowledge may lack the capabilities to collect open data and apply state-of-the-art technologies. This becomes apparent from the main focus of the studies that have fine-tuned BERT models for sentiment analysis in a political context. These studies have been fixated on the performance of the models and less on the actual implication of the findings (Catelli, 2022; Abercrombie & Batista-Navarro, 2020a). Therefore, research that draws conclusions on the result of sentiment analysis on political data is scarce, while Abercrombie & Batista-Navarro (2020b) stress the possibilities for research into the language changes in the political debate over time.

Additionally, a common flaw of research focused on training natural language processing models on political debate transcripts is the available data size (Thomas et al., 2006; Abercrombie & Batista-Navarro, 2018b). This study aims to add to the literature: 1) a large dataset consisting of the oral questions and interpellations in the Flemish Parliament over the last 20 years, and 2) exploring to what extent BERT models that have been fine-tuned for sentiment analysis can help to gain insights into the political debate over time.

The performance of three sentiment analysis models from Hugging Face will be compared. Afterwards, the best-performing model is used to analyse the sentiment of all oral questions and interpellations related to climate between 01-01-2000 and 01-11-2022. It is expected that sentiment has fluctuated over time but has increasingly become more negative. The findings of this study could help to hold parliamentarians accountable. Spikes in negative sentiment could indicate short-term panic instead of the long-term vision needed to build a sustainable climate policy. Furthermore, these insights increase transparency, one of the main goals of open data. Evaluating to what extent government data is open and transparent might address points of improvement from a government perspective. Also, being able to analyse what is done and said in the political arena first-hand, might decrease the amount of bias created by media that filter news without providing the complete picture. As a result, the application of new techniques could increase citizen involvement.

Research Goals

The goal of this research is threefold. First, this study aims to investigate how one could extract open government data from the Flemish Parliament's API in a way that makes it possible to conduct valuable analyses. Secondly, this study aims to investigate what sentiment analysis model from Hugging Face most accurately predicts the sentiment in parliamentary transcripts. The goal is to find a sentiment analysis algorithm that performs well enough so that conclusion can be drawn from the output. Lastly, this study will use the output of the sentiment analysis to answer several questions that focus more on the political science implications of performing a sentiment analysis on debates. The focus will be on oral questions and interpellations that concern climate change. However, the same approach can also be applied to other topics.

1. How to extract data from the Flemish Parliament's API in a way that makes it possible to perform valuable analyses?
    a. How should the data be collected?
    b. How should the data be pre-processed?

2. To what extent can sentiment analysis models that have been fine-tuned on text data from other contexts, accurately predict the sentiment in spoken text from the Flemish Parliament?
    a. Which BERT model (nlptown[2], DTAI[3], and gilesitorr[4]), fine-tuned for sentiment analysis, most accurately predicts the sentiment in parliamentary transcripts?

[2] https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment
[3] https://huggingface.co/DTAI-KULeuven/robbert-v2-dutch-sentiment
[4] https://huggingface.co/gilesitorr/bert-base-multilingual-uncased-sentiment-3labels

3. How has the sentiment towards climate change within the Flemish Parliament changed over the last twenty years according to the best performing model?
    a. How do the changes in sentiment differ across parties?
    b. What events have caused parliamentarians to speak more negative or positive about climate-related topics?

Literature Review

Political Polarization on Climate Change

The emergence of populist anti-political establishment parties has fuelled an increase of political polarization in countries across Europe (Casal Bértoa & Rama, 2021). The development towards more polarized societies can damage democracy since polarization discourages citizens from political participation and decreases political consensus. Furthermore, populist parties let their decisions depend on which way the wind blows which means that in most cases they lack a long-term policy perspective (Casal Bértoa & Rama, 2021).

After the IPCC published their *Fourth Assessment Report* that stressed that global warming was very likely due to human activities, climate change became a hot topic (IPCC, 2007). However, social attention does not equal social consensus and Dunlap et al. (2016) found that partisan polarization among the Republican and Democrat public has increased over the period 2007 to 2016.

In Europe, right-wing parties are more likely to be sceptical about climate change and oppose policies designed to combat climate change (Kulin et al., 2021). Schaller & Carius (2019) mapped the agendas of right-wing populist parties across Europe. They found that in Belgium, the right-wing party Vlaams Belang [Flemish Interest] voted against most climate policy proposals. The party even refused to participate in the parliamentary debate on Flemish climate policy in 2014. Based on these findings, a difference in sentiment towards climate-related topics is expected across political parties. Sentiment analysis on the oral questions and interpellations in the Flemish Parliament over the last 20 years could reveal how the sentiment within the Parliament and across parties has developed over time.

With regards to what is causing partisan polarization, some research points towards the role of the media and its disability to present complex and multidimensional problems (Freudenberg & Muselli, 2010; Fisher et al., 2013; Dunlap et al., 2016). Balčytienė, & Juraitė (2015) argue that the growing personalized access to information reinforces media fragmentation, driving audience segmentation and increasing political and social polarization across various nations in Europa.

Through content selection, news media have the power to affect the public opinion. Fan et al. (2019) found that media sources make a selection of politicians' opinionated quotes as a way to bring across their own opinions. This phenomenon undermines political transparency, as citizens mostly rely on media outlets to follow politics. The Open Data Directive, which entered into force in 2019, is a way to combat the public dependency to media outlets. The increasing amount of government data that is now available allows anyone to analyse politics without relying on secondary sources, such as mainstream or social media, for information (*European Legislation on Open Data*, 2019).

Sentiment Analysis

Sentiment analysis or opinion mining is defined as finding authors' opinions about specific entities (Feldman, 2013). Opinion mining and sentiment analysis are used interchangeably and, for example, used to determine a document's sentiment polarity (Abercrombie & Batista-Navarro, 2020b). Sentiment analysis can identify whether a text is positive, negative, or neutral. The author's sentiment can be determined on different granularity levels: document level, sentence level, and feature level (Abercrombie & Batista-Navarro, 2018a). The first two levels have a downside. Authors might express different sentiments for different entities in the same document or sentence. To understand the author's sentiment towards a specific entity, the higher the level of granularity, the more certainty we have in determining one's sentiment towards a single entity. On the other hand, by using a high granularity, you might miss some of the sentiment hidden in other parts of the sentence, paragraph or document.

Research into sentiment and position-taking analysis of parliamentary debates has attracted attention from researchers from computer science and political and social science backgrounds (Abercrombie & Batiste-Navarro, 2020b). For example, Dahal et al. (2019) used sentiment analysis on Twitter data to challenge surveys as a common approach to measure public opinions. They collected around 2 billion tweets from many different countries to gain insights into the sentiment towards climate change. The sentiment was tracked over multiple months. They found it to peak negatively around events such as the US withdrawing from the Paris Agreement and a hurricane hitting Cuba. Although these peaks are visible and the overall sentiment was negative, the average sentiment stayed within the same bandwidth for most of the researched period.

Similar research by An et al. (2014) showed that opinions of Twitter users towards climate change could change over time and in the aftermath of a specific event, in this case, an IPCC report. They stress that Twitter users are not representative of all social groups. Hence, investigating how sentiment changes among politicians over time and in the aftermath of certain events might better represent public opinion. The aftermath of an event, in particular, is interesting since a significant change in sentiment might indicate that politicians are reactive instead of proactive concerning climate change.

Among South Korean elite, Han (2022) found a trend towards increasing political polarization. Significant political events such as the impeachment of a President, are expected to have influenced the level of polarization at a specific point in time significantly. Although the context is very different, the same could be true for significant events concerning climate such as the publication of IPCC reports. These could cause the sentiment to peak negatively.

To analyse political transcripts, researchers have used different technological approaches to perform sentiment analyses. On one hand, the lexicon based approach, and on the other hand the approaches based on machine learning and deep learning (Catelli, 2022). Evidence shows that BERT models, fine-tuned for sentiment analysis have superior performance over other machine learning algorithms and over the lexicon based approaches (Catelli, 2022; Abercrombie & Batista-Navarro, 2020a; Alaparthi & Mishra, 2021; Nair et al., 2021; González-Carvajal & Garrido-Merchán, 2020). BERT, a Bidirectional Encoder Representations from Transformers, designed to pre-train and fine-tune deep bidirectional representations from an unlabelled text by jointly conditioning both left and right context in all layers, was introduced by Devlin et al. (2018). The base model has been trained on a huge dataset. This allows users to use this model and fine-tune it for various NLP tasks. These fine-tuned models are very effective for performing sentiment analysis on text in various contexts (Catelli, 2022; Abercrombie & Batista-Navarro, 2020a).

Abercrombie & Batista-Navarro (2020a) fine-tuned BERT embeddings on transcripts of debates from the United Kingdom (UK) Parliament. On a large corpus of speech transcripts, their model, which included the BERT architecture, outperformed the often-used support vector machine (SVM) and the multi-layer perceptron (MLP) models. Their fine-tuned BERT model achieved an accuracy of 67% on the full pre-processed corpus. However, this model was fine-tuned on transcripts of the UK Parliament specifically and may, therefore, not be applicable outside the UK. Thus, other models are needed in order to examine how language or sentiment in the political debate outside the UK has changed over time. Whether the fine-tuned sentiment analysis models available on Hugging Face can match the performance of the model used by Abercrombie & Batista-Navarro (2020a) will be examined in this study. The performance of three sentiment analysis models from Hugging Face will be evaluated in the next chapter.

Methodology & Experimental Setup

In the following chapter, the methodology of the research is explained. First, a summary of the research steps is given. These steps are visually presented in figure 3. Second, the data collection process will be discussed, followed by the pre-processing steps. Fourth, the sampling process is explained and the annotations of the sample dataset are evaluated using the inter-annotator agreement (IAA). Hereafter, the three models, used to perform the sentiment analysis, are described and the performance of each model on the sample dataset is presented and discussed. In the last section, two alternative approaches are explored.

Research setup

The Flemish Parliament stores all parliamentarian documents in a database. From this database, all kinds of documents can be retrieved including motions, debate transcripts, questions and interpellations, documents that are provided with motions, etc. The data in this database can be collected through an API. In R, there is a package available called Flempar that allows querying the web API of the Flemish Parliament. Because retrieving the data from the API takes a long time it makes sense to not be limited to the computing power of a local machine, the data collection was performed on a Google Cloud virtual machine. As the Flempar package and the necessary sentiment analysis tools are not included in any of the Docker Images on Dockerhub, a Docker Image was created specifically for this study. This Image is also made available on GitHub to ensure the reproducibility of this study.

Figure 3 presents a flowchart of the entire research methodology, from data collection to analysis. The research approach is as follows. The first step is collecting all data using the Flempar package in R. The result are four datasets which are then joined together to end up with one dataset. This dataset includes all oral questions and interpellations from the plenary sessions in the Flemish Parliament over the last 22 years. The next step is pre-processing the data so that the text is in the right format and only the oral questions and interpellations regarding climate change remain. This dataset is called Climate & Environment. To be able to compare the performance of different sentiment analysis models, a sample is taken and manually annotated. Then the best performing model is selected and employed on the entire Climate & Environment dataset. The last step is visually presenting and analysing the results.

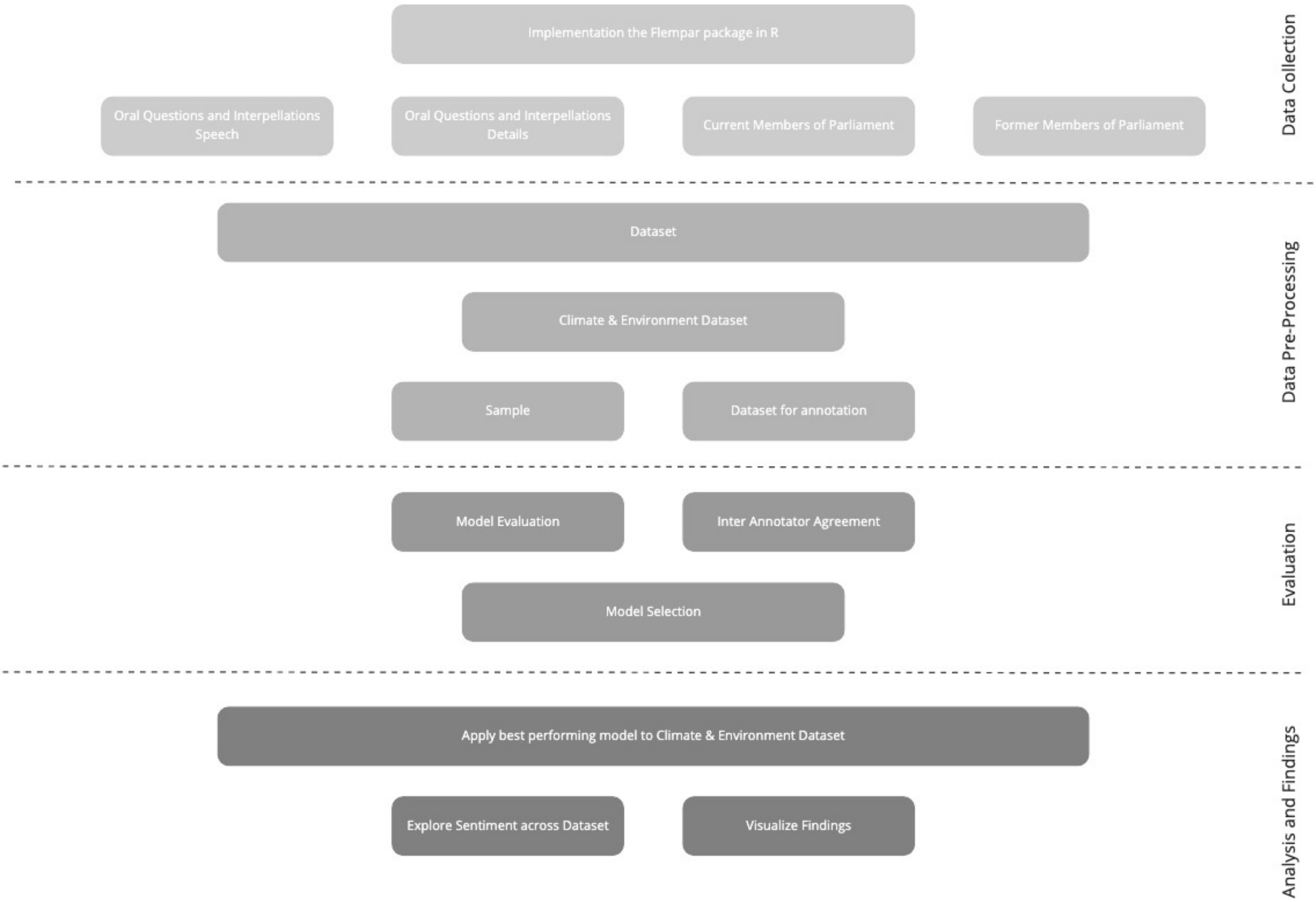

Fig. 3 *Flowchart of research methodology*

Data collection

The Flemish Parliament API is limited to processing 10000 pages per call to reduce the probability of having to process an unnecessary amount of queries that would cause the API to crash. This makes it impossible to extract all data from the entire time period in one go. Besides, the API could experience downtime or server-side issues that can obstruct the data collection. To avoid losing all progress when the connection to the API is lost, the code to collect the data includes some safety measures. The following paragraph will describe the data collection process step-by-step.

First, create a dataframe consisting of each interval's start and end dates. In this case, the interval is set to 365 days. Second, create a list with the same length as the date dataframe. Then, use a for loop to iterate over the number of rows of the date dataframe. Use each row's start and end date as input for Flempar's get_work() function. This function allows to access the API and retrieve specific information through a set of parameters. The output is written to the list created earlier. In addition to writing the output of each interval to the list, save the output as an rds file, this way, no progress is lost in case of errors occur.

Using the for loop explained above, all data can be collected. But, as mentioned earlier, the connection to the API could get lost. Usually, it is possible to reconnect after several minutes. To avoid having to restart the collection manually, the try() function can be incorporated in the for loop. This works as follows. If the connection to the API is lost and an error occurs, the combination of try() and sleep() allows to put our loop to sleep for a given period. After this time has passed, the try function executes the get_work() function again. When the connection is still lost, this process is repeated to a maximum of three times. If, after sleeping three times, the get_work() function still produces an error, the loop is stopped.

The steps above result in a folder with 23 dataframes, one for each time interval. After combining the separate files, the combined dataframe, that includes all the plenary oral questions and interpellations, has 82459 rows and four columns. Each row represents one oral question or interpellations, the transcript is in the text. The other columns include information on who is the speaker, including the speakers unique id. As is, this dataframe does not include all the information necessary to answer the research questions. More information is needed. By using the same structure but different parameter for the get_work() function, additional information such as the date is retrieved.

Data concerning the personal information of each MP is collected using a different function called get_mp(). Two calls are made to get all information on the current and former MP's. The output of these function is rather large and consist of several nested lists. Several transformations, such as unnesting multiple variables, are needed to end up with a square dataframe. To account for the fact that some MP's have switched parties over time, the MP dataframe is exploded to a dataframe in where each year an MP has been in the parliament under a party appears as one row.

In the following section, the data pre-processing is described in detail. For this study specifically, data pre-processing is essential for three reasons. First of all, the data collection results in four different datasets: 1) a dataset that includes all the speech of the oral questions and interpellations, 2) a dataset that includes all the details such as the date and the id specific to each oral question or interpellation, 3) a dataset including all information of current members of parliament, and 4) a dataset that includes all the information of former members of parliament. These four datasets need to be joined together to create a complete dataset that holds all the valuable information required for further analysis. Second, because this study focuses on the sentiment on climate change, climate-related oral questions and interpellations are filtered out. Third, Parliamentary information is complex as long text fragments can hold information on multiple topics. Therefore, the text fragments are split into sentences. Additionally, this helps to avoid the input sequences being too long for the chosen models, which have a maximum sequence input length of 512 tokens.

As mentioned, the data collection process results in four datasets. The speech and details of the oral questions and interpellations are first joined together. The joined dataset consists of 148429 rows and 24 columns. Then, the dataset that includes the information on current and former MP's is joined to the oral questions and interpellations.

Since this study focusses on the sentiment around climate change, the rows need to be filtered based on their theme. There are six theme variables. If the value of any of these six variables is labelled as 'Natuur en Milieu' ['Nature and Environment'], the oral question or interpellation is regarded to be somehow related to climate policy. Overall, there are 14286 speech fragments related to climate. However, these speech fragments also include those of the chairman of the parliament. As the chairman is expected to stay neutral, his or her comments are excluded. The final dataset consists of 14286 rows and 36 columns.

To properly perform sentiment analysis on text data, several pre-processing techniques have to be considered. Pre-processing text data is a research on its own, and proper pre-processing has found to positively affect the performance of sentiment analysis models (Alam & Yao, 2019). In R, rows in the text column are disconcerted of weird character combinations and html tags. After all text is cleaned, the manipulations in R are finished. The final dataset is saved as a CSV file to be able to load it into Python and perform the analysis.

Compared to social media posts and reviews typically targeted for sentiment analysis, parliamentary transcripts are inherently more complex as they hold far more information, covering different topics in one speech fragment (Abercrombie & Batista-Navarro, 2020a). Therefore, the decision is made to split the speech fragments to sentence level and increase the granularity. Besides increasing the granularity, this step also helps to overcome the issue that the BERT based sentiment analysis models from Hugging Face have a maximum sequence input length of 512 tokens. Applying it to complete speech fragments with a length of more than 512 could result in a loss of valuable information. The sentence-level dataset consists of 120267 rows, representing all the sentences and 36 columns. Most of these columns have been discussed earlier. Any columns that have not been discussed but are later used for analysis, will be discussed in the following sections.

The lack of sentiment labels in the final dataset makes it impossible to finetune a BERT model on this dataset specifically and hinders the possibility to compare the performance of the nlptown, DTAI, and gilesitorr models from Huggingface. Therefore, the last pre-processing step is to take a sample of 100 sentences that can be manually annotated.

Sample

As mentioned, the lack of labelled data hinders the possibility to evaluate performance of different models. To still be able to evaluate the performance of different finetuned BERT models, a sample of 100 sentences is taken from the dataset. These sentences are manually annotated with a negative (-1), neutral (0) or positive (1) label. Using a majority vote, the final label is calculated. These final labels are then used to evaluate the performance of our model.

To reduce the effect of annotator bias, the labels are given by three different annotators. The annotations are then compared to each other to calculate the inter-annotator agreement (IAA). The most common measure to calculate the IAA is to count the number of identical annotations and report this number as a percentage of our sample; this is called the raw agreement (Arstein, 2017; Bayerl & Paul, 2011). Although the raw agreement is easy to understand, agreement in itself does not necessarily imply that the annotation is reliable. Some agreements may be accidental, and this accidental agreement could be very high (Arstein, 2017).

Of the 100 sentences, the annotators consider 47 sentences neutral, 34 sentences are considered negative, and 19 sentences are considered positive. As mentioned, these final labels are calculated using a majority vote, there are just a few sentences that are considered negative, neutral or positive by all three annotators. Table 1 presents some of the sentences that are part of the sample dataset, including their annotations. Based on these few sentences it becomes clear that full consensus among annotators is limited. The raw agreement of the manually labelled sentences by the three annotators is 29%. This shows that, even for humans, labelling this data based on sentiment is a complex task.

To properly evaluate the IAA, this study uses Fleiss's $\boldsymbol{\kappa}$, where N is the total number of labels given to the annotated data by all annotators, and where $\mathbf{n}_k$ is the total number of labels of category $\boldsymbol{k}$ given by all annotators (Arstein, 2017). Let:

$$Fleiss's\ \boldsymbol{\kappa}: \qquad A_e = \frac{1}{N^2}\sum_k (\mathbf{n}_k)^2$$

The Fleiss's $\boldsymbol{\kappa}$ of the labels in this dataset is 0.356. According to the benchmarks provided by Landis & Koch (1977), the strength of agreement falls in the range of kappa that is considered fair. However, based on the limited amount of possible labels, more agreement was expected. The goal of IAA is not only to validate the annotation scheme, but also identify ambiguities or difficulties in the data (Arstein, 2017). The conclusion that can be drawn from the relatively low strength of agreement is that the sentences are relatively hard to label. Therefore, it is not expected that the pretrained models will perform particularly well in finding the sentiment.

Sentences	Label 1	Label 2	Label 3	Label
nederland, frankrijk, duitsland, spanje, italië, zelfs polen: allemaal willen ze uit dat verdrag stappen [the Netherlands, France, Germany, Spain, Italy, and even Poland, all want to withdraw from the treaty]	-1	-1	-1	-1
dat was niet de eerste keer, maar de twaalfde keer dit jaar [that was not the first, but the twelfth time this year]	-1	-1	-1	-1
daarin zitten maatregelen [it contains measures]	0	-1	1	0
het zal gaan over de financieringswet [the finance law will be discussed]	0	0	0	0
minister, u weet ongetwijfeld dat het anb een digitale campagne is gestart: doe de #natuurmove [minister, you undoubtly know that the anb has started a digital campaign: do the #naturemove]	0	1	0	0

ik ben ervan overtuigd dat op die lijn uiteindelijk ook een resultaat moet worden geboekt *[i am convinced that along this line, a result must ultimately be achieved]*	0	1	1	1
ze zorgt ervoor dat het op een goede, correcte wijze gebeurt *[she makes sure that it is done in a good, correct way]*	1	1	1	1

Table 1. *Manually annotated sentences*

Models

The first model is the nlptown/bert-base-multilingual-uncased-sentiment model (*Nlptown/Bert-base-multilingual-uncased-sentiment · Hugging Face*, n.d.). This model has been finetuned on 80.000 product reviews and achieved an accuracy of 57%. The second model, called DTAI-KULeuven/robbert-v2-dutch-base, is a finetuned model based on RobBERT(v2) (*DTAI-KULeuven/robbert-v2-dutch-sentiment · Hugging Face*, 2022). For sentiment analysis, this model has been finetuned on book reviews and news articles. The output is either negative (-1), neutral (0), or positive (1). Like the nlptown model, the third model is also based on bert-base-multilingual and is called gilesitorr/bert-base-multilingual-uncased-sentiment-3labels (*Gilesitorr/Bert-base-multilingual-uncased-sentiment-3labels · Hugging Face*, n.d.). The model is finetuned on a total 105879 words, it is unknown how many of these are Dutch. However, unlike the other two models, gilesitorr outputs the same three labels used to manually label the sample.

Model Evaluation

The output of the nlptown model is a sentiment score between one and five, based on the amount of starts given to a product. Because the sample data has just three categories, the five are converted to three categories before evaluation. An output of one or two are considered negative (-1), an output of three is considered neutral (0), and an output of four or five is considered positive (1). Based on these assumptions, nlptown achieved an accuracy of 50% on the sample dataset. The results are presented in table 2. The nlptown model performed the sentiment analysis on 100 sentences in 28 seconds. Based on the precision scores, this model most of all struggles with correctly predicting the positive class.

	Precision	Recall	F1-score	Support
Negative (-1)	0.514	0.529	0.522	34
Neutral (0)	0.606	0.426	0.500	47
Positive (1)	0.375	0.632	0.471	19
Accuracy			**0.500**	100
Macro avg	0.498	0.529	0.497	100
Weighted avg	0.531	0.500	0.502	100

Table 2. *Classification report nlptown*

On the labelled sample data, the accuracy of the DTAI model is worse than a random classifier, namely 32%. What strikes most in the classification report of the DTAI model is the extremely low recall score of the neutral class low precision scores for the neutral and positive class. This is most likely due to the class imbalance of the sample, where 47 of the sentences are labelled as neutral and just 19 are positive.

	Precision	Recall	F1-score	Support
Negative (-1)	0.538	0.412	0.467	34
Neutral (0)	0.300	0.064	0.105	47
Positive (1)	0.234	0.789	0.361	19
Accuracy			**0.320**	100
Macro avg	0.358	0.422	0.311	100
Weighted avg	0.369	0.320	0.277	100

Table 3. *Classification report DTAI*

The gilesitorr model achieves an accuracy of 47.0%. The classification report of this model is presented in Table 4. Just as in the classification report of the DTAI model, the effects of class imbalances in the labelled dataset become apparent.

	Precision	Recall	F1-score	Support
Negative (-1)	0.470	0.706	0.565	34
Neutral (0)	0.538	0.447	0.488	47
Positive (1)	0.2	0.105	0.138	19
Accuracy			**0.470**	100
Macro avg	0.403	0.419	0.397	100
Weighted avg	0.451	0.479	0.448	100

Table 4. *Classification report gilesitorr*

Taking into account the performance of all three models, nlptown is considered to perform best on this set of political transcripts. Although the accuracy of the nlptown model on the labelled dataset cannot be considered high, it outperforms a random classifier where the accuracy would be just 33% (1/3 labels). Also, considering the fact that manually, the sentences are also hard to predict - raw agreement and fleiss's **κ** are just 29% and 0.356 – an accuracy of 50% is considered acceptable for the nature of this study. Therefore, the finetuned model by nlptown is used to analyse the entire dataset of 120267 sentences. The results of this analysis will be presented and discussed in the following chapter. What has to be taken into account is that the presented findings are the result of a model with an accuracy of just 50% in addition to an already low IAA.

Alternative Approaches

Throughout this research, several decisions have been made that collectively lead to the selection of the nlptown model and the findings discussed in the following chapter. Numerous other choices could have been made. One alternative approach is splitting the text fragments into paragraphs instead of sentences, decreasing the level of granularity. This way, the model would have more input to predict the sentiment score while staying within the input size limits of BERT. The issue with this approach is that it would result in significantly longer pieces of text to label manually. This could make labelling even more complex, resulting in lower IAA.

A second alternative approach would be to avoid manual labelling altogether. Instead, all three models could have been applied to the entire dataset. A larger labelled dataset could have been created using a majority vote. This approach was only tested on the sample because running *the nlptown* model on the entire dataset took nearly an entire day. The three models agreed on the sentiment of the 21% of the sentences.

Results

As mentioned, the *nlptown* model performed best on the annotated dataset. Therefore, this model is applied to the complete Climate & Environment dataset to perform the sentiment analysis. In the following chapter, the findings of this sentiment analysis are presented. Several visualizations show how the sentiment regarding climate has changed over time. This includes how the sentiment has changed in the Flemish Parliament as a whole, across parties, and how the sentiment differed between the opposition and coalition.

Findings

Although the model accuracy of 50% causes the need to interpret the outcomes with caution, some interesting patterns can be found in the data. Figure 4 presents the average sentiment score of all parliamentarians over the past 16 years. Alike the findings by Dahal et al., 2019, the sentiment score remains fairly neutral and within the same bandwidth for the largest part of the researched period. The trendline however, shows a slight change towards a more positive sentiment.

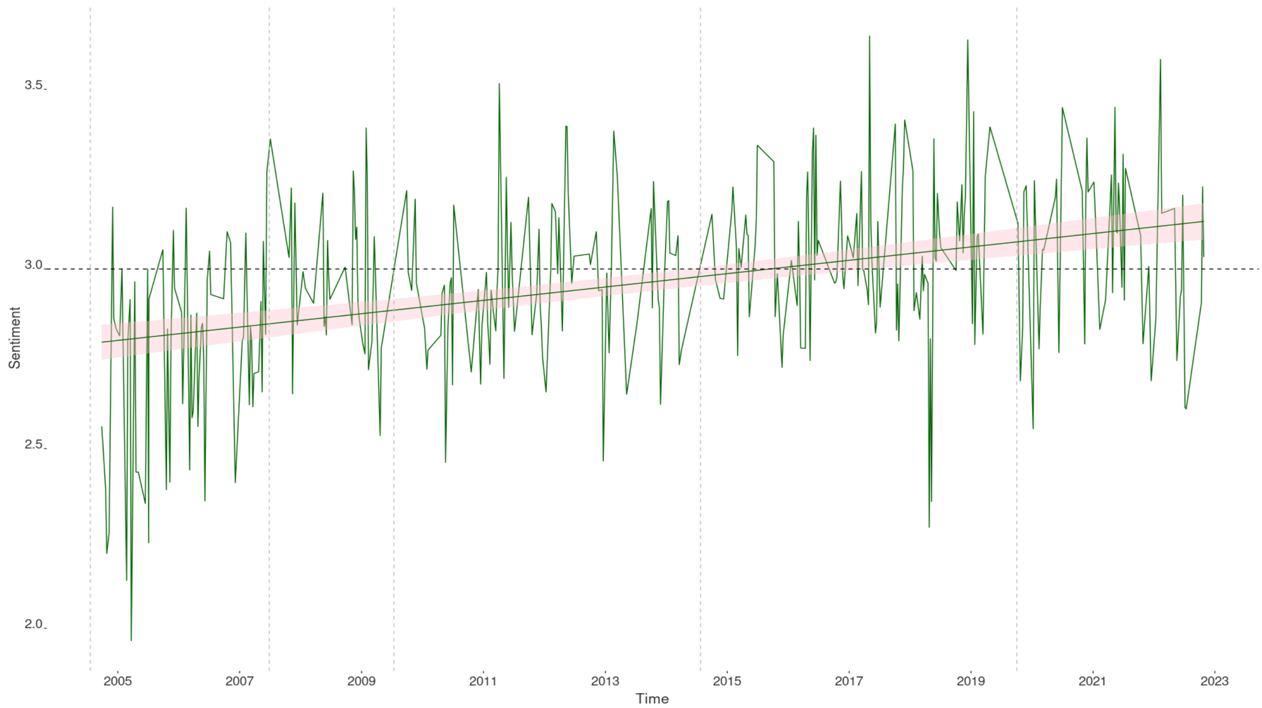

Fig. 4 *Sentiment among parliamentarians over time*

Note that the Y-axis represents the sentiment score. A sentiment score of three is considered neutral. A sentiment score of one is considered very negative and a score of five is associated with a very positive sentiment.

Across parties

Figure 5 presents the difference in sentiment across parties over time. Shown are the six largest political parties. Several parties have changed names over time, the sentiment of these parties has been merged and included under the present party name. From Figure 5 it becomes clear that there are significant differences in sentiment across parties. Overall, Vlaams Belang has the lowest sentiment score, implying its parliamentarians speak most negatively on topics related to climate. Second most negative scores Groen, a party which pursues doing good with regards to our climate. This might explain their negative sentiment. They strive for policy change in favor of our planet, resulting in critically addressing current climate policy. This critical tone of voice is likely to result in a more negative sentiment.

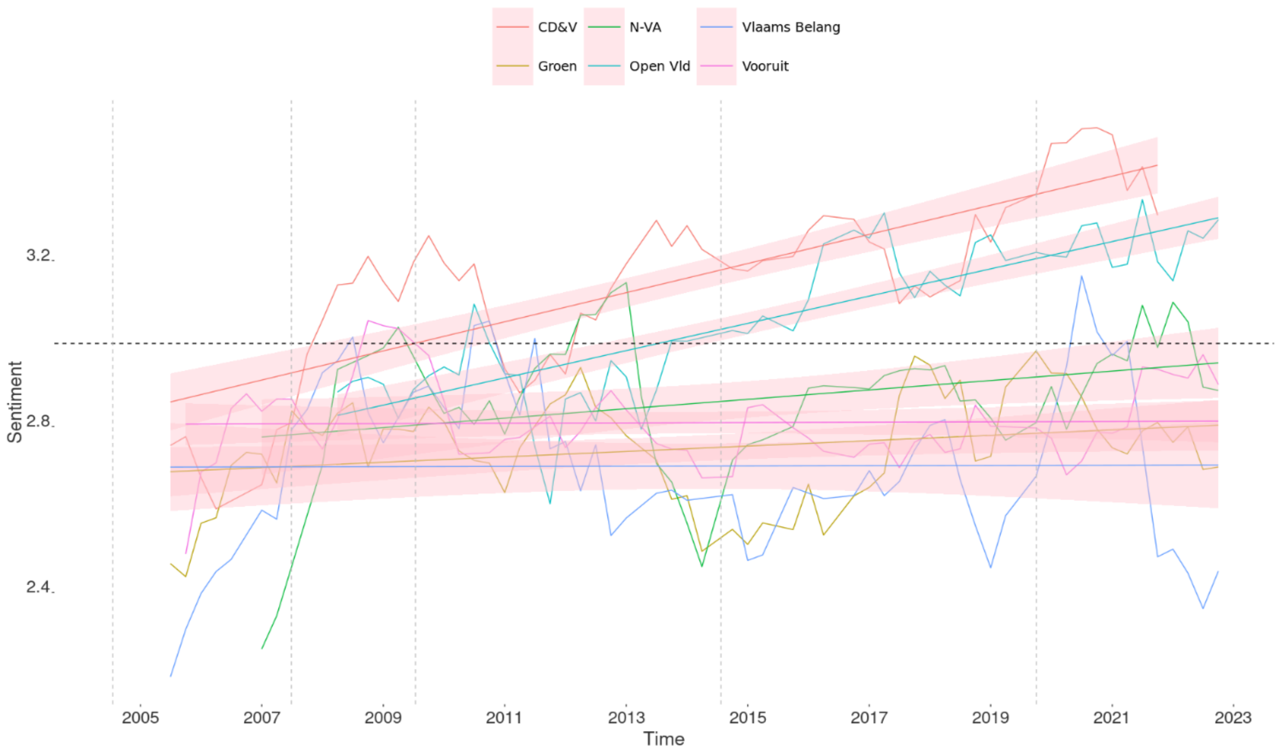

Fig. 5 *Sentiment across parties over time*

Coalition government vs Opposition

Figure 6 presents the sentiment concerning climate related topics of the parties that are part of the coalition and parties that classify as the opposition. For both groups the sentiment regarding climate related topics, is becoming more positive over time. In general, the coalition speaks more positively than the opposition and this trend has increased over the last 18 years. The pink area around the trendline represents a 95% confidence interval. From this, it can be concluded that since 2008, the coalition's sentiment is significantly more positive than the opposition's sentiment.

Several factors could cause this effect, but the nature of the oral questions and interpellations is most likely the reason why the coalition is significantly more positive. This nature is as follows: A member of parliament has a question related to a certain topic. This question is addressed to, in most cases, a minister. Logically, these questions are based on certain societal events or the political agenda of the coalition. Because of the latter, it is expected that the parties that are part of the coalition, in general speak with a more positive tone of voice, regardless of the topic. The opposition make remarks on the policy. Therefore, the opposition is expected to speak more negatively.

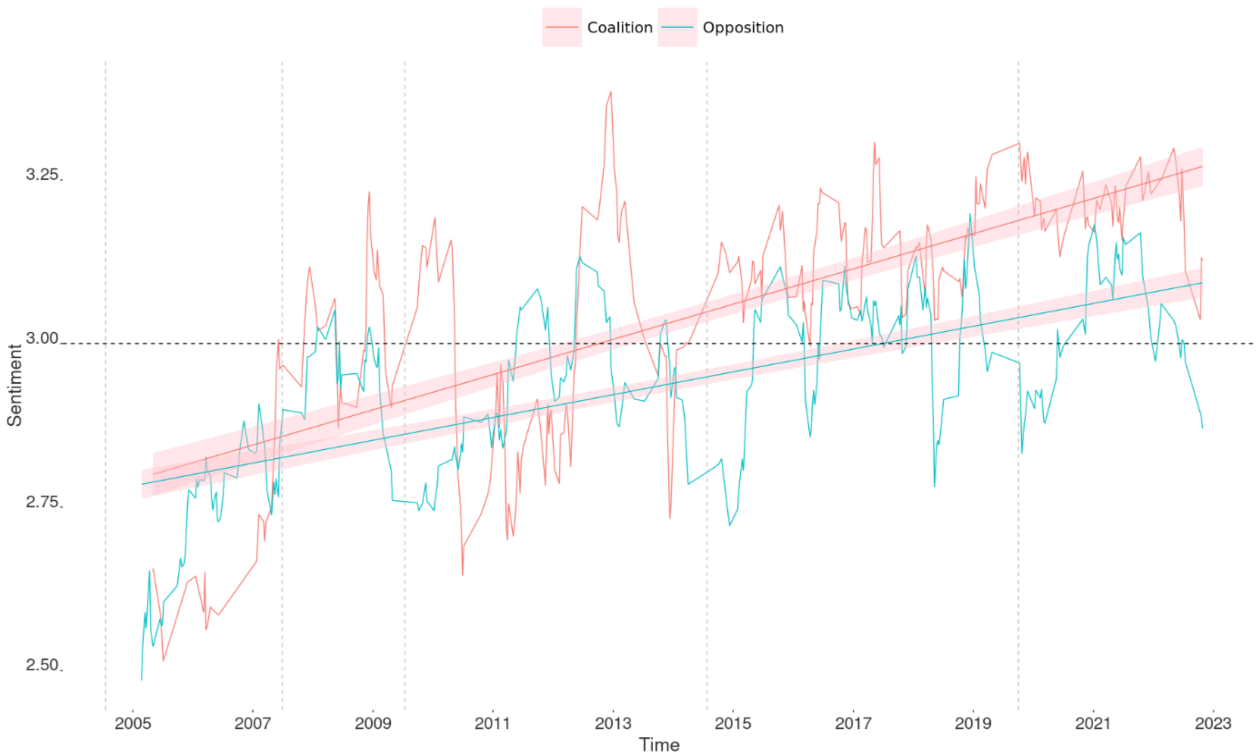

Fig 6. *Sentiment of coalition and opposition over time*

At certain points in time, the sentiment peaks, either positively or negatively. The peak that stands out the most is the big change towards a positive sentiment within the coalition in the end of 2012. Diving back into the data, positive sentiment scores among the coalition are found regarding the suspension of an environmental permit for Uplace. In 2012, the real estate development company Uplace received a permit to build a mall near Brussels. This development got a lot of attention due to objections by local entrepreneurs and environmental organizations (Nws, 2012). The permit was finally given by a CD&V minister. The peak in positive sentiment can be explained by the fact that this minister and its fellow party members had to defend this decision, and therefore have used language that classifies as more positive.

The peak in the negative sentiment among the opposition linked back to the Flemish contribution to the UN climate fund. This contribution sparked a debate about climate in which the opposition expressed their concerns. In the transcripts, the following interpellation is found by a member of the opposition Hermes Sanctorum-Vandevoorde: '*Minister, u betreurt het samen met ons, zegt u, dat de milieubewegingen het overleg hebben verlaten, maar u hebt ze natuurlijk ook weggejaagd. Als ze deelnemen aan dat overleg, verwachten ze dat ze au sérieux worden genomen in plaats van een eenzijdige aanpak ten voordele van enkele belangenorganisaties. Het is normaal dat zij op hun rechten staan. Minister, de heer Vandaele heeft het net aangehaald: fosfaat is een torenhoog probleem. U spreekt wel over een verbetering van de waterkwaliteit, maar dat geldt absoluut niet voor fosfaat. Dat is nu net een van de elementen waarom de milieubeweging uit het overleg stapt want er wordt onvoldoende perspectief geboden om die fosfaatproblemen op te lossen. Minister, wat zegt Europa over uw voorstel van aanpak van de fosfaatvervuiling?'* [*'Minister, you say that we both regret that the environmental movement have left the consultation, but of course you have also chased them away. If they participate in that consultation, they expect to be taken seriously as opposed to one-sided approach in favor of few interest groups. It makes sense for them to insist on their rights. Minister, as Mr. Vandaele just mentioned: phosphate is a huge problem. You do speak about an improvement in water quality, but that absolutely does not apply to phosphate. That being one of the elements why the environmental movement is withdrawing from the consultation, because insufficient prospects are offered for solving these phosphate problems. Minister, what does Europe have to say about your proposal for tackling phosphate pollution?'*]. This clearly is a concern and negative comment that the sentiment analysis brought to light correctly.

The dotted vertical lines represent the end and start date of a coalition. Over this period, there have been five different coalitions (*Overzicht Van De Vlaamse Regeringen Sinds 1981*, n.d.). Of these coalitions, Peeters II between 13-07-2009 and 25-07-2014 seems the most turbulent, based relatively large changes in sentiment over this period. Another remarkable pattern that appears in figure 6 also relates to the changes in coalition. It seems that at the start of a new coalition, the sentiment of the opposition becomes more positive. The contrary applies to the end of a coalition, where the sentiment of the opposition becomes more negative.

Minister vs Parliamentarian

Figure 7 presents the sentiment of the ministers and the parliamentarians over time. From this figure it becomes clear that overall, sentiment among ministers is higher compared to the sentiment of parliamentarians. This difference is significant and while the sentiment among both groups becomes more positive over time, the difference remains somewhat the same. The difference in sentiment can be explained by the fact that the oral questions and interpellations in most cases concern policy. These questions are initiated by the parliamentarians, directed to and answered by the minister. In most cases minister will defend their policy or point of view, resulting in a more positive way of speaking.

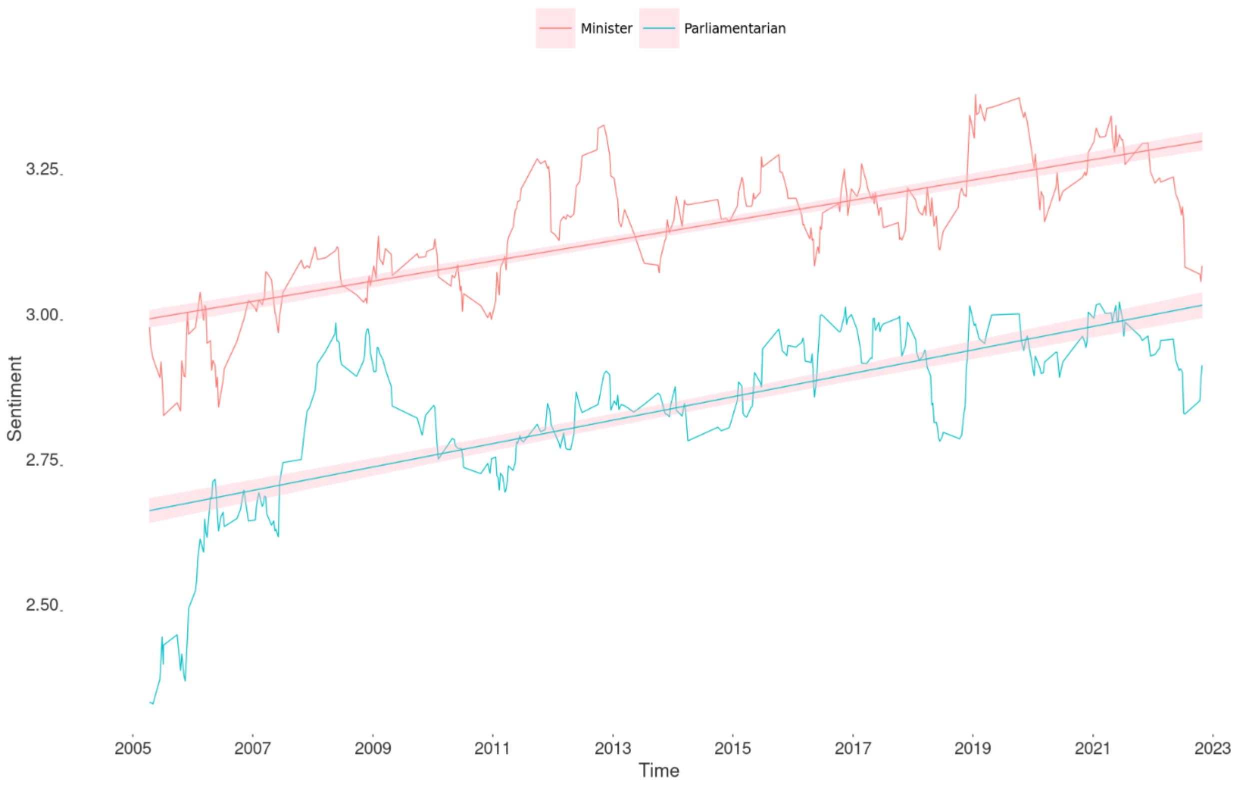

Fig. 7 *Sentiment of minister and parliamentarian over time*

Discussion & Conclusion

The answers to the three research questions elaborated on below will cover the extent to which this research successfully achieved the goals set out in the introduction . Next, the limitations of this research will be discussed. Followed by the scientific and societal implications and this recommendations for future research.

Research Questions

RQ1: How to extract data from the Flemish Parliament's API in a way that makes it possible to perform valuable analyses?

The Flempar package allows users to query the API of the Flemish Parliament. Through various functions, it is possible to collect a wide variety of data, such as speech fragments from all plenary sessions or commissions, oral questions and interpellations of plenary sessions, and information on the MPs. On their own, the datasets that these functions put out are not necessarily valuable for further analysis. It takes various pre-processing steps to create a dataset that holds enough information to be used for conducting valuable analyses. The pre-processing includes: 1) joining the speech fragments to session details such as the date, the MP who is speaking, and the subject of the text fragment, and 2) joining the personal information of the MPs to the speech and details.

Getting all the information of the MP is not as straight forward as collecting the session details of speech fragments. Numerous MPs have changed parties over time and to connect the correct membership to the correct period, several steps, such as unnesting multiple lists, have to be taken.

All things considered, it is possible to extract a sufficient amount of data from the API that is suitable for further valuable analysis. However, the collection and pre-processing of the data require advanced knowledge of the R software and its packages. As a result, the data is not considered open to the general public, but only to a select group of people. It does not achieve the increase in transparency aimed for with the *Open Data Directive*.

RQ2: To what extent can sentiment analysis models that have been fine-tuned on text data from other contexts, accurately predict the sentiment in spoken text from the Flemish Parliament?

The lack of sentiment labels in the dataset makes it impossible to fine-tune a BERT model for sentiment analysis on Flemish Parliament data. This study compared three sentiment analysis models from Hugging Face to overcome this issue. To compare the performance of these models, a small dataset with sentiment labels was created by taking a sample of 100 sentences and having these sentences manually annotated by three annotators. The annotators classified each sentence as negative (-1), neutral (0), or positive (1). Based on the low IAA score, detecting the sentiment in these speech fragments is difficult, even for humans.

By using a majority vote, a final sentiment label was given to each sentence. On this labelled dataset, the *nlptown* model, which is fine-tuned on product reviews, performed best. The model achieved an accuracy of 50%. An accuracy of 50% cannot be considered good, and compared to fine-tuned model of Abercrombie & Batista-Navarro (2020a), the *nlptown* model performs much worse.

RQ3: How has the sentiment towards climate change within the Flemish Parliament changed over the last 20 years according to the best performing model?

Although the *nlptown* model did not predict the sentiment of the labelled dataset as accurately as the fine-tuned model of Abercrombie & Batista-Navarro (2020a) did. When applied to the entire dataset, the sentiment analysis did bring to light some interesting trends and patterns. Furthermore, the model did seem to detect significant changes in sentiment around certain events.

In contrast to what was expected, the sentiment has become more positive over last two decades. The results show some differences in sentiment across parties. Large parties such CD&V and Open Vld have a significantly more positive sentiment. Based on the fact these parties have been part of the coalition, it makes sense they use language that classifies as more positive. Most of the time, the oral questions are initiated by the opposition, who aim to challenge the coalition and their policy initiatives. This format explains a more negative tone of voice coming from the opposition. The parties Groen and Vlaams Belang had the most negative sentiment score during the same period. This also makes sense as these parties have been large opposition parties, likely to have challenged the coalition.

Members of the coalition spoke significantly more positive in late 2012. This peak could be traced back to a climate permit granted for a mall near Brussels. From the transcripts it became clear that the minister defended her decision to grant the permit, causing the sentiment score to become more positive. The peak in negative sentiment among the opposition in late 2014 could be traced back to some comments by the opposition. It seems that the BERT model correctly identified these events as more positive and more negative. Although BERT seems to have correctly classified some of the peaks, the result must be interpreted with caution since the accuracy on the small labelled dataset was just 50%.

Limitations

As mentioned, this research has its limitations. Most of them are inherent to the exploratory nature. First of all, because the BERT model is not finetuned on political speech specifically, and sentiment classification accuracy is based on a sample of 100 sentences, the findings have to be interpreted with caution. The most robust way of overcoming this limitation is by creating a larger labelled dataset. Ideally, this labelled dataset has lower granularity. For instance, by labelling the sentiment of whole paragraphs instead of sentences. This way, the annotators have more context, which might help to properly label.

The ability to draw a conclusion from sentiment scores is the second limitation. A negative sentiment regarding climate does not necessarily indicate that one is against combatting climate change. It could also indicate that a parliamentarian is very concerned about climate change and would prefer stricter regulations and may therefore use many words that classify as negative. On the other hand, a parliamentarian not in favour of stricter regulations is also likely to use words that classify as negative. This shows that context knowledge is necessary to be able to draw conclusions.

The third limitations is that is impossible to know whether all the oral questions and interpellations that have been analysed actually concern climate related topics. The questions are filtered on the 'Natuur en Milieu' ['Nature and Environment'] theme. However, each question can have six themes and if 'Natuur en Milieu' [Nature and Environment] is the sixth theme, the question might be related to this theme, but not solely about climate change. An option would be to only include questions that have 'Natuur en Milieu' ['Nature and Environment'] as the first theme, but this way, valuable information could be lost as fewer questions remain. A different approach to select only the oral questions and interpellations that concern climate change, is to use a lexicon of words most frequently used in the climate change debate. However, putting together such a lexicon could be difficult tasks as words like 'klimaat' ['climate'] are used in a lot of different context besides climate change (e.g. ondernemersklimaat [entrepreneurial climate], arbeidsklimaat [working climate]).

Implications & Future research

Due to the limitations, the implications of this study are primarily theoretical. This research shows that sentiment analysis on parliamentary transcripts could yield interesting results. In order to verify the findings, future research could focus on annotating the dataset that was created in this study. The dataset could be labelled manually or using different sentiment analysis models. For manual labelling, follow the annotation guidelines in Abercrombie & Batista-Navarro (2018b): have multiple annotators – all L1 Dutch speakers, university graduates, and being familiar with Flemish politics and the Flemish Parliament – annotate a large enough part of the dataset to fine-tune BERT for sentiment analysis. To label a dataset using sentiment analysis models, take multiple models and use a majority vote to determine the sentiment of each text fragment. If the fine-tuned model can predict the sentiment more accurately, it can be used to validate the findings of this study.

More contribution can be made when researchers from computer science and political science backgrounds work together. If more cross-over is created, when working together political scientist and computer scientist could push the research into the analysis of parliamentary debates forward (Abercrombie & Batista-Navarro, 2020b). As mentioned, accurately analysing these large bodies of text could drastically increase political transparency, which in turn is likely to serve more citizen involvement. This research is relevant for society because it shows that even though it is possible to extract government data, this does not necessarily result in increased transparency, as only a select group of people can gain insights from this data. A way to improve transparency is also to publish analyses of the data (e.g. dashboards) so that it appeals to the general public instead of to the scientific literature only.

References

Abercrombie, G., & Batista-Navarro, R. T. (2018a). 'Aye'or'No'? Speech-level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts. In *LREC 2018, Eleventh International Conference on Language Resources and Evaluation*. European Language Resources Association.

Abercrombie, G., & Batista-Navarro, R. (2018b). A sentiment-labelled corpus of Hansard parliamentary debate speeches. *Proceedings of ParlaCLARIN. Common Language Resources and Technology Infrastructure (CLARIN)*.

Abercrombie, G., & Batista-Navarro, R. T. (2020a). ParlVote: a corpus for sentiment analysis of political debates. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 5073-5078).

Abercrombie, G., & Batista-Navarro, R. (2020b). Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. *Journal of Computational Social Science*, *3*(1), 245-270.

Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. *Journal of Marketing Analytics*, *9*(2), 118-126.

An, X., Ganguly, A. R., Fang, Y., Scyphers, S. B., Hunter, A. M., & Dy, J. G. (2014, August). Tracking climate change opinions from twitter data. In *Workshop on data science for social good* (pp. 1-6).

Artstein, R. (2017). Inter-annotator agreement. In *Handbook of linguistic annotation* (pp. 297-313). Springer, Dordrecht.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government information quarterly*, *32*(4), 399-418.

Bayerl, P. S., & Paul, K. I. (2011). What determines inter-coder agreement in manual annotations? A meta-analytic investigation. *Computational Linguistics*, *37*(4), 699-725.

Casal Bértoa, F., & Rama, J. (2021). Polarization: What Do We Know and What Can We Do About It?. *Frontiers in Political Science*, *3*, 56.

Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics*, *11*(3), 374.

Chowdhary, K. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649.

Cloud Natural Language API. (n.d.). Google Cloud. Retrieved September 27, 2022, from https://cloud.google.com/natural-language/docs/reference/rest

Dahal, B., Kumar, S. A., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social network analysis and mining*, *9*(1), 1-20.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.

Dunlap, R. E., McCright, A. M., & Yarosh, J. H. (2016). The political divide on climate change: Partisan polarization widens in the US. *Environment: Science and Policy for Sustainable Development*, *58*(5), 4-23.

European legislation on open data. (2022, May 19). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data.

Fan, L., White, M., Sharma, E., Su, R., Choubey, P. K., Huang, R., & Wang, L. (2019). In plain sight: Media bias through the lens of factual reporting. *arXiv preprint arXiv:1909.02670*.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82-89.

Fisher, D. R., Waggle, J., & Leifeld, P. (2013). Where does political polarization come from? Locating polarization within the US climate change debate. *American Behavioral Scientist*, *57*(1), 70-92.

Freudenburg, W. R., & Muselli, V. (2010). Global warming estimates, media expectations, and the asymmetry of scientific challenge. *Global environmental change*, *20*(3), 483-491.

González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. *arXiv preprint arXiv:2005.13012*.

Han, S. (2022). Elite Polarization in South Korea: Evidence from a Natural Language Processing Model. *Journal of East Asian Studies*, *22*(1), 45-75

IPCC, 2007: *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K. and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.*

DTAI-KULeuven/robbert-v2-dutch-sentiment Hugging Face. (2022, November 18). https://huggingface.co/DTAI-KULeuven/robbert-v2-dutch-sentiment

gilesitorr/bert-base-multilingual-uncased-sentiment-3labels Hugging Face. (n.d.). https://huggingface.co/gilesitorr/bert-base-multilingual-uncased-sentiment-3labels

nlptown/ber-base-multilingual-uncased-sentiment Hugging Face. https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

Kulin, J., Johansson Sevä, I., & Dunlap, R. E. (2021). Nationalist ideology, rightwing populism, and public views about climate change in Europe. *Environmental politics*, *30*(7), 1111-1134.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Lin, W. H., Wilson, T., Wiebe, J., & Hauptmann, A. G. (2006). Which side are you on? Identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)* (pp. 109-116).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Nair, A. J., Veena, G., & Vinayak, A. (2021, April). Comparative study of twitter sentiment on covid-19 tweets. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1773-1778). IEEE.

Nws, V. (2012, May 30). *Project Uplace krijgt milieuvergunning*. vrtnws.be. https://www.vrt.be/vrtnws/nl/2012/05/30/project_uplace_krijgtmilieuvergunning-1-1314711/

Open Government Data - OECD. (n.d.). https://www.oecd.org/gov/digital-government/open-government-data.htm

Overzicht van de Vlaamse regeringen sinds 1981. (n.d.). www.vlaanderen.be. https://www.vlaanderen.be/organisatie-en-werking-van-de-vlaamse-regering/overzicht-van-de-vlaamse-regeringen-sinds-1981

Schaller, S., & Carius, A. (2019). Convenient truths. *Mapping climate agents of right wing populist parties in Europe. Berlin*.

The Paris Agreement. (n.d.). United Nations. https://www.un.org/en/climatechange/paris-agreement

Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts.

Turian, J., Ratinov, L., & Bengio, Y. (2010, July). Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 384-394).

Vachudova, M. A. (2019). From competition to polarization in Central Europe: How populists change party systems and the European Union. *Polity*, *51*(4), 689-706.

de Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., & Nissim, M. (2019). Bertje: A dutch bert model. *arXiv preprint arXiv:1912.09582*.