universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Analysing Political Developments by Predicting Party Affiliations from Austrian Parliamentary Speeches

verfasst von | submitted by

Sarah Sulollari BA BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 647
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Digital Humanities
Betreut von | Supervisor:	Univ.-Prof. Dr. -Ing. Benjamin Roth B.Sc. M.Sc.

Acknowledgements

I want to thank my Professor, Ben, his Phd Student Anastasiia and Postdoc Vanja for all the discussions, advice and feedback. Further my partner Rob and his parents Melanie and Nick for proofreading. Thank you!

Abstract

This master thesis examines the speech behaviour of members of the Austrian National Council between 1996 and 2018. A fine-tuned German BERT-Model assigns every speech in the data set a probability distribution between all parties, based on the ParlSpeechV2 data set. One of the five folds is being left out during training, in order to receive a prediction for each speech in the data set. The probability distributions are aggregated on speaker and party level and chronologically examined. By visualising the results, we see the distances between MPs and the clustering of parties. Going further into detail, I looked at all the MPs that switched party, as well as the MPs that were misclassified. Instead of seeing the misclassifications as a performance issue, the thesis uses them as information about the MPs and the political developments. The work has three main insights: First, parties that govern together become more similar in their speech behaviour. Second, parties that are in opposition move further away from governing parties in their speech behaviour. Third, MPs who switch party adapt their speech behaviour usually after they switched.

Kurzfassung

Diese Masterarbeit bespricht das Redeverhalten des Nationalrats zwischen 1996 und 2018. Ein finetuned german-bert-base-cased BERT-Model, auf Basis der Reden des Nationalrats, (dem ParlSpeechV2 Datensatz), ordnet jeder Rede im Datensatz eine Wahrscheinlichkeitsverteilung aller Parteien zu. In den fünf trainierten Modellen wird jeweils ein Fünftel der Trainingsdaten nicht genutzt, um den gesamten Datensatz zu klassifizieren. Die Wahrscheinlichkeiten, zu welcher Partei eine Rede gehört, werden auf Abgeordnetenebene und Parteiebene aggregiert und im zeitlichen Verlauf untersucht. Mit Visualisierungen werden die Näheverhältnisse der Parteien dargestellt und das Redeverhalten von Abgeordneten, die Partei wechseln, untersucht. In einer vertiefenden Auswertung werden die falsch klassifizierten Abgeordneten nicht als Perfomancemangel hingenommen, sondern als Information genutzt, um mehr über die Entwicklungen in dem untersuchten Zeitraum zu verstehen. Die Arbeit zeigt, dass das Redeverhalten von Parteien in Koalition sich annähert, in Opposition entfernt und Abgeordnete meist erst nach einem Parteiwechsel ihr Redeverhalten anpassen.

Contents

Acknowledgements i

Abstract iii

Kurzfassung v

List of Tables ix

List of Figures xi

1 Introduction 1

2 Related literature 3

3 Data 9
- 3.1 Parties . . . 9
- 3.2 Time covered . . . 10
- 3.3 Sittings . . . 10
- 3.4 Speeches . . . 11
  - 3.4.1 Dealing with chairperson . . . 11
  - 3.4.2 Speeches per party . . . 12
  - 3.4.3 Speech length . . . 14
  - 3.4.4 Speech length per party . . . 14
- 3.5 Speakers . . . 15
  - 3.5.1 Members of the National Council . . . 18
  - 3.5.2 Ministers . . . 19
  - 3.5.3 Speakers activity . . . 20
- 3.6 Excursion speech-vote proportion . . . 20

4 Data preprocessing and training 31
- 4.1 Data preprocessing . . . 31
  - 4.1.1 Speakers . . . 31
  - 4.1.2 Distribution of training data . . . 32
- 4.2 Training . . . 32
  - 4.2.1 Training folds . . . 33
  - 4.2.2 Model architecture and parameters . . . 33

5 Evaluation and Results **37**
5.1 Evaluation . . . 37
5.1.1 Raw model scores . . . 37
5.2 Results . . . 38
5.2.1 Evaluation of speaker prediction . . . 39
5.2.2 Evaluation of party prediction . . . 41
5.3 Independent speakers . . . 42
5.4 Overview speakers . . . 43

6 Political developments **45**
6.1 New Parties - Old MPs . . . 45
6.1.1 BZÖ . . . 46
6.1.2 STRONACH . . . 51
6.1.3 NEOS . . . 53
6.1.4 PILZ . . . 54
6.1.5 Other MPs that switched parties . . . 55
6.2 Misclassified MPs . . . 56
6.2.1 SPÖ . . . 56
6.2.2 ÖVP . . . 60
6.2.3 FPÖ . . . 61
6.2.4 GRÜNE . . . 64
6.2.5 LIF . . . 64
6.2.6 BZÖ . . . 65
6.2.7 NEOS . . . 66
6.2.8 STRONACH . . . 67
6.2.9 PILZ . . . 68
6.3 Learnings from model scores . . . 69

7 Conclusion **73**

Bibliography **77**

8 Appendix **81**

List of Tables

3.1 Legislative periods in the Austrian Republic 11
3.2 Number of sittings per legislative period 11
3.3 Speakers containing "Schriftführer/Schriftführerin" or "Präsident/Präsidentin" 12
3.4 Yearly average speech activity per party 13
3.5 Members of National Council not in data 19
3.6 Ministers that were never members of the National Council 20

4.1 List of names that were replaced in data 32
4.2 Speech counts per party 32
4.3 Comparison of context window sizes with accuracy and weighted f1-score . 34

5.4 Average scores 38
5.5 Scores for speeches >30 terms 39
5.6 Scores for MPs 40
5.7 Scores with MPs over 11 speeches 40
5.8 Predicted party (columns) and actual party (rows) 41

6.1 Party affiliation score of misclassified SPÖ members 57
6.2 Party affiliation score of misclassified ÖVP members 60
6.3 Party affiliation score of misclassified FPÖ members 62
6.4 Party affiliation score of misclassified GRÜNE members 64
6.5 Party affiliation score of misclassified LIF members 65
6.6 Party affiliation score of misclassified BZÖ members 65
6.7 Party affiliation score of misclassified NEOS members 66
6.8 Party affiliation score of misclassified STRONACH members 67
6.9 Party affiliation score of misclassified PILZ members 69

List of Figures

2.1 Assessment ÖVP left to right 7
2.2 Assessment SPÖ left to right 8

3.1 Distribution of seats in Parliament per session 12
3.2 Contains speeches held by Parliamentary president and scribes 13
3.3 Distribution of Speeches between all parties and legislative periods 14
3.4 Average speech length with upper and lower quantile 15
3.5 Average speech length with upper and lower quantile 16
3.6 Distribution of speakers. OK stands for "without Parliamentary Group" . 16
3.7 Distribution of Speakers that switched parties. OK stands for "without Parliamentary Group" 17
3.8 Comparison of numbers of MPs joining or leaving a party. OK stands for "without Parliamentary Group" 18
3.9 Average speeches per MP 20
3.10 Disproportionality of seats to speeches 22
3.11 Disproportionality of seats to speeches 23
3.12 Disproportionality of seats to speeches 24
3.13 Disproportionality of seats to speeches 25
3.14 Disproportionality of seats to speeches 26
3.15 Disproportionality of seats to speeches 27
3.16 Disproportionality of seats to speeches 28
3.17 Gallagher index for speech-votes disproportionality over all governmental periods in the data set 29

4.1 Distribution of parties amongst the folds that were each left out 33

5.1 Relation of MPs to each other 43

6.1 Probability of speeches by the BZÖ MPs 46
6.2 Probability of speeches by FPÖ MPs 47
6.3 Probability of speeches 47
6.4 Probability of speeches being FPÖ and BZÖ 48
6.5 Probability of speeches being FPÖ and BZÖ 48
6.6 Probability of speeches being FPÖ and BZÖ 49
6.7 Probability of speeches being FPÖ and BZÖ 49
6.8 Probability of speeches being FPÖ and BZÖ 50
6.9 Probability of speeches being FPÖ and BZÖ 50

6.10 Probability of speeches being FPÖ and BZÖ 50
6.11 Probability distribution of speeches 51
6.12 Probability distribution of speeches 51
6.13 Probability distribution of speeches 52
6.14 Probability distribution of speeches 52
6.15 Probability distribution of speeches 52
6.16 Probability distribution of speeches 53
6.17 Probability distribution of speeches 53
6.18 Probability distribution of speeches 54
6.19 Probability distribution of speeches 54
6.20 Probability distribution of speeches 55
6.21 Probability distribution of speeches 55
6.22 Probability distribution of speeches 59
6.23 Proportion of SPÖ speeches being classified as ÖVP or FPÖ 70
6.24 Proportion of ÖVP speeches being classified as ÖVP or FPÖ 70
6.25 Proportion of FPÖ speeches being classified as ÖVP or FPÖ 70
6.26 Proportion of GRÜNE speeches being classified as ÖVP or FPÖ 71
6.27 Proportion of BZÖ speeches being classified as ÖVP or FPÖ 71

1 Introduction

Research questions and motivation

Analysing political language is the subject of many fields including political science, linguistics and Natural Language Processing (NLP). Due to limited time and resources, I narrowed the study field of this thesis down to the speeches of the Austrian Parliament between 1996 and 2018. There is a data set containing all speeches of the Austrian Parliament from 1996 to 2018 that includes the date, the role of the speaker (chairperson or Member of Parliament: MP [1]), the party affiliation of the MP, the speech and the name of the speaker. This data set does not cover a comprehensive understanding of ideological leanings within language. The questions that could be solved with it concern parliamentary parties and how the individual speech behaviour of MPs corresponds to their party affiliation, and what differences there are between government/opposition and changes in parliamentary speeches over time. Based on the approximately 200,000 speeches, I will train models that predict the party affiliation of each MP and each speech. My concrete research questions are: How does party affiliation and language used in speeches correspond to each other in the Austrian Parliament? Could one predict a certain MP leaving their party or having differing views to the official party positions?

The use and relevance of this concrete field

Training a model that can predict the party affiliation based on speeches of the Austrian Parliament can help to better understand the dynamics of the Parliamentary debate. These speeches are public and represent a core part of our democratic society. Because of the vast number of speeches, quantitative approaches on the data can build a bridge from the National Council directly to the public, as well as research in other fields relying mostly on qualitative methods.

- In terms of a public interest, a trained model accompanied by an explorative visualisation could help engaged citizens explore dynamics in the Parliamentary discussions.
- In academic research both the fields of social sciences and linguistics could be interested in finding more about language change and use within the Parliament.

Overview

In this thesis I trained a model that predicts the party affiliation of a speech. As a model result for each speech, I received a probability distribution between all parties that existed

[1] In this thesis I will use the term MP, Members of Parliament meaning Members of the National Council.

in the given time span. This text classification task was solved by training five BERT models, where one fifth of the data was left out. This means that each speaker is not part of the training data in one of the five models. I then aggregated the speech predictions at both MP and party level to evaluate the model and inspect developments in the Austrian Parliament. I used various visualisations to show those developments and derived three main findings:

- 1. The speech behaviour of parties that govern together becomes more similar.
- 2. Parties that are in opposition move further away from the governing parties.
- 3. MPs mostly adapt their speech behaviour according to their party. The idea that we might be able to see if an MP is going to switch party because they always spoke differently, does not show in the data.

Another observation that cannot fully be validated with this model is the following: From 1996 to 2018 speeches overall are moving more towards the centre-right. This is not fully investigated, but there are strong hints that shifts towards centre-right are visible in the data.

The structure of my thesis starts with a brief overview of the related literature. It gives an overview of NLP literature on political speech classification. As this is a fast evolving field, I will only mention some of the relevant papers that represent the current research. I will also cover some interesting citizen science projects around the Parliament and Parliamentary work from the field of both political science and data science in Austria.

The second chapter will talk about the data set and give a descriptive overview of it. It will provide some information on the political situation of that time span and give the relevant contextual information for the later chapters.

In the chapter about data pre-processing and training, the pipeline from the original data set to my models' predictions will be explained. I will describe my choices in training and show the architecture of my model.

Evaluation will be dealt with in a separate chapter covering the raw model scores, the aggregated scores on MP and party level, as well as the independent speakers. This chapter will also present a visualisation of all MPs in relation to each other.

By looking at all individual misclassifications and all MPs that switched from one party to another, I will derive the three previously mentioned findings.

Finally, I will conclude with a chapter showing how far my thesis could answer my research question. It will point out weaknesses of the model and potential future work. Additionally, it will give an outlook on how the data could be used by political scientists for social research.

2 Related literature

In this chapter I will cover the computational linguistic related literature in depth and briefly mention works that use computational linguistic methods for political science as well as data driven citizen science projects.

Text classification is a commonly used task and well developed field of NLP that usually uses supervised machine learning. I am interested in finding the party affiliation of an MP or party based on their speeches.

Fundamentally, language models always work on the distribution of words within the corpus. By having a text classifier that predicts the probability of each MP belonging to the parties in Parliament, I am assuming that the language used in those speeches differs between the parties or politicians. This does not directly prove that political parties use "ideologically biased" language. However, it might show that there is statistically significant variation in the language use of the Parliamentary parties. The other hypothesis that will be tested is whether it is possible to train a political text classification model using cross validation. Using computational linguistic analyses in social science is titled text-as-data approach in some of the following papers. The next section will give a short introduction into text classification.

Text as numbers

As with any approach using language for computations, we need to represent text as numbers. The field of text representation is a huge and rapidly developing one. Usually the texts are split into words or subwords, called tokens. There are many ways of representing words and their meanings by numbers. This idea was first introduced by Harris in 1954 as distributional hypothesis. The main concept is that words that often occur in similar contexts have similar meanings [Har54]. Jurafsky and Martin give a great overview of these developments in their book Speech and Language Processing [JM26]. The simplest way for numeric values to represent text, is to look at the co-occurrences of the words and create a matrix with all words as columns and rows. This way you have a vector for each word that counts how often a given word occurs with all the other words. In NLP literature this is called Bag-of-words (BOW) approach, because it does not consider anything else other than frequency. The representation of words into vectors is called embeddings and applies the idea that the distribution of words and their meaning are connected. The methods for representing text as numbers progressed from simple word count models to weighted models that consider not only the co-occurrences of the word and its neighbourhood, but also the distribution of words amongst all texts that are being used. One popular method to do this is term frequency-inverse document frequency (tf-idf), where the basic idea is to deal with outliers or very differently distributed words.

That could be words that occur very often but only in very few texts, or words that rarely occur but are in every text [JM26]. It was a common thing to train a language model with different kinds of representations of the words or tokens (embeddings). The two mentioned above were sparse matrices, meaning that there were many empty cells in the matrix. Before the advances in the architecture, papers were mostly talking about how to achieve better word representations. Instead of using sparse matrix representations of the co-occurrence of words, new algorithms introduced more dense matrices. Not every word needs to be represented separately, the information is reduced to fewer dimensions. The most common method and library is word2vec [MSC+13] [MYZ13]. This creates word embeddings that are static and represent the word and its overall context. Another development allowed a word to have different vectors depending on the concrete context they are used in. They change dynamically depending on the sentence or text in which they are being used. Those developments will be covered in this chapter.

Text classification

Logistic Regression

A common and longer established way to tackle the task of text classification is using Logistic Regression. Compared to Large Language Models (LLMs) these model types still have the advantage of being fully transparent, as we can follow the steps one by one and directly map the decisions on which class was predicted to the weights in the model. Originally I tried using scikit-learn Logistic Regression. My input was the tf-idf values for each speech. Even today, Logistic Regression has its relevance and is used commonly as it is still a main building block of further developed methods like neural networks.

LLMs

With the advance of Large Language Models (LLMs), the transformer architecture was introduced and embeddings developed from vectors that would always be the same values, to more dynamic vectors that could consider the context words more flexibly by using a method called attention mechanism [VSP+23]. An LLM usually entails an encoder and a decoder. The encoder learns vector representations for each word or token based on the input training texts it is given. The decoder takes an input, for example the beginning of a sentence or a question and creates a text by predicting the next word or token one at a time. The basic mechanism of generative LLMs is to generate text based on the words that occurred before in that sentence, or the given prompt. This type of language model is called an autoregressive language model because it is based on the given words.

BERT and MMLs

We can learn the representation of words by considering all words before and after the given word. The encoder does not need to be autoregressive, it can learn the representation of words by considering all words before and after the given word. The Bidirectional Encoder Representations from Transformers (BERT) introduced the idea of training an

LLM by masking out words within the text and finding predictions for those masked words, rather than predicting the next word given the words before [DCLT19]. Models like that are called a masked language model (MLM) and are often better suited for analysing text. This masked language modelling can be used to learn more complex representations of words and was introduced for German by deepset in 2019 [CSM19], and by other further works in 2020 [CSM20], and most recently some advances in 2024 [SFT+24]. In this thesis I am using this type of model, the bert-base-german-cased by deepset, to analyse the speeches of the Austrian National Council.

These pretrained models can be fine-tuned to a specific task by training another layer with labelled data for a specific task. In this case the task is the text classification, or more specifically the classification of speeches to parties.

BERT, LLMs or traditional approaches

There have been lively discussions around using fine-tuned BERT or Logistic Regression Models. Some examples are the following papers, mentioned later too by Reinig et al. (2024) [RRP24]. Other papers comparing BERT and traditional approaches to text classification have been around since BERT was used widely, such as [GMGBGC23] and [RRP24]. All these papers promoted the use of fine-tuned BERT for text classification. In contrast to that stance, the paper published by Pätz et.al. 2025 [PBS+25] uses the approach of creating customised embeddings trained on the German Parliamentary speeches (Bundestag) by using word2vec for topic modelling and tf-idf for sentiment analyses. They emphasise the importance of using high quality training data that is manually labelled by experts, as well as the customised use of from scratch trained models for working with political speeches. In their view, classifying speeches in Parliament needs to consider the language register, including the grammar and rhetorical peculiarities of political speech. They acknowledge, in addition to the amendable model performance, limitations that are significant for many research works in the field of social sciences, or when working with experts. Since it is a manual labelling process, even with the utmost care, it remains subjective to a certain extent [PBS+25]. Arguably, this is a problem that is less of an issue for the work of this thesis, as I am using the political party affiliation that was noted down in the Parliamentary protocols, which we can take as an objective ground truth. Still, a parallel to my research is that the sentiment distribution over time shows strong patterns of shifts in tone correlated to shifts in government and opposition, as seen in Figure 2 of that paper [PBS+25].

LLMs versus BERT

After the introduction of even more large scale language models, there was a shift to use generative LLMs for problems like text classification as well. An example of this is the paper about exploring elite polarisation in the European Parliament by actor and subject detection from 2025 [Iak25]. Also the paper by Vajjala and Shimangaud (2025) that gives on overview on the current state of text classification and LLMs. In their paper they experiment with different setups and compare 32 data sets in eight languages.

They conclude that apart from sentiment classification, for now supervised training and fine-tuning still outperforms the use of LLMs [VS25].

The community has now started to compare the performance of big LLMs and fine-tuned BERT models. As summarised in the papers below, the current research shows that there are many text classification cases where fine-tuned BERT models might be the better solution, both in performance and efficiency. In the paper by Zhang et al.(2025) they used both BERT and LLMs for six different text classification tasks and found that BERT-like models perform better when contextual knowledge is needed [ZHL+25]. The authors argue that LLMs do not have that contextual knowledge and perform better in tasks that need a more common sense understanding of the world.

Political sciences use of NLP

Political science has leveraged those advances already when working with political speech texts. For example the paper on GREEK-BERT that fine-tuned the model on both a binary and a multiclass text classification task for party prediction [KS23]. Also the paper on stance classification in the Australian Parliament [NZY+25], where they used both BERT and tf-idf based models and showed that, for stance classification and a set of representative features, BERT is both outperformed in performance and most clearly in computational costs with the added advantage of transparency.

In the paper by Timoneda and Vera (2025), different transformer models and their use in political sciences were evaluated with an important finding for my work: The need for balanced training data for fine-tuning. This is a problem for many fields in political science, for example in cases where small parties and fewer speeches are available to train the model. The GREEK-BERT paper, as well as my work, show weaknesses with smaller parties, probably due to imbalanced training data for fine-tuning.

In Germany there were studies made on polarisation, sentiment and topic modelling as described above. Another example is a paper by Reinig et al.(2024). They used an approach from pragmatic linguistics, the concept of the speech act. They annotated all speeches according to their speech act by manually labelling a part according to their annotation scheme, and then using both a classic approach with tf-idf with a Support Vector Machine and a fine-tuned BERT with some variations. The BERT models outperformed the baseline model. One of the insights in the paper's analysis chapter was that the parties follow different rhetorical strategies depending on whether they are in opposition or in government [RRP24]. This connects to some main findings on the opposition/government behaviour, which we will see in chapter 6. Another hot topic in political science that is making use of computational linguistics and machine learning, is research on the rise and behaviour of illiberal parties as well as polarisation [WSY25].

Data visualisation

Another interesting approach that is slightly related to my work, is the combination of computational linguistics and data visualisation. The authors of GoViz propose a tool to automatically create dashboard-like visualisations that helps political scientists to get

an overview over their data and explore speeches. I would also like to give three more honourable mentions of websites enabling democratic participation and control. First, an active example for a visual tool exploring Parliamentary speeches is the parliamentfyi project [HC26]. They have a dashboard, where one can explore the speeches of the Austrian Parliament by resolution, topics, parties and persons, as well as sentiment and stance. A new feature allows citizens to combine categories and interact with the data, enabling users to answer their own research questions with the data. Second is the website "Meine Abgeordneten" [DEZ26], where users can look through all current politicians with mandates in Austria on the national and federal level, although it has no data on speech behaviour. Third is the project by the University of Graz, Politikmonitor; here the citizen gets an overview on what political promises were made and whether or not they have been implemented. Again, there are no computational linguistic methods used in this project, but rather a collection of manually labelled data. They too observe a move towards the centre-right ÖVP in terms of which agendas have been evoked or passed.

Having looked at the computational linguistic development of text classification, its implementation in political science and some examples of usage for democracy and public interest, I want to point to a panel study by the University of Vienna, AUTNES, which regularly investigates the opinions and stances on topics relevant to democracy and politics in Austria [vdTP26].

One of the results I want to highlight here are the perceived positions of the parties within the Austrian political system. Those panel studies are only available from 2017, but show a clear movement as depicted in the graphs in Figure 2.1 and Figure 2.2:

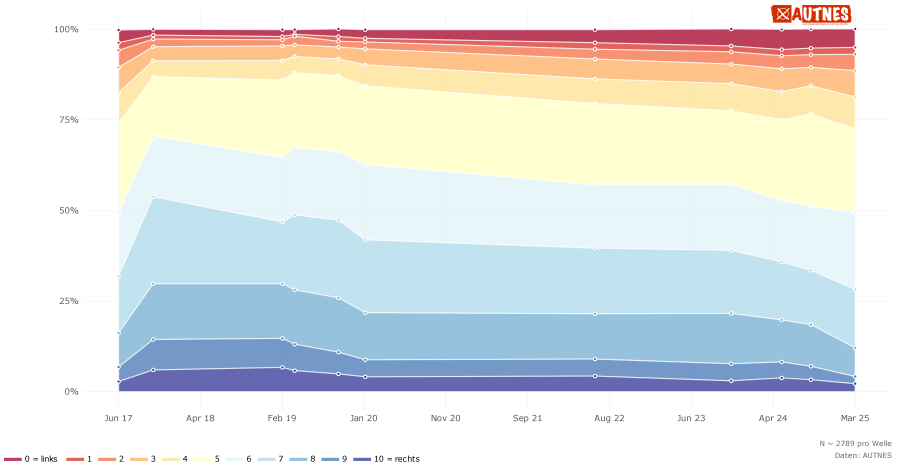

Figure 2.1: "In politics one talks about "left" and "right". Where would you put the following party, ÖVP, on a scale of 0 to 10 if 0 is "left" and 10 is "right"?"

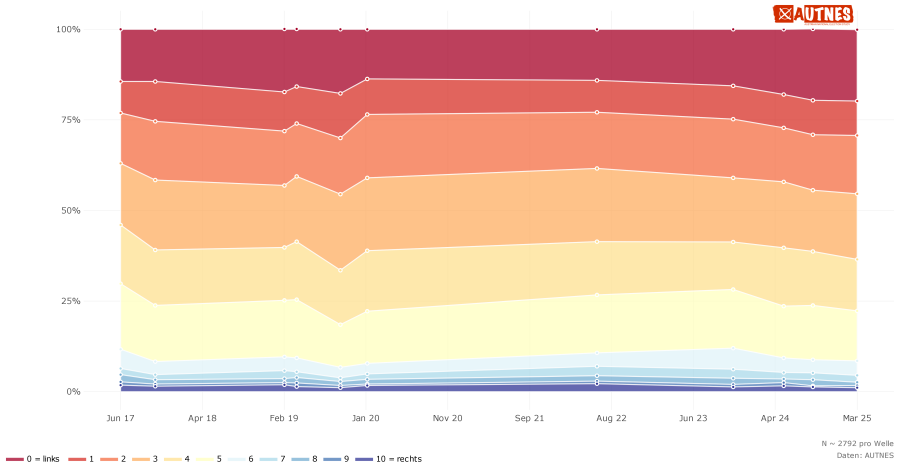

Figure 2.2: "In politics one talks about "left" and "right". Where would you put the following party, SPÖ, on a scale of 0 to 10 if 0 is "left" and 10 is "right"?"

We can see in Figure 2.2 that SPÖ is being perceived as more right in their time in Parliament as part of the government. In the same time span ÖVP in Figure 2.1 is viewed as less right.

The most relevant work for this thesis is the ParlSpeech V2 data set by Christian Rauh and Jan Schwalback: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies. This is the basis of my thesis and will be discussed in the following chapter [RS20].

3 Data

In this chapter I will give a short overview of the data set as well as some contextual knowledge important for the later chapters. First I will cover metadata around the data set as well as its structure. Further on I will talk about the political parties covered in the data as well as the amount of speakers, Parliamentary sittings and finally speeches. I will end this chapter with a short overview on the relation of party size and speech activity.

ParlSpeech V2 data set

The data set used in this thesis is the ParlSpeech V2 data set, containing speeches of nine European Parliaments, including the Austrian National Council (Nationalrat). The Austrian data was scraped from the web on 26 November 2019 and includes speeches from January 1996 until December 2018. They are available as PDFs on the Austrian Parliament home page and were automatically transcribed and then automatically annotated using an R environment. For both steps, the scripts were tested on samples across the data. The data is available as an RSD file containing a data frame with eleven variables: date shown as (YYYY-MM-DD); agenda; speech number (running order of speeches per session); speaker with slight changes/variation; party also as string character with slight variation; party.facts.id to be able to map parties that changed their name; chair a Boolean character to tell if the respective speech was given in the role of being a chairperson; terms counting the number of words; agenda a column that was added secondary to the creation of the dataset to specify the type of debate and topic; the text as a raw text only having removed the boilerplate as well as the parliament and isocountry to identify the country and parliament [1].

date	agenda	speechnumber	speaker	party	party.facts.id	chair	terms	text	parliament	iso3country

3.1 Parties

There are eight parties in the data set. SPÖ, the centre-left party, ÖVP, the centre-right party and FPÖ, the right wing party, were represented in all electoral periods. The Green Party (GRÜNE) was in Parliament except for the last XXVI legislative period, when a new party, PILZ formed by the former GRÜNE MP, Peter Pilz, entered Parliament. LIF, the neoliberal party, was in Parliament for the XX and the XXIII legislative periods, and as well as LIF-NEOS in the XXV legislative period. In the following XXVI legislative period they ran as NEOS. BZÖ was the first party to get into government without being elected.

[1] General information about the data set:`https://doi.org/10.7910/DVN/L4OAKN`.

They split off from FPÖ during the XXII legislative period. Finally, Team Stronach (STRONACH) was another Parliamentary faction that was formed before running for elections in the XXVI legislative period. The party, formed by the billionaire Frank Stronach, managed to get into Parliament with the elections for the XXV legislative period, but dissolved itself before the next election.

3.2 Time covered

The time covered in the data set contains seven legislative periods as shown in Table 3.1, of which only three had the regular length of five years[2]. The other four legislative periods ended with elections called early. To call an early election, the National Council needs to be dissolved by a simple majority in Parliament. Starting in 1986, SPÖ was in a coalition with ÖVP until the XX legislative period ended after three years and nine months. The reason the coalition ended was that the two governing parties, SPÖ and ÖVP, could not agree on a budget for the next year. After the 13 years of "grand coalition" [3] between SPÖ and ÖVP, the next two periods were the first time that FPÖ was in government. The first coalition between ÖVP and FPÖ in the XXI legislative period was shorter, due to internal conflicts within FPÖ that resulted in the end of the coalition. After the elections ÖVP and FPÖ formed another government, the XXII legislative period, which nearly lasted the full four years. The next government in the XXIII legislative period, was a coalition of SPÖ and ÖVP that only lasted two years due to conflicts between the parties. The XXIV legislative period was the only one that lasted the full five years since the legal change from four to five years. It was followed by another grand coalition in the XXV legislative period that ended due to internal conflicts in ÖVP that changed their name to ÖVP–Liste Kurz. Finally, in the last legislative period covered in the data set, FPÖ makes it back into government in a coalition with ÖVP. This government ended due to the "Ibiza scandal", in which compromising videos of the Vice President Heinz-Christian Strache were published.

3.3 Sittings

Within this time span there were 1,051 sittings which are distributed between the periods as shown in Table 3.2. The amount of sittings is relatively stable through all governments, so the difference is mostly due to the length of each period.

Each sitting has an agenda that can include different items but usually starts with a discussion on current topics. Another frequent item is the *question time* when MPs can address government Members with questions that should be answered immediately. Further possible items include *urgent questions*, *urgent motions*, as well as motions calling

[2]The ordinary legislative period was extended from four to five years in 2007, beginning with the National Council Elections in 2008.

[3]The term grand coalition, in German "große Koalition" meant the coalition between SPÖ and ÖVP. Since those two parties were traditionally the two largest parties in Austria, the term was used to describe their coalitions. This may not be an accurate term any more, as FPÖ is similarly big.

Legislative Period	start date	end date	governing parties
XX. Gesetzgebungsperiode	15.01.1996	28.10.1999	SPÖ, ÖVP
XXI. Gesetzgebungsperiode	29.10.1999	19.12.2002	ÖVP, FPÖ
XXII. Gesetzgebungsperiode	20.12.2002	29.10.2006	ÖVP, FPÖ, later BZÖ
XXIII. Gesetzgebungsperiode	30.10.2006	27.10.2008	SPÖ, ÖVP
XXIV. Gesetzgebungsperiode	28.10.2008	28.10.2013	SPÖ, ÖVP
XXV. Gesetzgebungsperiode	29.10.2013	08.11.2017	SPÖ, ÖVP
XXVI. Gesetzgebungsperiode	09.11.2017	22.10.2019	ÖVP, FPÖ

Table 3.1: Legislative periods in the Austrian Republic

Legislative period	Number of sittings
XX. Gesetzgebungsperiode	183
XXI. Gesetzgebungsperiode	118
XXII. Gesetzgebungsperiode	164
XXIII. Gesetzgebungsperiode	76
XXIV. Gesetzgebungsperiode	220
XXV. Gesetzgebungsperiode	200
XXVI. Gesetzgebungsperiode	90

Table 3.2: Number of sittings per legislative period

for a short debate on a written answer by a government Member. Additionally, MPs can request special sittings on urgent matters if requested by twenty Members of Parliament. Every sitting is chaired by the Presidents of the National Council. They are responsible for keeping the sittings in compliance with the Rules of Procedure [nRoA23]. The rules include regulations on speaking time, the order of speakers and the order of points in the agenda.

Figure 3.1 shows the distribution of MPs for each sitting. We see that independent MPs as well as other smaller emerging parties come up in the course of the sittings and do not necessarily concord with elections in 1999, 2002, 2006, 2008, 2013 and 2017.

3.4 Speeches

The following section will give a short overview on the distribution of speeches and Members of Parliament (MPs) in the data.

3.4.1 Dealing with chairperson

As shown in Figure 3.2 most speeches are held with no party affiliation, hence NA_character. When a speaker is making a speech in the role of a chairperson, the data set does not provide the party affiliation. As a chairperson, the speech is not bound to a party, but to the proceedings and rituals of Parliamentary sessions. Equally, scribes speeches noted in

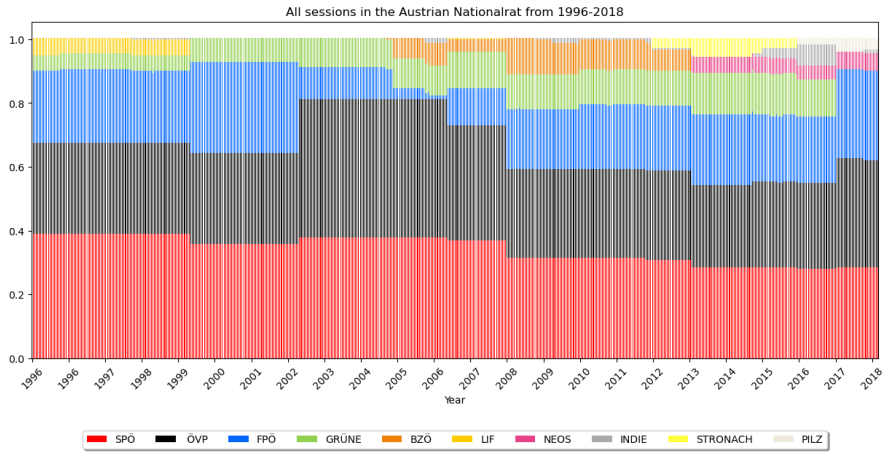

Figure 3.1: Distribution of seats in Parliament per session

the Parliamentary protocols usually include questions on the protocol and transcription. Those non-defined values are excluded from the further work as they should not be mixed up with speeches that are affiliated with a party or speeches that were held by MPs without a party. The 109,707 speeches not used further are mostly chairperson speeches, with only a few scribes speaking.

Role	Number of Speeches
Scribe	152
President of Parliament	109555

Table 3.3: Speakers containing "Schriftführer/Schriftführerin" or "Präsident/Präsidentin"

3.4.2 Speeches per party

When looking at the speeches per party in Figure 3.3, it is important to consider that the parties have varying amounts of time and seats in Parliament. Parties that were in Parliament since 1996 (SPÖ, ÖVP and FPÖ) have the most speeches followed by GRÜNE, which was in Parliament from 1996, apart from the XXVI legislative period (9 November 2017 to 22 October 2019), when they were not in Parliament. In this period, PILZ was in Parliament consisting of some former GRÜNE MPs as well as one former member of SPÖ. STRONACH only stayed in Parliament for one legislative period (XXV) but lost so many members to other parties or as independents, that by the end of the period they were no longer a Parliamentary Group. Both NEOS and LIF only took part in two legislative periods. Additionally, the number of seats each party had in Parliament differs greatly. All three top parties, SPÖ, ÖVP and FPÖ were in government for some of the time covered in the data. By looking at the yearly average speech activity per party, the predominant position of SPÖ, ÖVP and FPÖ is still strong. Table 3.4 shows the number

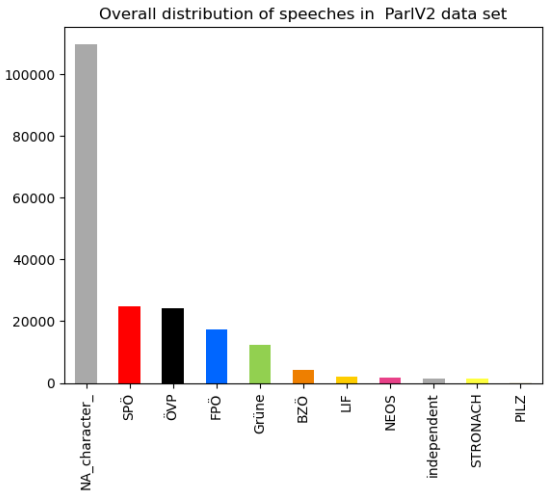

Figure 3.2: Contains speeches held by Parliamentary president and scribes

of years a party was in Parliament as well as the average number of speeches held by a party per year.

party	time in Parliament	average speeches per year
SPÖ	24 years, 9 months	996
ÖVP	24 years, 9 months	968
FPÖ	24 years, 9 months	695
Grüne	22 years, 9 months	545
BZÖ	8 years, 6 months	490
NEOS	6 years	310
LIF	6 years, 6 months	279
PILZ	2 years	172
STRONACH	3 years, 9 months	352
Independent	18 years, 9 months	83

Table 3.4: Yearly average speech activity per party

Table 3.4 above points out that both SPÖ and ÖVP speak the most, followed by FPÖ, BZÖ and GRÜNE, which each still make more speeches than LIF and NEOS. The fewest speeches are made by STRONACH, followed by the chairperson speeches that are over 13 terms. The speech activity is a normalised measure for how active each party in the data set is according to the time it spent in Parliament. Taking the length of the period the party was in the National Council (years rounded to quarter years) as N_y and the sum of all speeches of a party as N_s, we get the speeches per year per party as follows:

$$N_s y = N_s / N_y$$

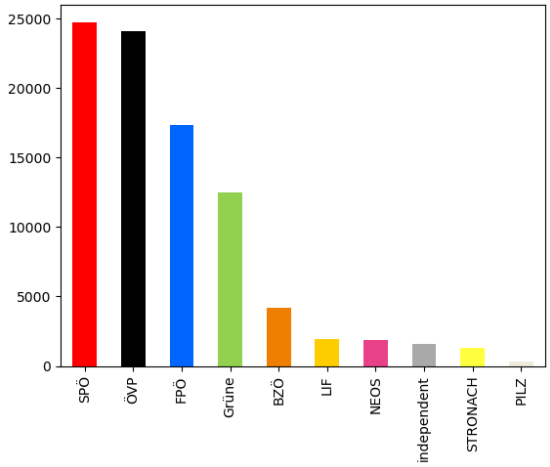

Figure 3.3: Distribution of Speeches between all parties and legislative periods

3.4.3 Speech length

Not only does the distribution of speeches vary between parties, speeches themselves also vary widely. The average length is 305 words, with a standard deviation of 528 words. The data contains speeches that have as few as one word and up to 62,490 words. The distribution of words per speech is wide ranging from the lowest quantile at 13 words and to the third quantile at 471 words, as shown in Figure 3.4. The shortest 5% of speeches contain overwhelmingly formal language such as "Ich gelobe" or other formal phrases that are part of the Parliamentary protocol.

There is no overall rule on speaking time in Parliament, it can vary from five minutes to 20 minutes depending on the point in the agenda. Sometimes the National Council can shorten or lengthen the speech time upon request. The conditions under which this can happen are also handled in the Rules of Procedure [nRoA23].

3.4.4 Speech length per party

On average LIF has the longest speeches (762 terms), followed by GRÜNE (568 terms) and NEOS (423 terms). Considering that LIF and NEOS have common roots and partially overlapping MPs, it does confirm the tendency towards longer speeches. STRONACH, GRÜNE and PILZ have a higher average speech length than the parties that were in government for at least some periods. We also see in the box plots in Figure 3.5 that the speeches in the chairperson position are usually short. As pointed out in the section on dealing with the chairperson position, they often serve a moderating function rather than actually contributing to a discussion.

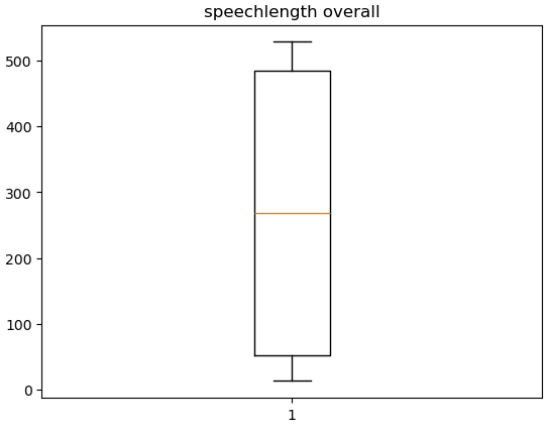

Figure 3.4: Average speech length with upper and lower quantile

3.5 Speakers

In total the data set includes 708 speakers, of which 691 are MPs. They are distributed amongst the parties as seen in Figure 3.6. The graph in Figure 3.7 shows that MPs changed their party affiliation multiple times. In total, 55 MPs changed their party affiliation, including the MPs that became independent majority members of the National Council. The highest number of MPs left FPÖ and the fewest left LIF. Most MPs left a party and became independent. The party that was most joined by other MPs was the BZÖ.

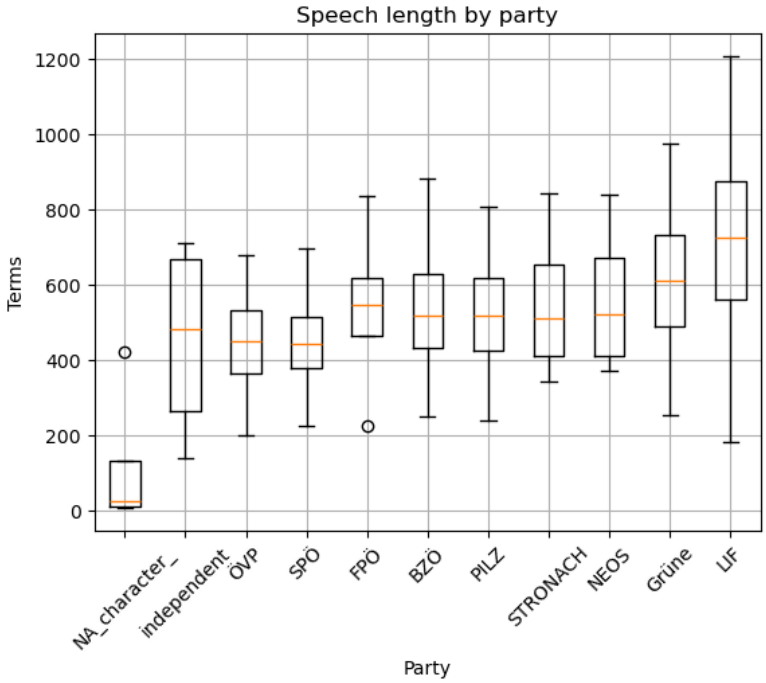

Figure 3.5: Average speech length with upper and lower quantile

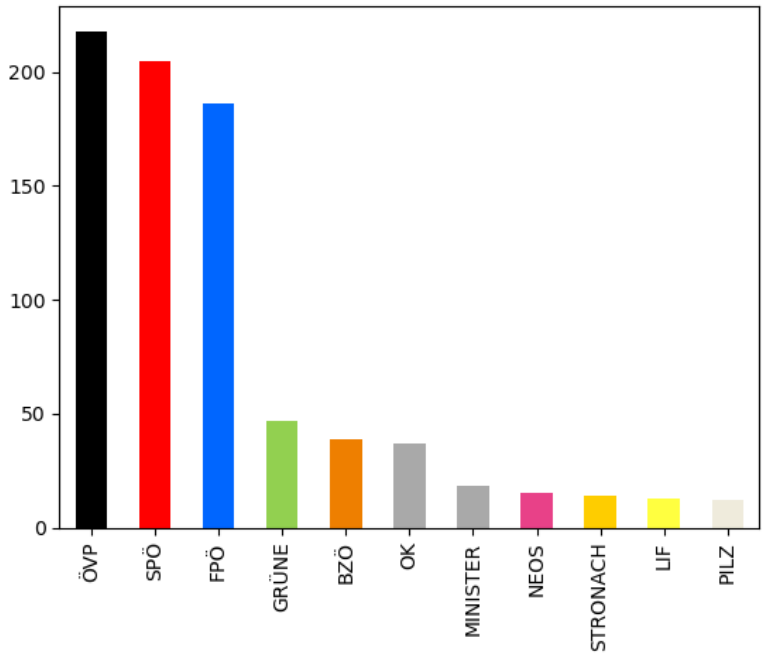

Figure 3.6: Distribution of speakers. OK stands for "without Parliamentary Group"

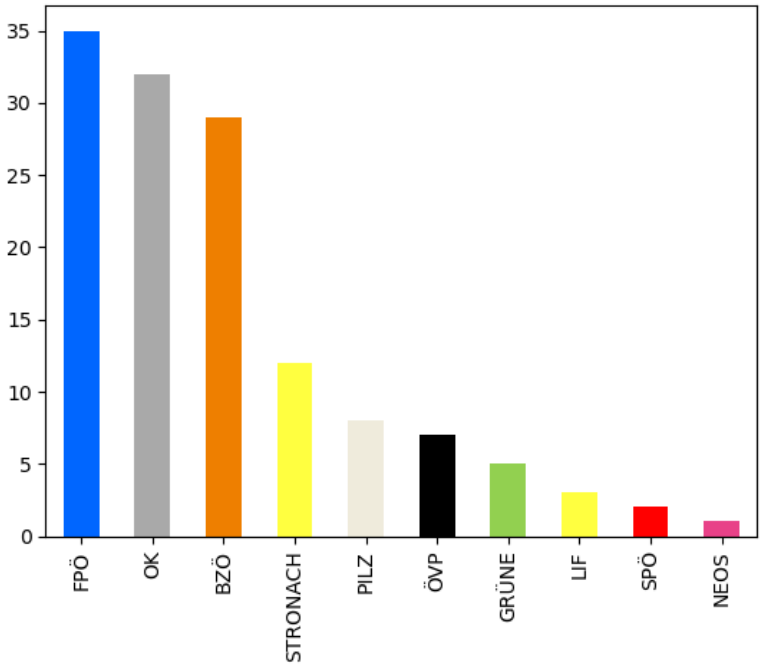

Figure 3.7: Distribution of Speakers that switched parties. OK stands for "without Parliamentary Group"

The next Figure 3.8 compares the number of MPs that joined a party and left a party. Here we can clearly see that some parties have little to no movement and some are very flexible.

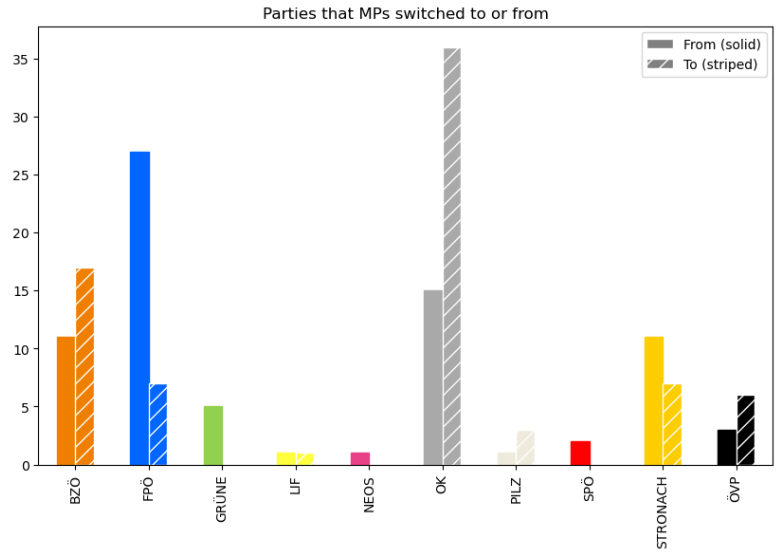

Figure 3.8: Comparison of numbers of MPs joining or leaving a party. OK stands for "without Parliamentary Group"

3.5.1 Members of the National Council

The National Council holds 183 Members. They are distributed amongst the parties according to the election results. Each party publishes lists with candidates on regional, federal and national levels. A party needs to have 4% of the votes[4] in order to enter Parliament. In the National Council a party needs five MPs to form a Parliamentary Group which can share the workload and duties of the elected Members. MPs within the National Council have a free mandate. They can follow their duties without any directive forcing them to vote or speak in a certain way. Nevertheless, there is a tradition of "Klubzwang", meaning that the members of a party faction in the Parliament are expected to vote together and support the same positions. The idea behind this is that MPs should represent the political positions of the party that the voters voted for and therewith follow the voters will.

From the XX to the XXVI legislative period inclusive, there was a total of 718 MPs in the National Council, not all of whom had speeches included in the data set. Table 3.5 lists 24 people who were MPs, but are not in the data set. Most of these were not in the data set because they were about to leave Parliament shortly after the beginning of the collected data in January 1996. Some however only joined in autumn 2018 or even later, and were therefore not in the time span of the data. A few MPs were technically in the ParlSpeechV2 data set but did not have any speeches except for "Ich gelobe", which translates to "I vow", a phrase Members of the National Council say when being appointed.

[4] or needs to win a direct mandate in a regional constituency. That has never happened, but is mathematically possible.

Member of Parliament	Party	Legislative Period in Data
Paul Rübig	ÖVP	XX.
Illona Graenitz	SPÖ	XX.
Christine Haager	SPÖ	XX.
Hilde Hawlicek	SPÖ	XX.
Heinrich Neisser	ÖVP	XX.
Ingrid Tichy-Schneider	ÖVP	XX.
Agnes Schierhuber	ÖVP	XX.
Maximilian Krauss	FPÖ	XXVI.
Klaus Lukas	FPÖ	XX.
Reinhard Rack	ÖVP	XX.
Karl Leutgöb	FPÖ	XX.
Doris Hager-Hämmerle	NEOS	XXVI.
Sandra Wohlschager	FPÖ	XXVI.
Laurenz Pöttinger	3 FPÖ	XXVI.
Thomas Dim	FPÖ	XXVI.
Lukas Brandweiner	ÖVP	XXVI.
Wilfred Tilg	FPÖ	XX.
Karl Schmidhofer	ÖVP	XXVI.
Christian Stocker	ÖVP	XXVI.
Wilhelm Niederhuemer	FPÖ	XXI.
Gerhard Abraham	SPÖ	XXI.
Konstanze Röhrs	FPÖ	XXI.
Kurt Nekula	SPÖ	XXIII.
Irina Schoettl-Delacher	FPÖ	XXI.

Table 3.5: Members of National Council not in data

3.5.2 Ministers

Some speakers are Ministers who have their party affiliation listed when speaking. One can be a Minister or be appointed a position in government without being a party member. On some occasions Ministers can speak in Parliament without speaking in their role as a Minister. In these particular cases a Minister who speaks in Parliament will have their party affiliation in the Parliamentary protocol. There are some people in the data set to whom this applies, as shown in Table 3.6.

minister	party
Andrea Kdolsky	independent, ÖVP
Claudia Plakolm	ÖVP
Claudia Schmied	SPÖ
Elisabeth Sickl	FPÖ
Erwin Buchinger	SPÖ
Gernot Blümel	ÖVP
Hannes Farnleitner	ÖVP
Hartwig Löger	independent
Heinz Grasser	independent
Hubert Gorbach	FPÖ, BZÖ
Josef Ostermayer	SPÖ
Karin Gastinger	independent, BZÖ
Liese Prokop	ÖVP
Margarete Schramböck	ÖVP
Monika Forstinger	FPÖ
Peter Doskozil	SPÖ

Table 3.6: Ministers that were never members of the National Council

3.5.3 Speakers activity

The majority of MPs in the data set has between 33 and 173 speeches. The minimum is one speech and the maximum 1,293 speeches. On average MPs have 201 speeches in the data set. This varies per party as seen in the following boxplots in Figure 3.9. GRÜNE has the highest speech activity out of all parties.

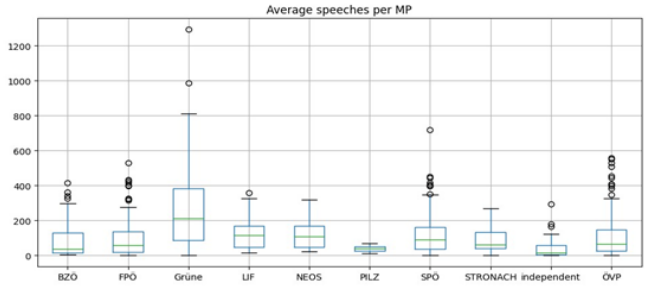

Figure 3.9: Average speeches per MP

3.6 Excursion speech-vote proportion

In order to look at how active a party is in relation to its size, I used the `poli_sci_kit` library to look at the disproportionality of speeches and seats given a party. This method

is used in statistics and political science to evaluate the relation between votes and seats in Parliament, as there are different ways to assign seats to parties. [5]

Instead of comparing total votes with seats, I compared total speeches with seats to look at how active a party is given their size. The width of the bars in the graphs in Figure 3.10 represents the percentage of speeches made by the party. Parties that are shown in the negative in the graph speak more than the proportion of seats they had in Parliament. In Figure 3.10 only FPÖ is close to being proportionally active in Parliament. Both ÖVP and SPÖ are less active according to their size and are the biggest outlier in this graph. The independent speeches are vastly overrepresented as they do not receive any votes but are either in government, or formerly affiliated with other parties. It is computed by subtracting the share of votes $V_i \sum(V)$ from the share of speeches $S_i \sum(S)$.

$$Disproportionality_{sv} = \frac{V_i}{\sum(V)} - \frac{S_i}{\sum(S)}$$

In order to look at party speech behaviour the seats per party should be compared to the number of speeches given. The dispersion of seats in Parliament and speeches is shown below in Figure 3.10 to Figure 3.16 for each legislative period.

Legislative Period XX

For the first legislative period in my data set GRÜNE as well as LIF are more active than SPÖ and ÖVP as seen in the graph in Figure 3.10. Both parties were in government which does not show in those speeches, as they are counted as independent (speeches of Ministers for example).

[5] In political science this was discussed in the paper by Gallagher, who compared used methods to assign seats [Gal91].

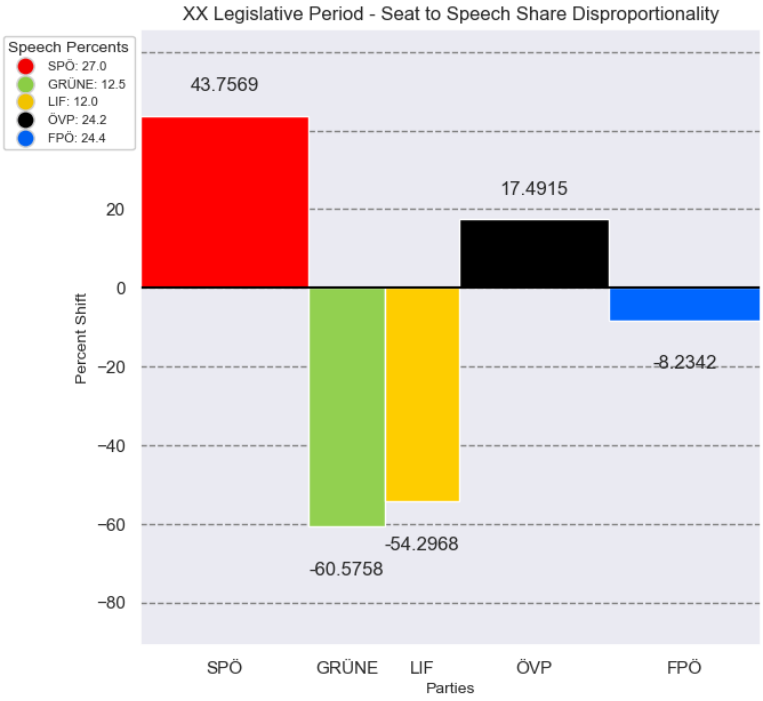

Figure 3.10: Disproportionality of seats to speeches

Legislative Period XXI

We clearly see in Figure 3.11 that in this legislative period GRÜNE took a more active role compared to the number of seats. There were only four parties in Parliament at that time, ÖVP and FPÖ constituted the government.

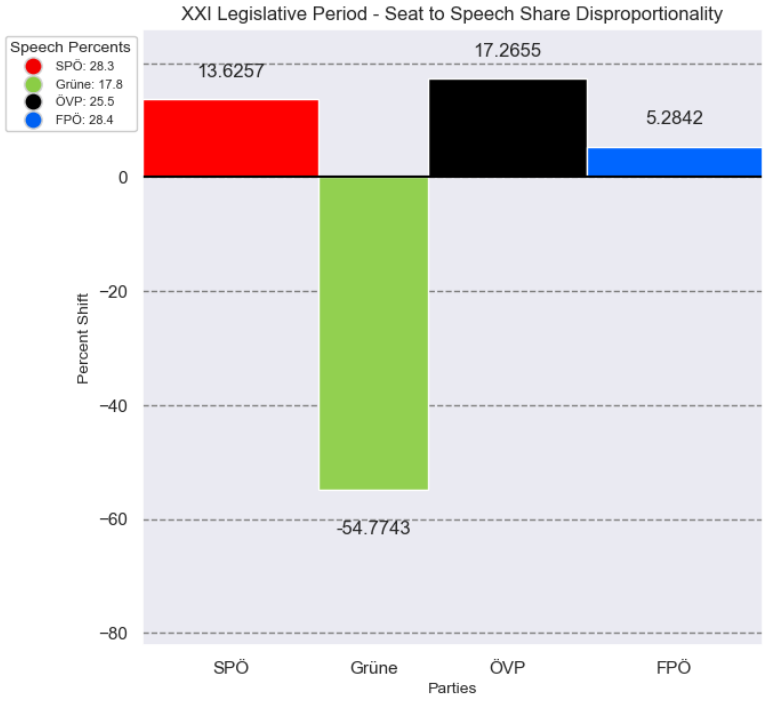

Figure 3.11: Disproportionality of seats to speeches

Legislative Period XXII

The graph in Figure 3.12 depicts the distribution of speeches according to the party affiliation by the end of that term. FPÖ split three ways into BZÖ, the independent and FPÖ. BZÖ and the independent have an especially strong disproportion to their votes as they were not voted into Parliament. This explains the high disproportionality of FPÖ. The first block of MPs leaving FPÖ was in April 2004, followed by more in 2005 and it ended with FPÖ only having one MP in Parliament. ÖVP and FPÖ were in a coalition until BZÖ replaced FPÖ in the final year. Compared to the other legislative periods, here SPÖ is speaking the least compared to the number of seats. It was also the first time they were not in government since 1970.

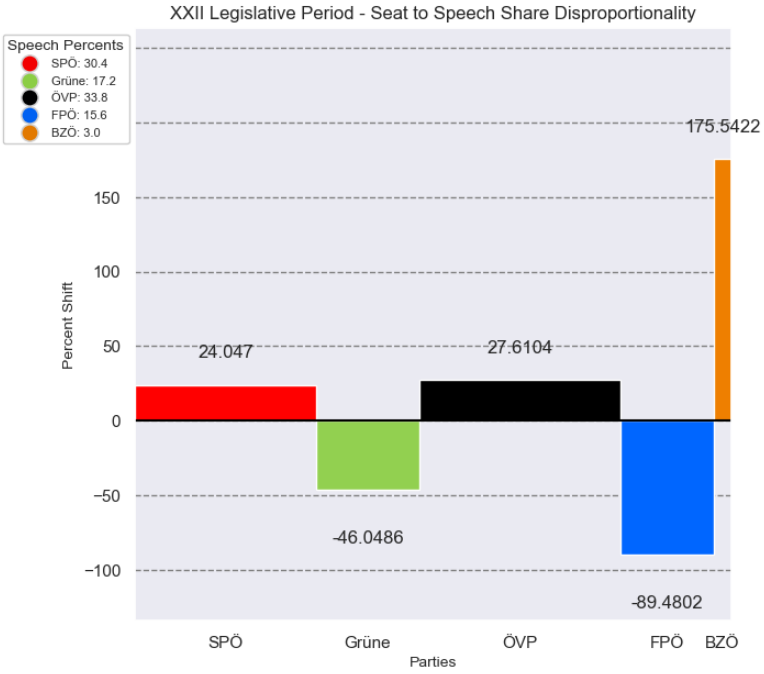

Figure 3.12: Disproportionality of seats to speeches

Legislative Period XXIII

Looking at the graph in Figure 3.13, we can see SPÖ again is speaking the least compared to the number of MPs. An outlier is an MP from LIF who was elected to Parliament for SPÖ but in agreement with SPÖ remained a LIF MP within the SPÖ Parliamentary Group. This government was a coalition between SPÖ and ÖVP. FPÖ and BZÖ made the most speeches compared to their seats in that period.

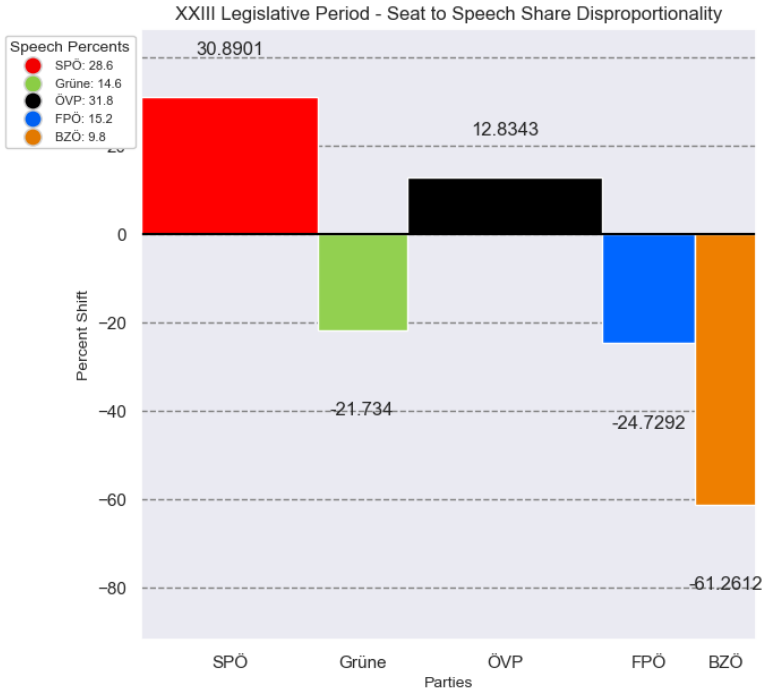

Figure 3.13: Disproportionality of seats to speeches

Legislative Period XXIV

Both parties in government, SPÖ and ÖVP, speak less compared to their number of seats. FPÖ in opposition is also less active as depicted in Figure 3.14. The two parties that make more speeches than seats are GRÜNE and BZÖ.

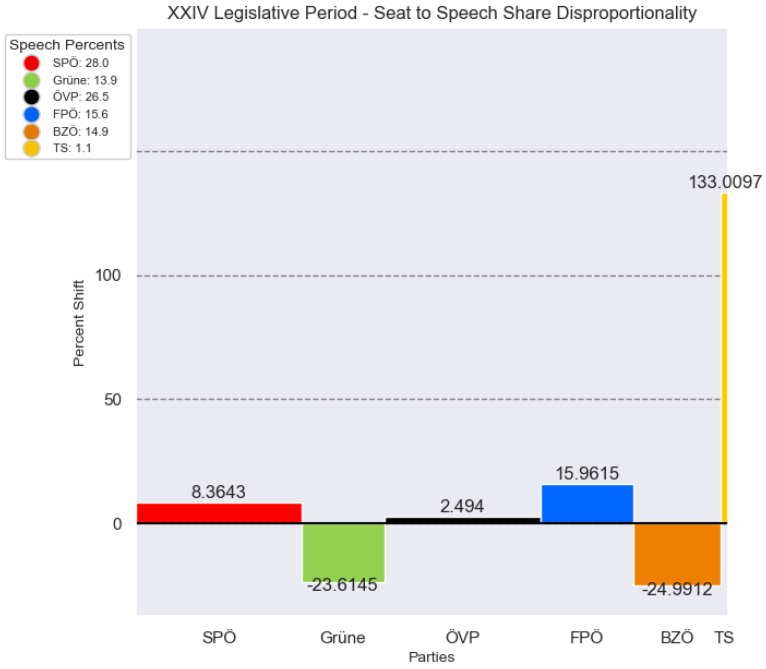

Figure 3.14: Disproportionality of seats to speeches

Legislative Period XXV

Another SPÖ ÖVP coalition, where SPÖ and ÖVP are less active in comparison to their number of seats. Here, in Figure 3.15 FPÖ is even less active according to their speeches. Also FPÖ with 40 seats is the only party not in government that is close to the two governing parties in number of seats. Two new parties come into Parliament at this election: NEOS and STRONACH. Both parties are very active.

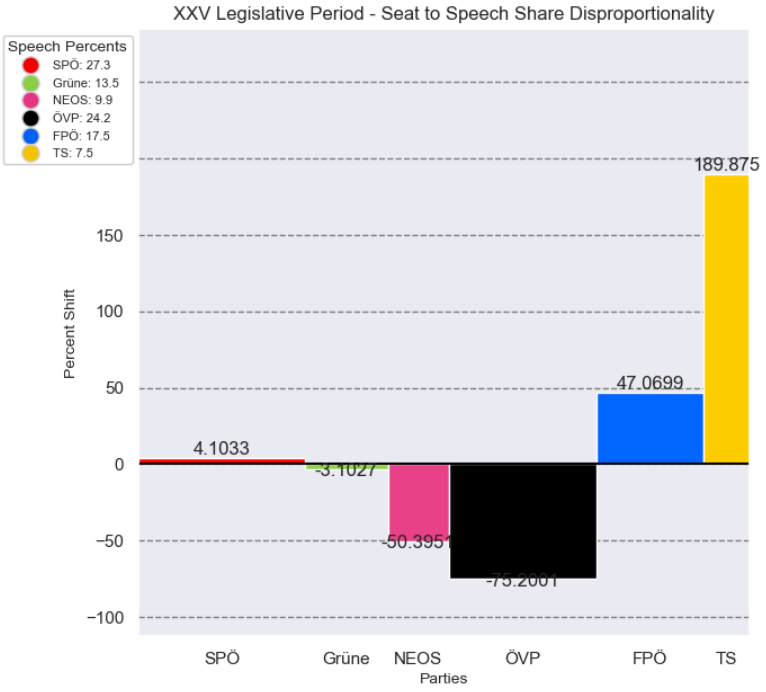

Figure 3.15: Disproportionality of seats to speeches

Legislative Period XXVI

In the last legislative period of the data set GRÜNE is not in Parliament. This is due to the split of GRÜNE into PILZ and GRÜNE. ÖVP and FPÖ are once more in a coalition, which also explains why FPÖ is again taking up more space and speeches in Figure 3.16.

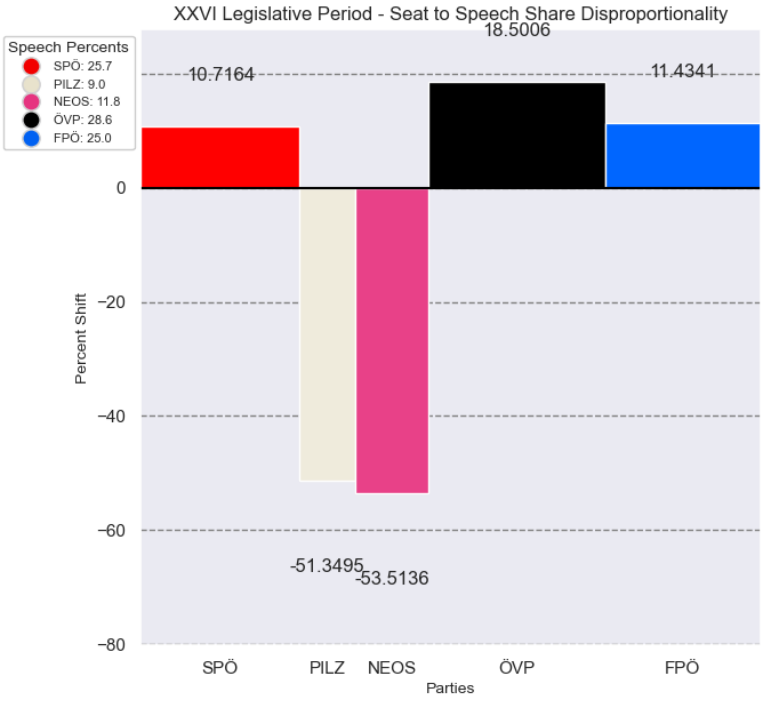

Figure 3.16: Disproportionality of seats to speeches

Overview on Speech-Votes Proportion

As depicted in the bar charts in Figures 3.10 to 3.16, the parties with most speeches take up most seats. The distribution is not directly proportional to the distribution of seats. The biggest two parties SPÖ and ÖVP take up less speeches than seats in percentage terms. FPÖ is balanced, taking up both 19% for speeches and seats over the whole corpus time. GRÜNE has far more speeches in percentage terms than seats. And all the smaller parties like NEOS, LIF, PILZ, STRONACH and BZÖ have double or more than double the percentage of speeches compared to their percentage of seats. In smaller parties individual MPs speak more often than in larger parties. Whether the parties are in government or not seems to not influence the overall speech number. One can see that by comparing SPÖ and FPÖ balances in speech behaviour.

In addition to looking at the disproportionality of speeches and seats per party, we can compare the governmental periods by using the Gallagher Index. The formula is as follows:

$$\text{LSq} = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(V_i - S_i)^2}$$

The Least square is the root of half of the sum of the squared difference between v (votes, but here speeches) and s (seats) for each party i. We can see that the disproportionality is slightly changing in Figure 3.17:

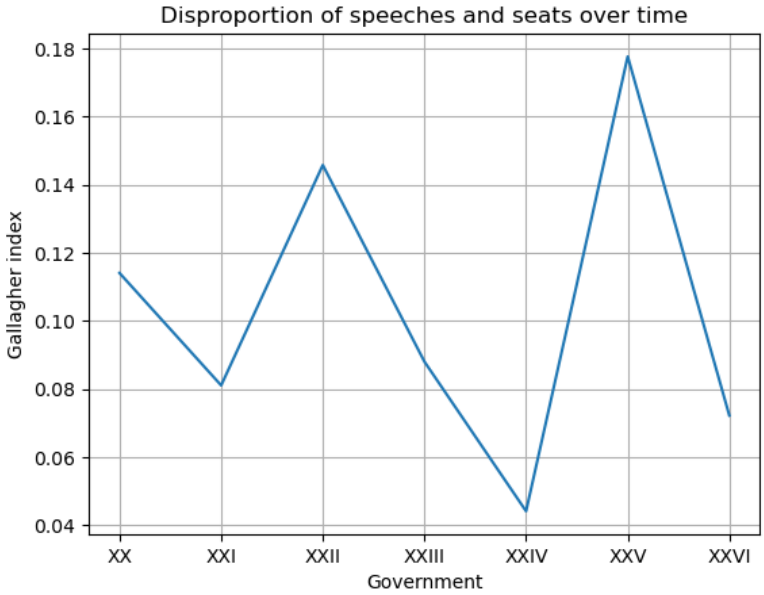

Figure 3.17: Gallagher index for speech-votes disproportionality over all governmental periods in the data set

In the XXI legislative period BZÖ was founded and in the XXV legislative period PILZ started, so maybe foundations of new parties and a higher amount of switching MPs influences the disproportionality of seats to speeches.

4 Data preprocessing and training

In this chapter I will describe what I did with the data set and how I trained my text classification model.

4.1 Data preprocessing

The data set needed to be read into Python from a ṙds file, a commonly used file type for storing R-data frames. The project that created the data set used an R-environment. I used the rpy2 library to read it into Python and saved it as a pandas data frame in a ṗkl file as it enables fast storage and reading.

For training, I used the speeches in the "text" column as well as the "party" column and the name of the speaker in the "speaker" column. I had to slightly clean the party names and speakers as described below.

Party names

In the column "party", the name of the party to which the speaker belongs at the time of the speech is noted. One party occurs with three different names in the data set: "JETZT", "PILZ" and "Jetzt-Liste Pilz", which I cleaned to "PILZ" using Regex. The speakers without party affiliation, including both the Ministers and the Members of the National Council without Parliamentary Group affiliation are noted as "independent". Those speeches were excluded from the training data as I only used speeches that had a party label.

4.1.1 Speakers

The column "speaker" contains the name of the speaker in the Parliamentary protocols. As some speakers changed their titles or roles, I extracted the last name and the name before the last name with a Regex match. Sometimes MPs would be called by their academic titles before their names, and sometimes the function they had in Parliament such as leader of a Parliamentary working group would be written down. The same speaker can occur with different written names in the protocols. After extracting with Regex there are only the first name (or the middle name) and the last name in the data. Some speakers married in the time span of my data. For those I replaced the former name with the new name listed in Table 4.1.

Name to be replaced	Name to replace
Beate Hartinger	Beate Hartinger-Klein
Karin Miklautsch	Karin Gastinger
Helmut Moser	Hans Moser
Michael Pock	Michael Bernhard
Elisabeth Aumayr	Elisabeth Achatz
Carmen Gartelgruber	Carmen Schimanek
Sylvia Prettenthaler	Sylvia Rinner

Table 4.1: List of names that were replaced in data

Some of the speakers changed their party affiliation. Those are detailed in chapter 3 and will be discussed in depth in chapter 6. That did not affect the training, as the model is trained on speech-chunk level. For training my text classification model, I excluded the independent speeches as well as the speeches by Ministers and scribes.

4.1.2 Distribution of training data

In total, I trained on 88,219 speeches distributed as listed in Table 4.2.

Party	Number of speeches	Percentage of all speeches in training data
SPÖ	24747	28%
ÖVP	24063	27%
FPÖ	17308	20%
GRÜNE	12455	14%
LIF	4170	5%
BZÖ	1942	2%
NEOS	1866	2%
STRONACH	1322	1%
PILZ	346	<0.1%

Table 4.2: Speech counts per party

As Table 4.2 shows, the most speeches were made by SPÖ, ÖVP and FPÖ, who each made between 20% and 28% of the speeches in the training data. GRÜNE are in the middle with 14% of the speeches. LIF, BZÖ, NEOS and STRONACH each made between 1% and 5% of the speeches in the training data. PILZ has by far the fewest with only 346 speeches.

4.2 Training

The goal of the trained model should be the prediction of the party affiliation based on text. Given a speech, the model should return the probability distribution amongst all parties in Parliament.

The task of the model is to classify the party affiliation of a speech based on speeches held in the Austrian Parliament.

Because the aim is to collect the prediction for every speech, a leave-one-fold-out training architecture was chosen.

In this thesis the multinomial text classification task will be solved by training five BERT models on 80% of the training data leaving out 20% each. I divided the data in folds (n=5), and trained five models leaving one fold out each.

4.2.1 Training folds

To ensure that the models are somewhat comparable, each fold needed roughly the same distribution of training data. I divided the data into five folds trying to create folds that have a similar distribution of parties compared to the data set as a whole. Additionally, I wanted to be able to predict each speaker's party affiliation. In order to be able to get a prediction, all MPs needed a model where their speeches were not included in the training data. So every MP is not in the training data of one of the five models.

First I sorted the speeches into speakers, and then I ranked the MPs with the number of speeches. Then the MPs and their speeches were looped over for each party and distributed. One by one the speeches of the MPs were distributed amongst the five folds. I ended up with folds with a distribution as follows in Figure 4.1:

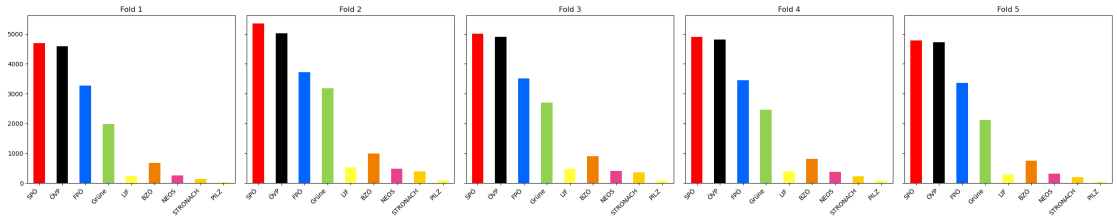

Figure 4.1: Distribution of parties amongst the folds that were each left out

For each model one fold was kept as the testing set. The other four folds were used to train the models. As one can see the smaller parties are more difficult to share between five, nevertheless each fold has every party.

4.2.2 Model architecture and parameters

After evaluating the approach of using a Logistic Regression with tf-idf, I decided to train a BERT model. Using tf-idf representation of the data resulted in MPs being classified for the topics they mostly speak about, rather than the roles they have. In the division of Parliamentary work, each MP has certain working groups and topics they are responsible for. This means that every party has someone that talks about environmental issues or foreign policies. The political speeches in Parliament cannot be compared with media or social media texts about politics. The agenda of the Parliamentary sittings dictate the topics that the speakers deal with in their speeches. Even by using elaborate word count

representations the model would classify a person that talks about a certain topic to the party that is most known for speaking about that topic. In the data, topics occur over all parties, as each party would speak in the sittings. Without consideration of syntax it is more difficult for the model to capture the correct party affiliation. BERT should be able to consider context, phrases and cover more linguistic information like syntax.

I used the hugging face PyTorch transformers library /https://pytorch.org/ to set up the training. In this thesis I fine-tuned the `bert-base-german-cased` pretrained Model. It was trained on the German Wikipedia, OpenLegalData and news articles with 12GB in total and was last updated in April 2020. The maximum sequence length for BERT is 512 tokens, including the padding tokens. It has 12 layers, as a bert-base model, and uses WordPiece-Tokenizer that was updated in 2020 because it had some problems with punctuation. There is another German BERT model by the Bavarian State Library. It seemed to me that both models performed similarly.

Context window size

Choosing the context length is a balance between efficiency and performance and for this case the length of 256 tokens was sufficient. In the following Table 4.3 I list the results for one epoch with one of my models with a 128, a 256 and a 512 context window size each. [1].

context size	accuracy	f1-score	inference time
128	0.4812	0.48	4.82 it/s
256	0.5274	0.53	2.69it/s
512	0.5705	0.56	1.58it/s

Table 4.3: Comparison of context window sizes with accuracy and weighted f1-score

The standard deviation of words lies around 528 words. Half of the speeches are under 300 words, so using a 512 token context window size would result in even more empty spaces. Roughly a quarter of speeches are between 300 and 600 words long. Because splitting is inevitable for those and the longer speeches, it is more convenient to split the speeches into smaller chunk sizes. That should give more balanced chunks than having a big chunk and a mini one. For every model I split each speech into chunks with n=256 tokens. My final training dataset consists of speech-chunks, their party label as well as an ID to aggregate the chunks back to speech level.

Model architecture

I fed the chunks to a BERT Classifier with a BERT layer, the pretrained model and a drop out layer with 10% drop out for regularisation. Finally, a linear layer with output classes for each party (n=9) to perform the classification.

[1] I kept the batch size at 20 for each run and tried it with the same data, leaving out the first fold.

```
    class BERTClassifier(nn.Module):
    def __init__(self, bert_model_name, num_classes):
        super(BERTClassifier, self).__init__()
        self.bert = BertModel.from_pretrained(bert_model_name)
        self.dropout = nn.Dropout(0.1)
        self.fc = nn.Linear(self.bert.config.hidden_size, num_classes)

    def forward(self, input_ids, attention_mask):
        outputs = self.bert(input_ids=input_ids, attention_mask=
  attention_mask)
        pooled_output = outputs.pooler_output
        x = self.dropout(pooled_output)
        logits = self.fc(x)
        return logits

```

Parameters

I used a batch size of 20 and two training epochs. With a smaller test sample I tried more epochs but there were no improvements in performance. I used the BertTokenizer for tokenization.

```
    bert_model_name= "bert-base-german-cased"
    num_classes = 9
    max_length = 256
    batch_size = 20
    num_epochs = 2
    learning_rate = 2e-5 #standard value
    tokenizer = BertTokenizer.from_pretrained(bert_model_name)


```

Training loop

The training loop passes a batch, the gradients are reset and the input is sent to CPU, then the attention_mask does the padding. The labels are sent to CPU too. In the forward pass the models logits are calculated and taken into the CrossEntropyLoss function. The loss is being backpropagated with an optimizer step and then returned.

```
  def train(model, data_loader, optimizer, scheduler, device):
    model.train()
    cross_entropy_loss = 0
    for batch in tqdm(data_loader,desc="Training"):
        # Reset gradients before first run - if training
        optimizer.zero_grad()
        input_ids = batch["input_ids"].to(device)
        attention_mask = batch["attention_mask"].to(device)
        labels = batch["label"].to(device)
        # forward pass
        outputs = model(input_ids=input_ids, attention_mask=
  attention_mask)
```

```
        loss = nn.CrossEntropyLoss()(outputs, labels)
        # backpropagation with optimizer step
        loss.backward()
        cross_entropy_loss += loss.item()
        optimizer.step()
        scheduler.step()
    cross_entropy_loss = cross_entropy_loss/len(data_loader)
    return cross_entropy_loss

```

I then collected the probability distribution for each chunk in the evaluation and aggregated the chunk predictions on speech level (creating a simple mean). The aggregated probability distribution of the speeches was linked back with my data set. The variable "predicted_party" contains the max probability party after taking the mean of all chunks. I did this for each of my five trained models and joined them back into a dataset that is used in this notebook to discuss the results. Later, I also tested the 1,555 speeches labelled as independent on all five models and took the mean of the predictions for each model to add to my final data set.

5 Evaluation and Results

In this chapter I will discuss the evaluation method, the model scores and the aggregated results. The model does not show high performance superficially. However, as I will show in chapter 6, it does perform well when discussed in detail. The t-SNE visualisation of the Parliament according to the predicted speech behaviour of the MPs will conclude this chapter.

5.1 Evaluation

5.1.1 Raw model scores

For evaluation, I used the raw logits of the prediction and normalised them with a softmax function. The output of the softmax function is the probability distribution of the parties between 0 and 1. The predicted party for a speech was then found by aggregating the raw logits and receiving a probability distribution on speech level and taking the maximum probability. I did this using the Speech-ID I created when splitting the speeches into chunks.

Those aggregated raw speech evaluation scores are not good by standard text classification metrics. FPÖ and ÖVP received the highest scores with roughly 60% F1-score, followed by SPÖ and GRÜNE. LIF, BZÖ and NEOS do not perform well with scores between 0.2 and 0.4. STRONACH and PILZ barely manage an F1-Score above 0.1.

The five models perform similarly "badly" as shown in Tables 5.1a to 5.3a below that depict each models performance as well as the last Table 5.4 showing the mean:

party	precision	recall	f1	support
SPÖ	0.55	0.61	0.58	11054
ÖVP	0.58	0.58	0.58	10516
FPÖ	0.58	0.62	0.60	9520
GRÜNE	0.59	0.44	0.50	5957
LIF	0.34	0.39	0.36	974
BZÖ	0.33	0.23	0.27	1760
NEOS	0.43	0.37	0.40	761
STRONACH	0.08	0.13	0.10	344
PILZ	0.05	0.03	0.04	115

(a) Model 1 Scores

precision	recall	f1	support
0.49	0.55	0.52	13850
0.55	0.58	0.56	12607
0.45	0.55	0.50	10135
0.59	0.37	0.45	9905
0.44	0.43	0.43	2072
0.31	0.33	0.32	2856
0.37	0.22	0.28	1339
0.15	0.10	0.12	1173
0.11	0.03	0.05	307

(b) Model 2 Score

party	precision	recall	f1	support
SPÖ	0.51	0.57	0.54	12102
ÖVP	0.54	0.56	0.55	11129
FPÖ	0.4	0.58	0.51	9927
GRÜNE	0.60	0.45	0.52	8568
LIF	0.52	0.27	0.35	1558
BZÖ	0.30	0.30	0.30	2627
NEOS	0.38	0.22	0.28	1214
STRONACH	0.25	0.05	0.09	1207
PILZ	0.06	0.05	0.05	185

(a) Model 3 Score

precision	recall	f1	support
0.53	0.56	0.55	11547
0.58	0.59	0.59	11974
0.52	0.58	0.55	9817
0.64	0.52	0.58	7571
0.47	0.23	0.31	1527
0.33	0.37	0.35	2392
0.44	0.37	0.40	1335
0.11	0.14	0.13	555
0.05	0.07	0.06	184

(b) Model 4 Scores

party	precision	recall	f1	support
SPÖ	0.55	0.54	0.54	11994
ÖVP	0.58	0.62	0.60	11160
FPÖ	0.50	0.57	0.53	9281
GRÜNE	0.59	0.47	0.52	7279
LIF	0.36	0.24	0.29	1019
BZÖ	0.36	0.34	0.35	2312
NEOS	0.28	0.29	0.28	804
STRNOACH	0.12	0.09	0.10	524
PILZ	0.16	0.11	0.13	157

(a) Model 5 Score

precision	recall	f1	support
0.53	0.57	0.55	12109
0.57	0.59	0.58	11477
0.49	0.58	0.54	9736
0.60	0.45	0.51	7856
0.43	0.31	0.35	1430
0.33	0.31	0.32	2389
0.38	0.29	0.33	1059
0.14	0.10	0.11	761
0.09	0.07	0.07	190

(b) Average scores per party

When looking at the weighted average accuracy Table 5.4, we see that the folds perform similarly with overall F1-score between 0.49 and 0.52.

fold	precision	recall	f1-score	support
1	0.53	0.53	0.52	41001
2	0.49	0.49	0.49	54244
3	0.50	0.50	0.49	48517
4	0.54	0.53	0.53	46902
5	0.53	0.53	0.52	44530

Table 5.4: Average scores

5.2 Results

For further evaluation, I excluded all speeches with less than 30 terms.

I originally did not exclude them from the data in my evaluation for my raw model scores, as it might have been helpful information on which parties are interjecting more. It could have been that those short speeches appear more often with certain parties and help classify the speech. It turned out that those speeches usually contained little information and consisted mostly of formal language.

party	precision	recall	f1-score	support
SPÖ	0.61	0.64	0.62	24218
ÖVP	0.63	0.65	0.64	23428
FPÖ	0.55	0.66	0.60	16766
GRÜNE	0.67	0.53	0.59	12093
LIF	0.51	0.37	0.43	1869
BZÖ	0.40	0.34	0.36	4108
NEOS	0.55	0.36	0.44	1833
STRONACH	0.17	0.09	0.12	1298
PILZ	0.05	0.01	0.02	338

Table 5.5: Scores for speeches >30 terms

The performance is better without those speeches as seen in Table 5.5. It improves especially for the smaller parties. This could be due to the distribution of formal speeches under 30 terms in the data. Most of them are held by parties that formed a government, which leads to the opposition parties being skewed towards them with those speeches.

Still the performance on speech level is not convincing. It might be that a single speech does not provide enough context to identify the party affiliation. Bad scores could also be caused due to common topics between parties. Parties might have overlapping positions on individual topics but on others they might disagree strongly.

5.2.1 Evaluation of speaker prediction

I aggregated the speeches with >30 terms on speaker level with a simple mean and chose the maximum probability as the predicted party. As a result 613 speakers were labelled fully correct. From the switchers, 21 speakers were partially correct, but the final party was not the predicted party. Out of those 21 wrongly labelled switching speakers, seven are completely wrong, not even predicting one of the parties the speaker belongs or belonged to. The performance immediately improves when looking at the score on this aggregated MP level, as shown in Table 5.6.

party	precision	recall	f1
SPÖ	0,89	0,89	0,89
ÖVP	0,85	0,95	0,90
FPÖ	0,86	0,90	0,88
GRÜNE	0,86	0,84	0,85
LIF	0,83	0,42	0,56
BZÖ	0,78	0,44	0,56
NEOS	0,90	0,64	0,75
STRONACH	1	0,17	0,29
PILZ	0	0	0

Table 5.6: Scores for MPs

In Table 5.6 the F1-score for all parties except PILZ and STRONACH are above 0.5, with the larger parties reaching something between 0.85 and 0.9.

After investigating the speeches on an MP level, it turned out that 74 MPs had less than ten speeches. I tried to use speakers that had more speaking time in Parliament, so I excluded the 10% of MPs with the fewest speeches. I only used speakers that had at least 11 speeches. This only improved the performance marginally, as seen in Table 5.7.

party	False Negative	True Pos	False Pos	True Neg	precision	recall	f1
SPÖ	19	169	21	425	0,89	0,90	0,89
ÖVP	10	188	29	407	0,87	0,95	0,91
FPÖ	12	136	25	461	0,84	0,92	0,88
GRÜNE	5	36	5	588	0,88	0,88	0,88
LIF	7	5	1	621	0,83	0,42	0,56
BZÖ	9	7	2	616	0,78	0,44	0,56
NEOS	5	9	0	620	1,00	0,64	0,78
STRONACH	10	1	0	623	1	0,09	0,17
PILZ	6	0	0	628	0	0	0

Table 5.7: Scores with MPs over 11 speeches

I looked at all MPs, particularly the wrongly labelled ones, in detail. The wrongly labelled speakers as well as the MPs that switched party will be discussed in depth in the next chapter. In general 613 MPs have been labelled correctly and 627 partially correct, meaning that at least one of the parties the MP belonged to was predicted.

Not considering the corrections in data quality, the model has an accuracy of 88.6%. Taking into consideration that the raw evaluation was between 50-60%, this is useful output.

Data quality

There are some errors in the ParlSpeechV2 Data Set which I checked with the stenographic protocols, that have been correctly assigned the actual party by my model. Bernhard Vock is a member of FPÖ, so is Peter Fichtenbauer. Their speeches are labelled ÖVP in the ParlSpeechV2, but the same speeches are clearly labelled FPÖ in the stenographic protocols. The same for Franz Löschnak(SPÖ) who is wrongly labelled FPÖ in the data set. Roughly 30 speeches of Johannes Moser in 2003 and 2004 are falsely labelled in the ParlSpeechV2 data set as LIF. This is an unfortunate mistake, as the Parliamentary protocols had him noted as *Hans* Moser. There is a Hans Helmut Moser, who was MP for FPÖ and LIF in the 1990s. Unfortunately, this could have been caused by "Hans" being a typical Austrian shortage for Johann, and the Parliamentary scribes just noting it down that way. Out of the 95 wrongly labelled MPs 19 had the other correct label. There are 43 MPs with mixed labels because they belonged to two or more parties or were independent and belonged to one party. So there are 76 MPs that are wrongly labelled, even though those sometimes are better than the labels of the stenographic protocol as in the case of SPÖ MP Alexander Zach, who is actually a NEOS (formerly LIF) member and never has been a member of SPÖ.

We can see that the predictions work quite well with a strong diagonal for the bigger parties. With the small parties it becomes more messy or unclear. Also, there is no MP predicted as PILZ.

5.2.2 Evaluation of party prediction

party	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	STRONACH
SPÖ	**185**	19	0	2	0	0	1	0
ÖVP	6	**211**	5	0	0	0	0	0
FPÖ	2	15	**157**	0	1	0	0	0
GRÜNE	6	1	0	**37**	0	0	0	0
LIF	3	0	2	2	**5**	0	0	0
BZÖ	0	1	8	0	0	**7**	0	0
NEOS	1	0	3	1	0	0	**9**	0
STRONACH	0	1	7	0	0	2	0	**2**
PILZ	5	0	0	1	0	0	0	0

Table 5.8: Predicted party (columns) and actual party (rows)

I further aggregated the speeches on party level. Table 5.8 is contingency table showing the actual parties in the columns and the predicted parties in the rows. We can see that the predictions work quite well with a strong diagonal for the bigger parties. For most parties it actually works well, it is only with STRONACH and PILZ that it gets more difficult for the model to classify. No MP was classified as PILZ and only two as STRONACH.

We can also see that MPs are usually classified towards the "same" political direction. Meaning that an MP from a left party is rather misclassified to another left party and an MP from a right party is rather misclassified to another right party. Also, often the parties are misclassified with their predecessor party (if they have one). For example, a LIF MP or BZÖ MP might be misclassified as an FPÖ MP or a PILZ MP as GRÜNE. An extreme example is BZÖ where most MPs were classified as FPÖ, not BZÖ.

5.3 Independent speakers

I also looked at the independent speeches which will be dealt with in this section. I predicted a party for each speech held as independent. As I did not train a class for independent speeches, my aim was to see if the MPs behaved differently when independent. Additionally, the speakers that never had a party affiliation could be politically mapped within the Parliamentary space.

Of the 33 speakers that held speeches labelled as independent, 8 were labelled wrongly. Most of them were labelled as FPÖ even though they belonged to the BZÖ or STRONACH. Only one was labelled wrongly as SPÖ and one wrongly as ÖVP. I will talk about these speakers in the next chapter.

The important information when looking at the independent speeches individually is that even though the speakers left their party or had not joined a new party yet, their speech behaviour did point to their affiliated parties.

In most of the cases of those wrongly labelled MPs at least the second-highest probability was one of the actual party affiliations of that speaker.

All in all the 1,555 speeches that were held as independent were distributed as shown in Table 5.9b. In Table 5.9a the maximum aggregated probability on MP level is listed.

party	speakers	proportion
SPÖ	3	0.09
ÖVP	7	0.21
FPÖ	16	**0.48**
GRÜNE	2	0.06
BZÖ	1	0.03
STRONACH	4	0.12

(a) Number of predictions of individual speeches

party	speeches	%
SPÖ	291	0,19
ÖVP	466	**0,30**
FPÖ	455	0,29
GRÜNE	71	0,05
LIF	2	0,00
BZÖ	115	0,07
NEOS	11	0,01
STRONACH	137	0,09
PILZ	7	0,00

(b) Distribution of speech predictions for individual speeches

It is apparent that there are more speeches that are labelled towards FPÖ and ÖVP. This might be the case because of the dynamical changes within FPÖ. We will discuss the formation of the BZÖ and their members in the following chapter. Another aspect of

analysing the independent speeches, are the ministers that have no party affiliation. They usually speak more similarly to the party they have been appointed or suggested by.

5.4 Overview speakers

Generally speeches were often misclassified towards the same political direction. Often the smaller parties were mistakenly classified as the bigger party in that political direction. We can see this by looking at the probability distribution of each speaker. In order to gain an overview of the speakers and their relation to each other I used t-SNE to reduce the distribution probability between the nine parties to two dimensions.

It is important to note that the axis labels on the graph in Figure 5.1 bear no meaningful information. Instead, the distance and relation of the speakers (here as points) to each other is important. The colours represent the parties the MPs actually belong to. To visualise the MPs that switched I chose the party they belonged to longer.

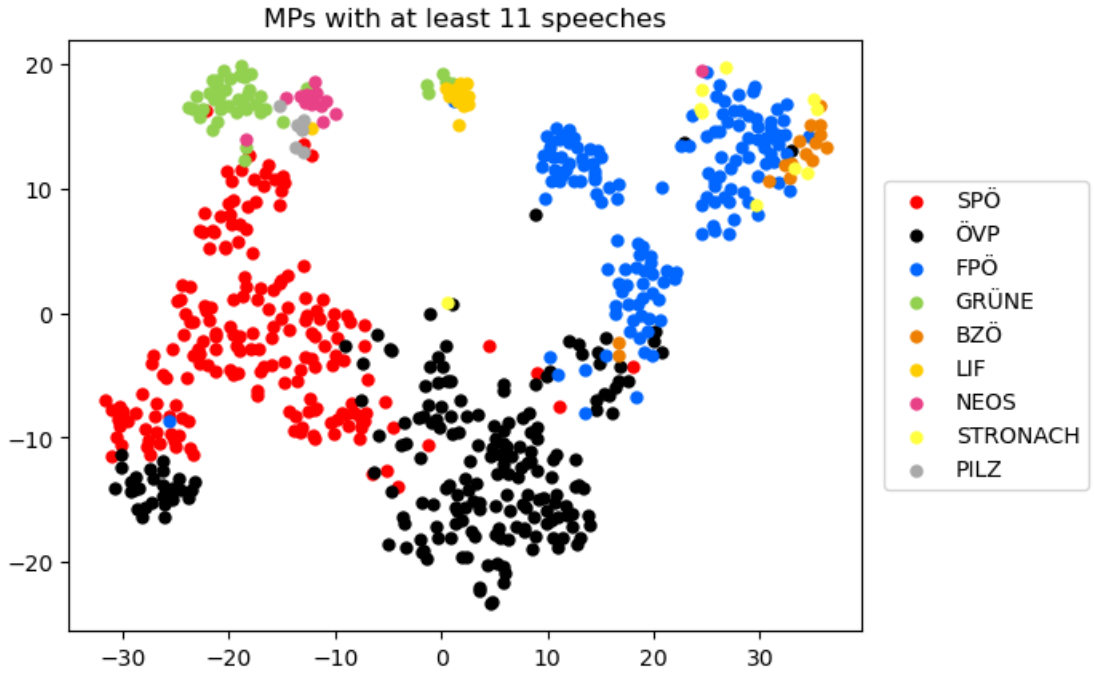

Figure 5.1: Relation of MPs to each other

We can see that the parties are clustered fairly well. The orange dots on the bottom of the t-SNE visualisation are the BZÖ, which was created by former FPÖ members. The yellow dots represent LIF, they are socially liberal but economically right wing.

The black dots that are separated by the red dots on top might be the "grand coalition" MPs of the 1990s. In that time the idea of "Sozialpartnerschaft" was very strong. It meant that both conservatives and social democrats worked for the better good. GRÜNE

is furthest from FPÖ and ÖVP which also makes sense, even though there are some more economically liberal Greens that are clustered with LIF. The dark brown dots are the STRONACH party, they blend in with FPÖ, BZÖ and ÖVP which are the parties its MPs switched from. All in all it would be interesting to have a time comparison of all those MPs per government.

6 Political developments

This chapter about political developments will cover MPs that were misclassified and those that switched party. I will conclude with the three findings central to this thesis. By looking at the speech behaviour we can observe:

- 1. MPs adapt their speech behaviour according to their party
- 2. Parties in a coalition adapt their speech behaviour towards each other.
- 3. Parties move away from governing parties when in oppositional situations.

In the first section about new parties, I will talk about MPs that left their old party for a new party. I will end that section with some MPs that changed multiple times and the MPs that were not involved with new parties, but still changed between parties. The second section of this chapter will cover the misclassified MPs on an individual level. The misclassifications will be set into context, which will prove that some of the misclassifications were actually correct. Finally, by aggregating the predictions for the speeches and considering the information about the new parties and misclassified MPs on an individual level, the listed key findings 1. to 3. will be dealt with in detail. After the first coalition between ÖVP and FPÖ, ÖVP speeches did not move back closer to its coalition partner SPÖ, causing a general move to the right in Parliamentary speeches.

Just to recap the governments in the legislative periods covered in the data: firstly there was a grand coalition between SPÖ and ÖVP, then the first coalition between ÖVP and FPÖ (black-blue), followed by another three periods with a grand coalition between SPÖ and ÖVP and finally the last period in the data is the second black-blue coalition between ÖVP and FPÖ [1]

6.1 New Parties - Old MPs

In the following section I will talk about the parties that were created by Members of Parliament founding new Parliamentary Groups, rather than being voted into Parliament for the first time. Three new parties, BZÖ, STRONACH and PILZ, were voted into

[1] To remind the reader of the approximate political positions of those governing parties; SPÖ is the social democratic, centre-left party. ÖVP is the conservative, centre-right party and FPÖ is a right-wing party. GRÜNE were in government after my data set's time span and are a left-wing environmentalist party.

Parliament after they already had already been in Parliament in the preceding legislative period without contesting an election.

6.1.1 BZÖ

BZÖ entered Parliament and government without elections. It was founded in April 2005 and was led by former FPÖ member Jörg Haider. It was the first party to be in a governing position without elections in the Austrian Republic. The Parliamentary Group changed its name from FPÖ to F-BZÖ on 28 April 2006. Not all FPÖ members switched to BZÖ, some remained FPÖ members but also were members of the F-BZÖ Parliamentary Group.

At that time FPÖ had 18 seats in the National Council and was in a coalition with ÖVP. Within one month, twelve MPs became BZÖ party members, six remained in the new F-BZÖ Parliamentary Group as FPÖ members. Three of those six left the Parliamentary Group to become independent MPs in March 2006. Here is a profile of the speeches held by FPÖ members (see Figure 6.2) and the speeches held by the split off BZÖ members, bearing in mind they shared the Parliamentary Group as F-BZÖ (see Figure 6.1).

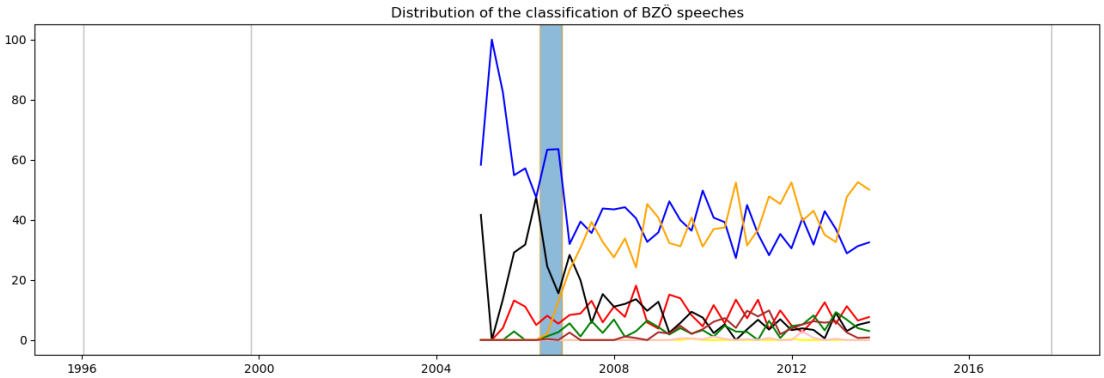

Figure 6.1: Probability of speeches by the BZÖ MPs

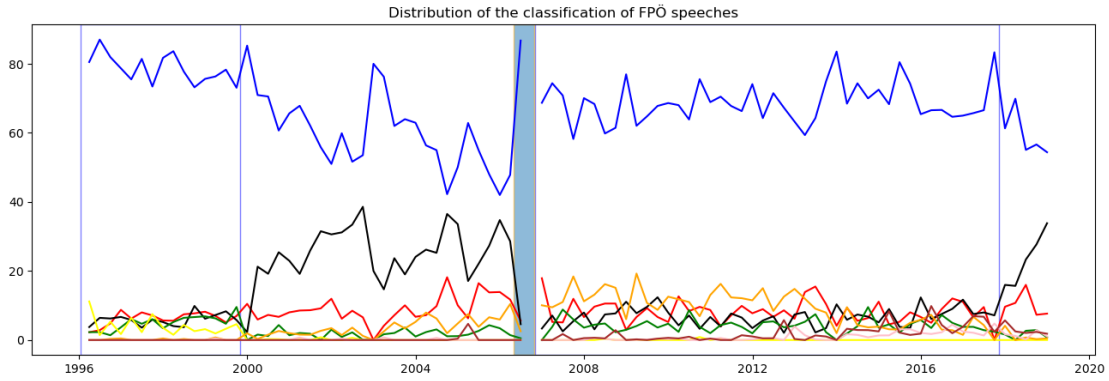

Figure 6.2: Probability of speeches by FPÖ MPs

Party founder Jörg Haider

At the time the BZÖ was founded, Jörg Haider was the Federal State Governor for Carinthia. He was in Parliament for FPÖ from 1979 until 1999.

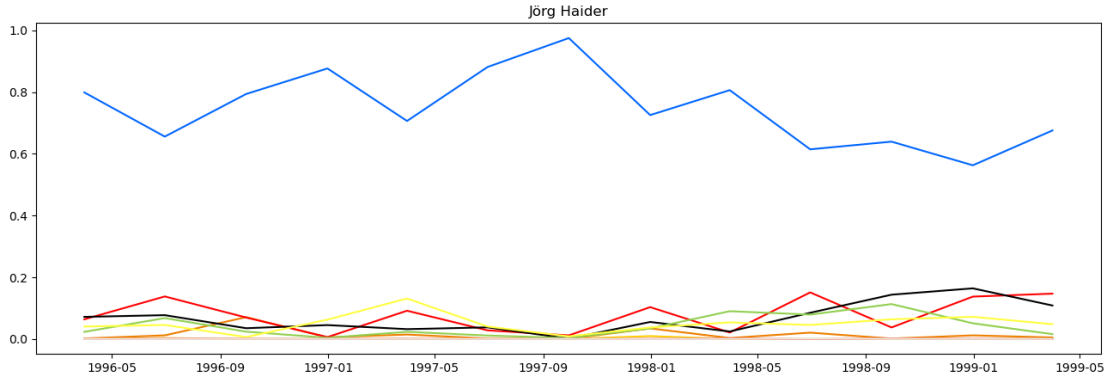

Figure 6.3: Probability of speeches

As seen in his probability scores for FPÖ in blue and the BZÖ in orange (Figure 6.3), there is very little BZÖ probability in his speeches. This differs from the MPs that joined the BZÖ in Parliament.

FPÖ → BZÖ

After the creation of the BZÖ some MPs changed their speech behaviour towards their new party.

As an example Josef Bucher, another long serving politician, tends visibly more towards the BZÖ around its creation date. That we can see as the orange line on Figure 6.4, where the orange line shows the growing probability of his speeches being BZÖ and the blue line shows the falling probability of his speeches being FPÖ.

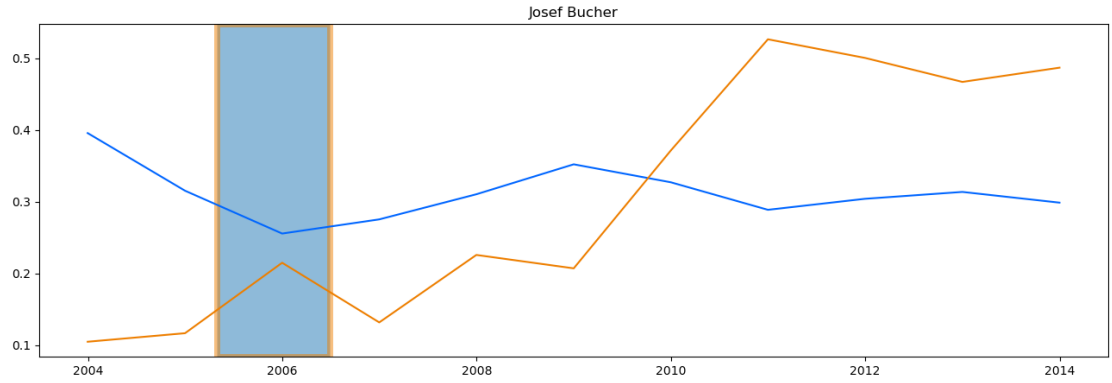

Figure 6.4: Probability of speeches being FPÖ(blue) and BZÖ(orange)

Another example for a rising BZÖ probability in speeches of an MP that switched to the BZÖ is Herbert Scheibner, whose speeches moved more strongly towards the BZÖ over time, after his switch (Figure 6.5).

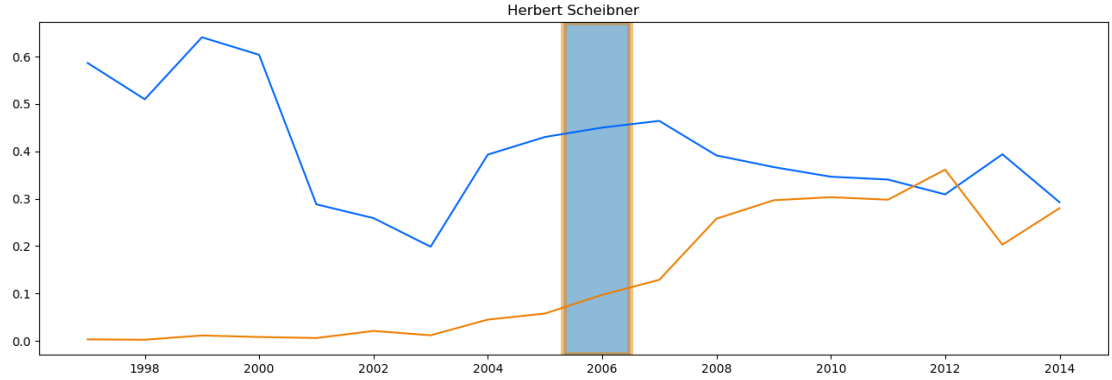

Figure 6.5: Probability of speeches being FPÖ(blue) and BZÖ(orange)

On the other hand the MPs that remained FPÖ members did not adapt their speech behaviour to that extent towards the BZÖ, like Eugen Bösch (Figure 6.6) and Barbara Rosenkranz (Figure 6.7) who left the Parliament in the next elections.

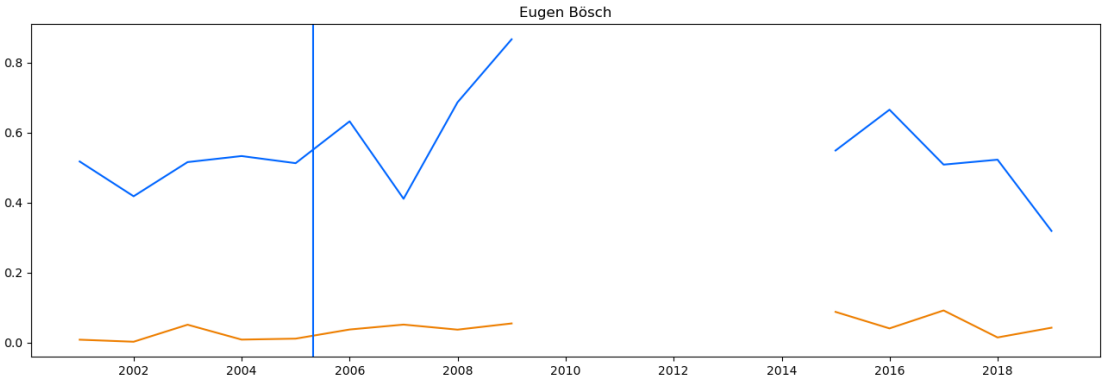

Figure 6.6: Probability of speeches being FPÖ(blue) and BZÖ(orange)

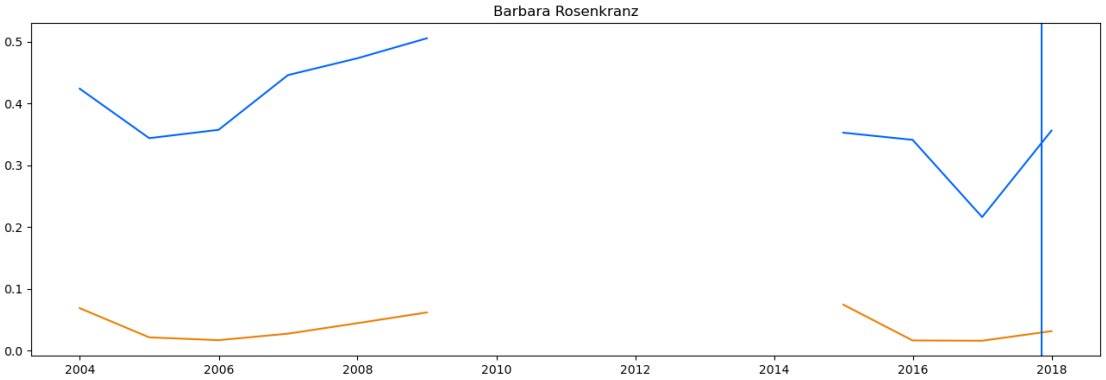

Figure 6.7: Probability of speeches being FPÖ(blue) and BZÖ(orange)

Maximilian Hofmann, Detlev Neudeck and Helene Patrik-Pablé are the three MPs that left only the Parliamentary Group F-BZÖ, but stayed FPÖ party members outside of the Parliament. Figures 6.8, 6.9 and 6.10 show changes in the party prediction of the speeches once they left the Parliamentary Group marked by the vertical line.

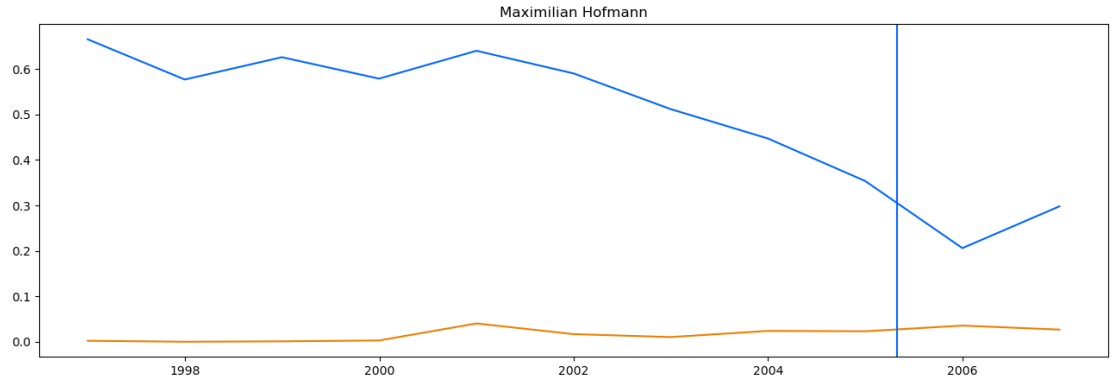

Figure 6.8: Probability of speeches being FPÖ(blue) and BZÖ(orange)

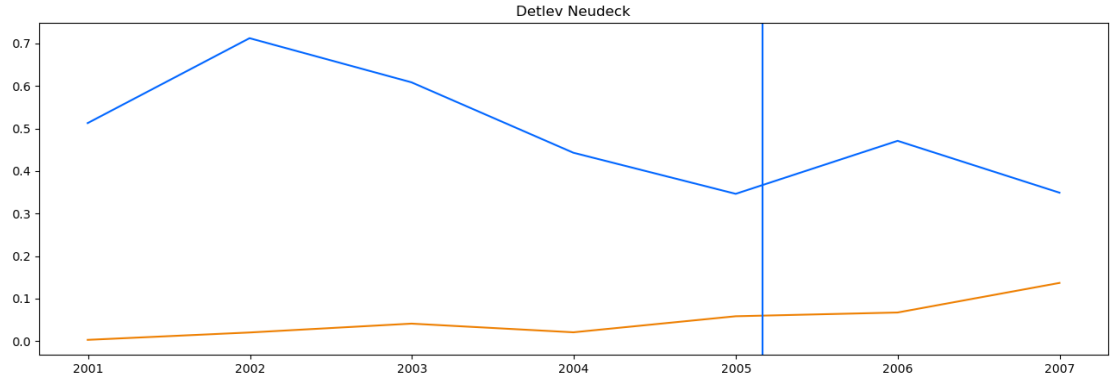

Figure 6.9: Probability of speeches being FPÖ(blue) and BZÖ(orange)

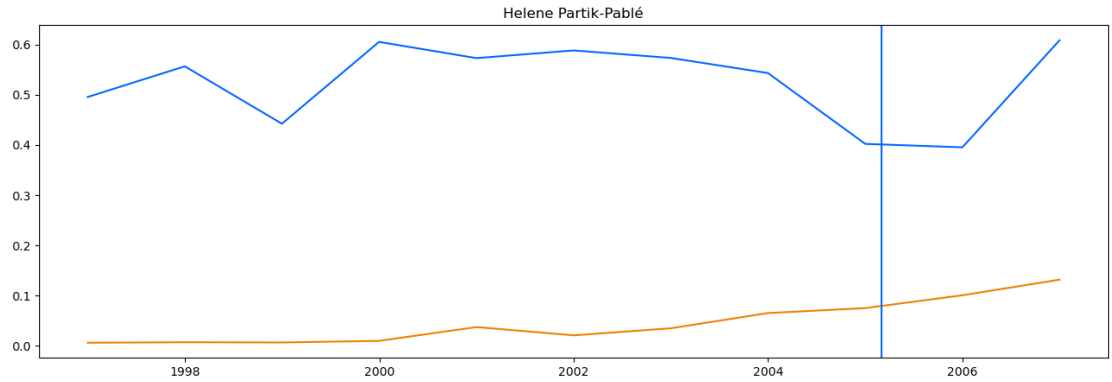

Figure 6.10: Probability of speeches being FPÖ(blue) and BZÖ(orange)

We can see that all three have a decrease in FPÖ probability before leaving their Parliamentary Group. Also there is a group of MPs who left FPÖ for the BZÖ and then

left to join another new party, STRONACH.

6.1.2 STRONACH

STRONACH is another party that first got into Parliament without contesting an election. The first session with STRONACH speeches was on 30 October 2012, a month after its formation. The Parliamentary Group started with five members [2]. They all came from BZÖ, even though two of them, Robert Lugar and Erich Tadler were independent for over one year in between. Gerhard Köfer (Figure 6.16) was another independent MP in Parliament who joined the party, but before the Parliamentary Group was formed. He left the SPÖ Parliamentary Group and became an independent MP until he left Parliament for regional politics.

The speech behaviour of the five founding members of STRONACH is shown in Figures 6.11 to 6.15:

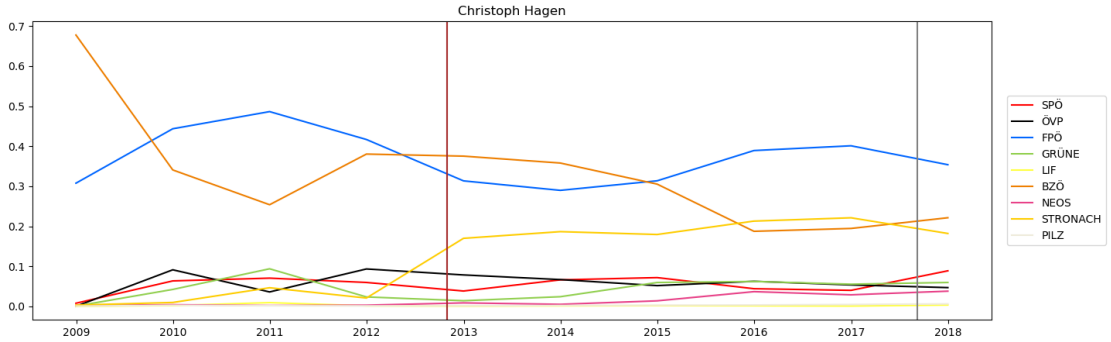

Figure 6.11: Probability distribution of speeches

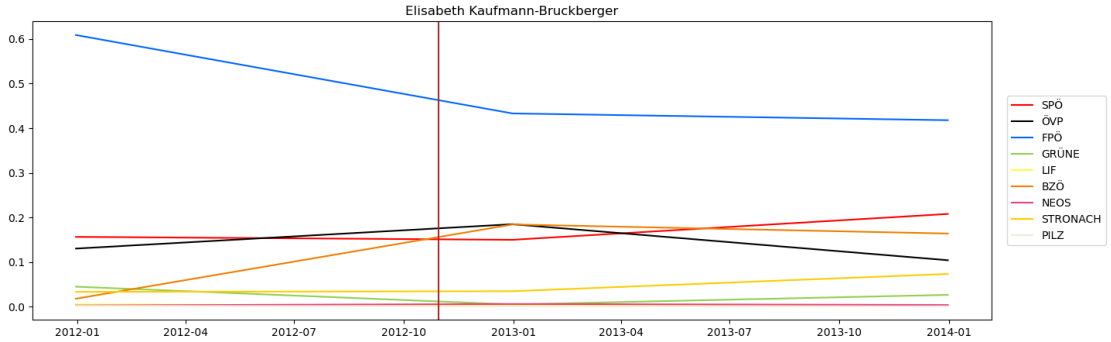

Figure 6.12: Probability distribution of speeches

[2] Christoph Hagen, Elisabeth Kaufmann, Erich Tadler, Robert Lugar and Stefan Markowitz started the STRONACH Parliamentary Group.

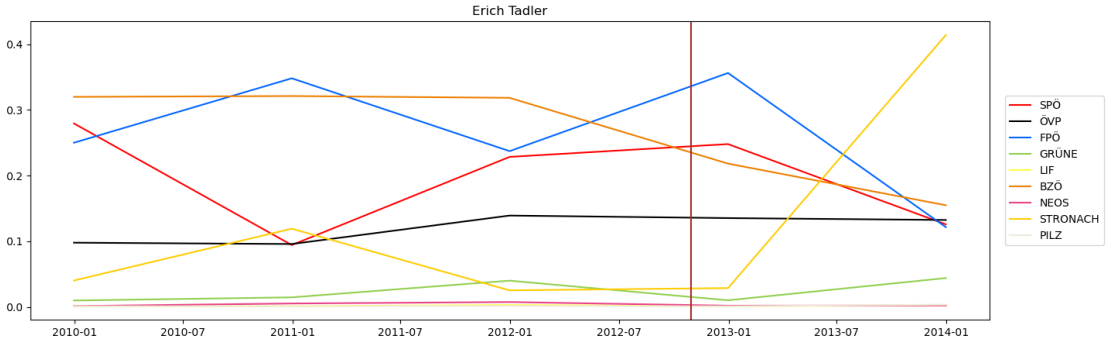

Figure 6.13: Probability distribution of speeches

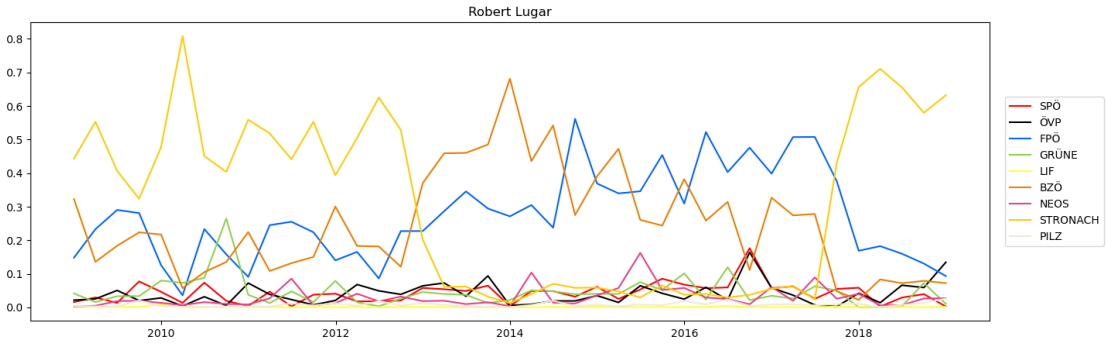

Figure 6.14: Probability distribution of speeches

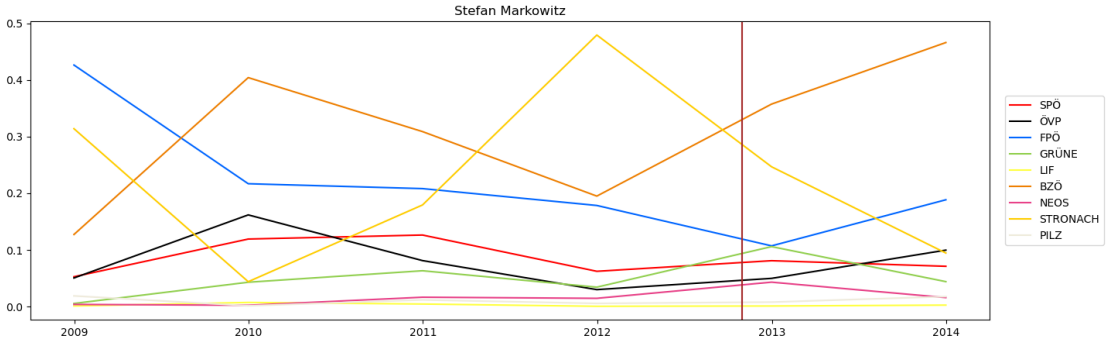

Figure 6.15: Probability distribution of speeches

We can see in the figures again that the brown line for STRONACH is rarely the highest. Generally that is due to fewer speeches being classified as STRONACH. However, in comparison to other MPs from other parties, there is at least a visible amount of speeches being classified as STRONACH. 735 speeches had STRONACH as the highest probability in total.

The former SPÖ Member, Gerhard Köfer left Parliament for a regional position before the Parliamentary Group STRONACH was formed. In the protocols and on the Parliamentary website he is quoted as an independent Member of the Parliament, even during the time he already was a party member of STRONACH.

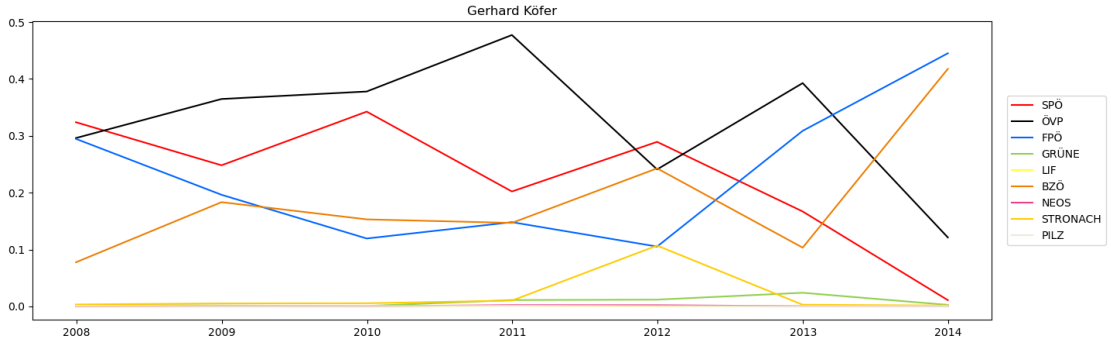

Figure 6.16: Probability distribution of speeches

6.1.3 NEOS

NEOS ran for the National Council elections in 2013, a year after they were founded. For the elections they joined up with LIF, with whom they formed the NEOS-LIF Parliamentary Group. Angelika Mlinar was the head of LIF and Matthias Strolz the head of NEOS. A year later they merged the two parties, with Matthias Strolz as the new party leader and Angelika Mlinar as the joint deputy leader alongside Beate Meinl-Reisinger from NEOS.

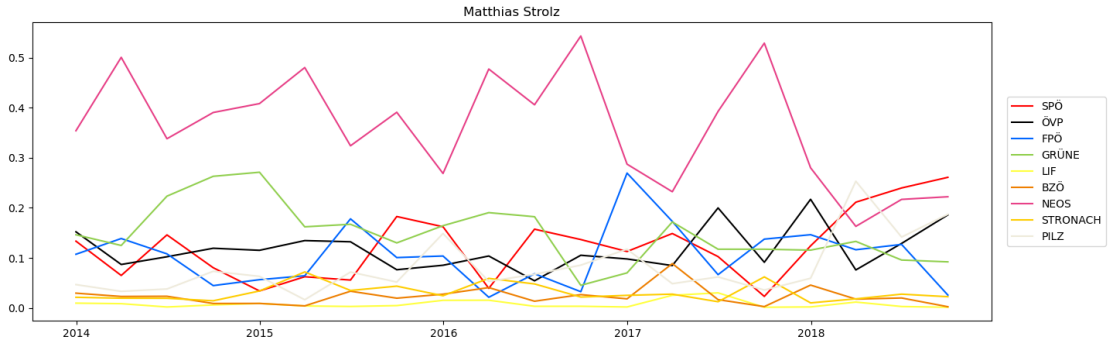

Figure 6.17: Probability distribution of speeches

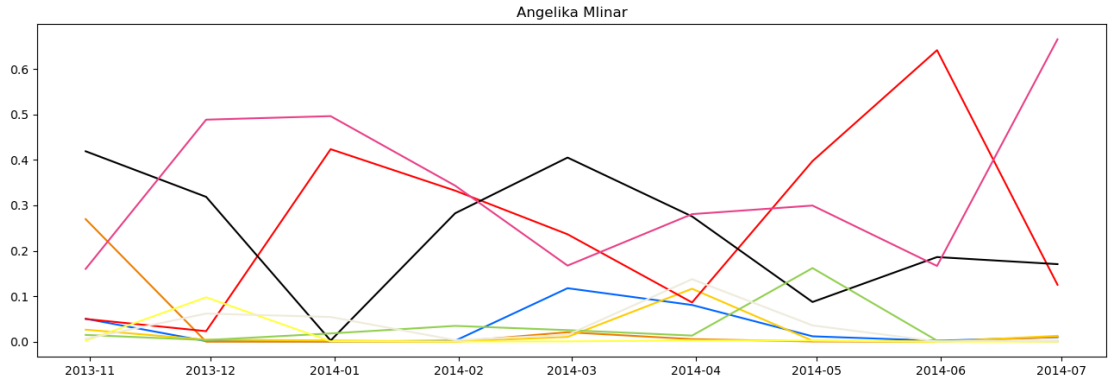

Figure 6.18: Probability distribution of speeches

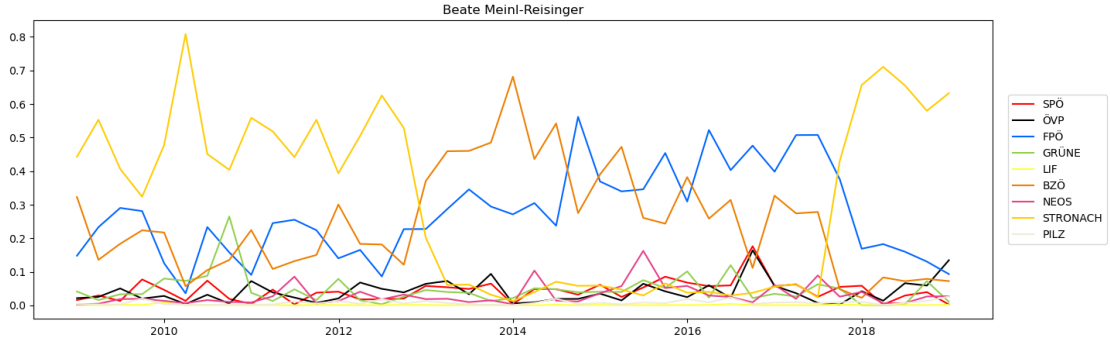

Figure 6.19: Probability distribution of speeches

Matthias Strolz (Figure 6.17) was also a member of ÖVP and their student organisation before he founded the new party NEOS.

Angelika Mlinar (Figure 6.18) was in Parliament for less than one year because in 2014 she successfully ran as the main candidate for NEOS in the EU Elections. She therefore left Parliament to join the EU Parliament. Her speech score is close between NEOS, SPÖ and ÖVP. Most speeches were labelled as NEOS, even though my model predicts her as SPÖ because on average she has a slightly higher probability for SPÖ.

Beate Meinl-Reisinger (Figure 6.19) was misclassified as she did not have enough speeches in her second period in Parliament. She shows clear strength in NEOS and GRÜNE within her speeches, also some familiarity with ÖVP, which hovers between 20% and 30% probability. She was a Member of ÖVP before she joined NEOS and was misclassified as ÖVP and FPÖ by the models.

6.1.4 PILZ

GRÜNE MP Peter Pilz (Figure 6.20) formed a new Parliamentary Group, after he lost an internal GRÜNE party election for their next electoral lists. He formed the new party

called PILZ on 25 July 2017. It should be noted that in the original data set there are seven speeches during that time, that are incorrectly attributed to GRÜNE, but in the Parliamentary protocols they are noted as independent. [3]

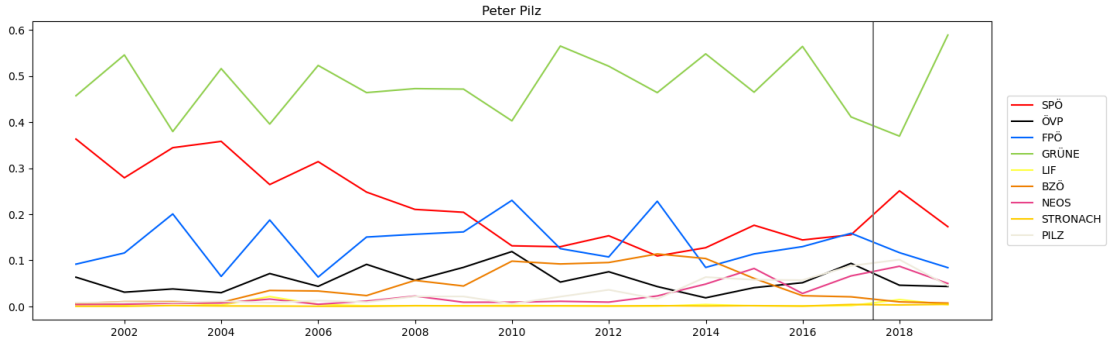

Figure 6.20: Probability distribution of speeches

One MP joined from SPÖ. Daniela Holzinger-Vogtenhuber (Figure 6.21) left SPÖ on 28 July 2017 and remained as an independent MP until she joined PILZ in the next electoral period. Two GRÜNE MPs also joined the newly formed Parliamentary Group PILZ. One of them was Wolfgang Zniggl who left GRÜNE on the same date and was elected into Parliament as a PILZ MP too. The other was Bruno Rossmann who also left GRÜNE and briefly served as an independent MP until he won a seat with PILZ at the next election.

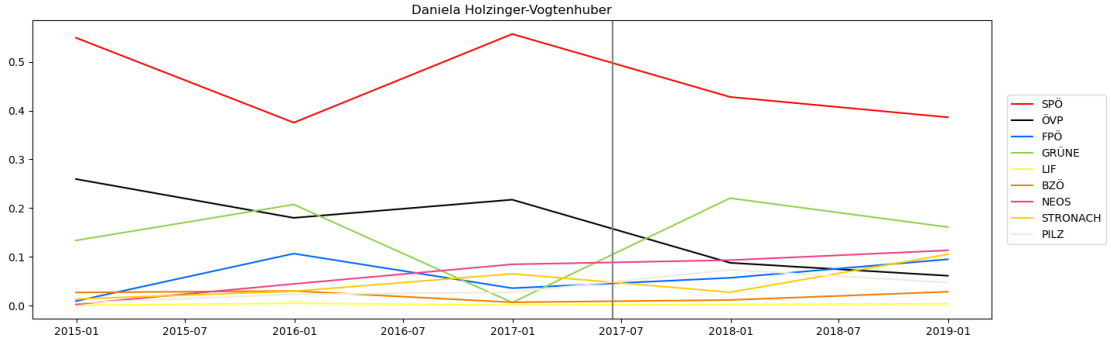

Figure 6.21: Probability distribution of speeches

6.1.5 Other MPs that switched parties

Some MPs just left their party to join other long established parties as opposed to new or failing ones. Efgani Dönmez, originally a politician for GRÜNE, won a seat as an MP for ÖVP. He was a Member of ÖVP until 4 September 2018 after which he became

[3] I checked the Parliamentary protocols of the sittings of 20 September 2017, 4 October 2017 and 13 October 2017.

an independent MP until 22 October 2019. The data set only shows his speeches as an ÖVP member or an independent member. He got completely misclassified, which will be discussed later.

Robert Lugar switched multiple times from BZÖ to STRONACH and left both parties to become independent in between. Christoph Hagen switched more than once too, he went from BZÖ to STRONACH and then to FPÖ and in between also had periods of being an independent MP.

6.2 Misclassified MPs

In this chapter I will discuss all the misclassified MPs where my model, in predicting their speeches, failed to correctly assign the party that they belong to. The aim is to find patterns in the misclassifications and use them for interpretation.

6.2.1 SPÖ

Table 6.1 shows that by aggregating the speech predictions on a personal level 22 SPÖ MPs were misclassified as belonging to a different party. The final column in the table shows the predicted party for each of them based on the highest party affiliation score from the model. 19 were predicted to be ÖVP MPs, two GRÜNE and one NEOS.

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Alexander Zach	0.26	0.07	0.05	0.23	0.04	0.01	0.26	0.01	0.06	NEOS
Heinz Fischer	0.22	0.14	0.15	0.45	0.04	0.00	0.00	0.00	0.00	GRÜNE
Ulrike Sima	0.29	0.04	0.16	0.47	0.00	0.03	0.00	0.00	0.00	GRÜNE
Alois Stöger	0.37	0.53	0.07	0.01	0.00	0.01	0.00	0.00	0.00	ÖVP
Andrea Gessl-Ranftl	0.42	0.44	0.06	0.03	0.00	0.03	0.00	0.01	0.00	ÖVP
Christian Hursky	0.32	0.46	0.14	0.04	0.00	0.02	0.01	0.00	0.00	ÖVP
Claudia Schmied	0.36	0.59	0.02	0.01	0.00	0.02	0.01	0.00	0.00	ÖVP
Gerald Klug	0.39	0.46	0.13	0.01	0.00	0.00	0.00	0.00	0.00	ÖVP
Gerhard Köfer	0.27	0.38	0.17	0.00	0.00	0.15	0.00	0.02	0.00	ÖVP
Harald Troch	0.34	0.35	0.15	0.08	0.00	0.04	0.02	0.00	0.01	ÖVP
Johann Hechtl	0.35	0.55	0.06	0.00	0.00	0.02	0.00	0.00	0.00	ÖVP
Josef Auer	0.28	0.29	0.24	0.04	0.00	0.05	0.05	0.04	0.00	ÖVP
Jürgen Schabhüttl	0.28	0.57	0.1	0.01	0.00	0.05	0.00	0.00	0.00	ÖVP
Maria Berger	0.33	0.5	0.03	0.04	0.00	0.09	0.00	0.00	0.00	ÖVP
Michael Ehmann	0.39	0.42	0.11	0.03	0.00	0.01	0.02	0.01	0.00	ÖVP
Michael Schickhofer	0.37	0.49	0.08	0.03	0.00	0.02	0.01	0.01	0.00	ÖVP
Norbert Darabos	0.27	0.37	0.28	0.04	0.00	0.03	0.01	0.00	0.00	ÖVP
Peter Doskozil	0.14	0.46	0.37	0.01	0.00	0.01	0.00	0.00	0.00	ÖVP
Philip Kucher	0.35	0.43	0.06	0.06	0.00	0.03	0.03	0.02	0.02	ÖVP
Rudolf Hundstorfer	0.38	0.43	0.1	0.05	0.00	0.02	0.01	0.00	0.00	ÖVP
Rudolf Plessl	0.33	0.43	0.16	0.03	0.00	0.02	0.01	0.02	0.01	ÖVP
Sonja Steßl-Mühlbacher	0.34	0.44	0.12	0.05	0.00	0.03	0.01	0.00	0.01	ÖVP

Table 6.1: Party affiliation score of misclassified SPÖ members

SPÖ → NEOS

Alexander Zach was actually a member of LIF. For the 2006 elections LIF cooperated with SPÖ with the aim to prevent a right wing majority between ÖVP and FPÖ. For practical reasons he was part of SPÖ Parliamentary Group, but politically he was completely independent. LIF merged with NEOS in 2014 so his prediction as NEOS member is not too far away. He was sent to the ORF Board of Trustees by NEOS.

SPÖ → GRÜNE

Two SPÖ MPs were misclassified as GRÜNE, Ulrike Sima and Heinz Fischer. Ulrike Sima's misclassification is a reflection of her political background prior to running to become an MP for SPÖ in the 1999 elections. She had been a member of the green student party, as well as working for an environmentalist NGO (Global 2000). She ran for GRÜNE in the election in 1996 and only joined SPÖ for the election in 1999, when she got into Parliament as an MP for SPÖ. Heinz Fischer has only two speeches with more than 30 words in the data set. Most of his speeches were held in the position of the President

of the Parliament or even the President of Austria. Those speeches were labelled one as GRÜNE and one as ÖVP. One speech is more administrative. In that speech he responds to an accusation of his organisational role as a President of the Parliament. Two speeches might be too few to say what party a politician belongs to.

SPÖ → ÖVP

The 19 SPÖ MPs who were misclassified as ÖVP can be separated into two groups. The first group includes those MPs who were appointed Ministers by the SPÖ when in coalition with ÖVP. This is by far the biggest group of MPs that has been misclassified with 19 MPs. The second group consists of MPs that joined after the first FPÖ-ÖVP government, when SPÖ was in coalition with ÖVP again. The Ministers that were in an SPÖ-ÖVP government before the first ÖVP-FPÖ coalition are Alois Stöger, Claudia Schmied, Maria-Margarethe Berger, Rudolf Hundstorfer and Norbert Darabos. Because those Ministers are in a coalition with the ÖVP, they might use similar language to argue for policies and agendas that are set by the SPÖ-ÖVP government. The need to argue for a common governmental policy could be a possible explanation for the misclassification of the speeches in this group of ministers. As an example for those Ministers I will discuss Claudia Schmied in more detail. Claudia Schmied was a Minister under SPÖ-ÖVP coalitions from January 2007 to December 2013. Most of her speeches were labelled ÖVP, but there are many SPÖ labelled speeches too. She often deals with topics around her Ministry of Education. It is remarkable that speeches with roughly the same topic and length were labelled differently. One was labelled strongly as SPÖ and one was labelled strongly as ÖVP. The speech about all-day child care where she emphasises the freedom of choice for parents was classified as ÖVP. A speech talking about the exact same topic with the argument that all-day child care helps to fight social inequalities was labelled SPÖ. This is an interesting example of two speeches speaking from the position of the government, but using different ideological view points to argue for the political agenda of all-day child care.

The Ministers that were in an SPÖ-ÖVP government after the first ÖVP-FPÖ government are Norbert Darabos, Hans Peter Doskozil, Rudolf Hundstorfer, Sonja Steßl and Gerald Klug. The fact that they were Ministers in a government coalition with ÖVP could hint towards SPÖ governing politicians adapting their speech behaviour towards their coalition partner ÖVP.

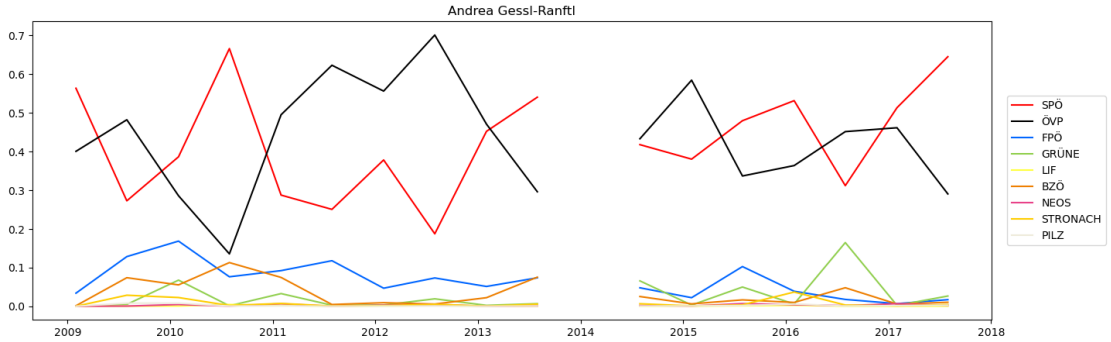

Figure 6.22: Probability distribution of speeches

There are similar dynamics amongst MPs that got into Parliament during a "grand coalition": Andrea Gessl-Ranftl, Christian Hursky, Johann Hechtl, Harald Troch, Josef Auer, Jürgen Schabhüttel, Michael Schickhofer, Michael Ehmann, Philip Kucher and Rudolf Plessel all got into Parliament in a coalition period with the ÖVP.

Andrea Gnessl-Ranftl was in Parliament for two legislative periods, both were "grand coalitions". As seen in Figure 6.22 her speech behaviour moved a bit more towards SPÖ when she was an MP for the second time. In the speeches that were strongly labelled ÖVP, she does comment on things often in a defensive manner. Either she explains what the government is already doing regarding a topic, or she is mentioning and explaining the complexity of certain demands.

Some of the SPÖ MPs misclassified as ÖVP were in Parliament for more than just the periods that SPÖ was in government. Harald Troch was in Parliament from October 2013 until October 2024. He experienced being an MP both in government and in opposition. His aggregated score for ÖVP and SPÖ is between 34-35%. Philip Kucher came into Parliament in October 2013 and is still there. With those MPs the speech behaviour of the coalition times might have influenced their oppositional speech behaviour as well. [4]

Gehard Köfer

He left SPÖ and became independent for the last nine months of his time in Parliament. In that time he joined the newly founded party Team Stronach. He switched to the federal government with his new party, before STRONACH became its own Parliamentary Group, which is why he was independent as an MP even though he already was a party member of STRONACH. That is why his speeches are labelled independent. His speeches after he left SPÖ are either FPÖ or ÖVP, which would fit the misclassification as ÖVP. After some scandals and conflicts with STRONACH, he founded his own regional party in Carinthia.

[4] It would be interesting to gather more data about how influential the first period in Parliament is for MPs and if the first period of an MP in Parliament shapes their speech behaviour.

6.2.2 ÖVP

ÖVP has eleven misclassified MPs as seen in Table 6.2. There are no oppositional ÖVP speeches as they were always in government in the timespan of the data set. Six MPs are misclassified as SPÖ and five are wrongly labelled as FPÖ.

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Franz Steindl	0.38	0.12	0.04	0.03	0.00	0.00	0.00	0.00	0.00	SPÖ
Hannes Farnleitner	0.33	0.03	0.01	0.05	0.00	0.00	0.00	0.00	0.00	SPÖ
Joseph Huainigg	0.3	0.1	0.14	0.00	0.07	0.01	0.03	0.01	0.00	SPÖ
Juliane Bogner-Strauß	0.4	0.06	0.03	0.00	0.02	0.02	0.01	0.01	0.00	SPÖ
Kira Grünberg	0.34	0.12	0.06	0.00	0.01	0.01	0.01	0.02	0.00	SPÖ
Werner Fasslabend	0.34	0.15	0.02	0.02	0.01	0.00	0.00	0.00	0.00	SPÖ
Bernhard Vock	0.15	0.4	0.05	0.00	0.23	0.02	0.04	0.01	2	FPÖ
Ernst Fink	0.32	0.37	0.01	0.01	0.01	0.00	0.00	0.00	2	FPÖ
Gerhart Bruckmann	0.41	0.46	0.01	0.00	0.01	0.00	0.00	0.00	2	FPÖ
Helmut Kukacka	0.32	0.36	0.04	0.01	0.03	0.00	0.00	0.00	2	FPÖ
Peter Fichtenbauer	0.12	0.47	0.12	0	0.07	0.01	0.01	0.00	2	FPÖ

Table 6.2: Party affiliation score of misclassified ÖVP members

ÖVP → SPÖ

Three of the six ÖVP MPs misclassified as SPÖ MPs were in Parliament in the 1990s while ÖVP was in coalition with SPÖ. These MPs are Franz Steindl, Hannes Farnleitner and to an extent Werner Fasslabend. Franz Steindl was a member of the ÖAAB which is the workers association of ÖVP. He was also active in the AK, which is the Workers Chamber. His topics on workers rights might be the reason many of his speeches were classified as SPÖ, even though his second-strongest party is ÖVP. Hannes Farnleitner was Minister for Economy from 1996 to 2000 whilst SPÖ was in government with ÖVP. Werner Fasslabend was Minister for Defence from 1990 to 2000 and Third Parliamentary President from 2000 to 2002. As for the three other ÖVP MPs misclassified as SPÖ, Juliane Bogner-Strauß was Minister for Women from 2017 to 2018; Joseph Huainigg was the speaker for disabled people within the ÖVP Parliamentary Group; and Kira Grünberg is an MP who speaks about inclusive sport/society from her own experience as a professional sports person with disabilities. The last two misclassifications might be due to both speakers emphasising the importance of inclusive sports as it may be that SPÖ talks more about including groups with disadvantages.

ÖVP → FPÖ

As mentioned in the chapter on data quality, two MPs were labelled as ÖVP in the Parliamentary protocols even though they were FPÖ members: Bernhard Vock and Peter Fichtenbauer.

The other three ÖVP MPs that are labelled as FPÖ by the model are Ernst Fink, Helmut Kukacka and Gerhart Bruckmann. In Ernst Fink's speeches, he campaigns against the ÖBB (Austrian Railway) and the railway workers rights and their pension system. Helmut Kukacka criticises the ÖBB in his speeches too. He was head of the Mittelschul-Kartell-Verband (Christian fraternity). As an oppositional party, FPÖ criticises governmental structures with regards to efficiency. Gerhart Bruckmann was the first pensioner MP. He was a long working researcher at IHS before becoming an MP. The FPÖ often push for privatisation and austerity politics.

6.2.3 FPÖ

As Table 6.3 shows 18 FPÖ MPs were misclassified, 15 as ÖVP, two as SPÖ and one as LIF. There are many MPs that actually switched party to the BZÖ, STRONACH or ÖVP.

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Franz Löschnak	0.61	0.14	0.12	0.05	0.07	0.00	0.00	0.00	0.00	SPÖ
Hans Müller	0.39	0.24	0.36	0.00	0.00	0.00	0.00	0.00	0.00	SPÖ
Alois Kainz	0.3	0.34	0.33	0.01	0.00	0.01	0.01	0.01	0.00	ÖVP
Christian Pewny	0.16	0.43	0.38	0.02	0.00	0.01	0.00	0.00	0.00	ÖVP
Christian Ragger	0.26	0.44	0.18	0.00	0.00	0.04	0.06	0.00	0.00	ÖVP
Christian Ries	0.14	0.45	0.32	0.00	0.00	0.06	0.00	0.01	0.01	ÖVP
Christian Schandor	0.24	0.44	0.29	0.01	0.00	0.01	0.00	0.00	0.00	ÖVP
Dieter Böhmdorfer	0.15	0.34	0.25	0.03	0.01	0.22	0.00	0.00	0.00	ÖVP
Elisabeth Sickl	0.1	0.46	0.42	0.01	0.00	0.01	0.00	0.00	0.00	ÖVP
Elmar Lichtenegger	0.2	0.37	0.34	0.03	0.00	0.04	0.00	0.02	0.00	ÖVP
Gerhard Hetzl	0.18	0.41	0.39	0.01	0.00	0.01	0.00	0.00	0.00	ÖVP
Heinz Grasser	0.16	0.59	0.22	0.00	0.00	0.02	0.00	0.00	0.00	ÖVP
Herbert L. Graf	0.23	0.43	0.3	0.03	0.00	0.01	0.00	0.00	0.00	ÖVP
Marialuise Mittermüller	0.13	0.46	0.28	0.00	0.00	0.12	0.00	0.00	0.00	ÖVP
Monika Forstinger	0.32	0.38	0.25	0.00	0.00	0.05	0.00	0.00	0.00	ÖVP
Petra Wagner	0.18	0.37	0.3	0.01	0.00	0.11	0.00	0.02	0.00	ÖVP
Roland Zellot	0.12	0.43	0.41	0.02	0.00	0.02	0.00	0.00	0.00	ÖVP
Willi Brauneder	0.19	0.13	0.25	0.16	0.26	0.00	0.00	0.00	0.00	LIF

Table 6.3: Party affiliation score of misclassified FPÖ members

FPÖ → SPÖ

Franz Löschnak was actually an SPÖ MP and was wrongly labelled in the data. Hans Müller was an MP for FPÖ from 1999 to 2002. He speaks about economical and fiscal topics and often criticises the government.

FPÖ → ÖVP

This is the biggest group of misclassified FPÖ MPs with 15 labelled as ÖVP. Some of them were Ministers or MPs during a coalition between FPÖ and ÖVP [5].

Beate Hartinger-Klein occurs in my data set twice; first as Beate Hartinger and later as Beate Hartinger-Klein. She was labelled correctly as FPÖ in her first period in Parliament from 1999 to 2002. In her second period she was wrongly labelled as ÖVP. This is an example of a Minister being misclassified as the party they are in government with, as ÖVP and FPÖ were coalition partners at that time. Dieter Böhmdorfer was Minister of

[5] They were Ministers in either one or more of the three legislative periods 1999 to 2002, 2002 to 2005 or 2017 to 2019.

Justice from 2000 to 2004 which was also during a time of coalition with ÖVP. Elisabeth Sickl was Minister of Social Security and Generations and later Minister of Work, Health and Social Agenda, again when ÖVP was FPÖ's coalition partner. Monika Forstinger was Minister for Transport for FPÖ from 2000 to 2002 in that "black-blue" coalition. Roland Zellot was an MP for FPÖ from 1999 to 2005, covering a large part of FPÖs time in government. His predictions for FPÖ are very strong but he is still wrongly labelled as ÖVP. Petra Wanger was in Parliament from 2017 to 2019, again during a time when FPÖ was in government with ÖVP. Alois Kainz was an MP from 2017 to 2019 and also is connected to the military.

There is a group of FPÖ MPs whose misclassification as ÖVP may relate to their speeches on the economy. This group includes Christian Pewny, Gerhard Hetzl and Herbert L. Graf. Christian Pewny who was an MP from 2017 to 2019 and a Member of the WKO [6], has only eight speeches, four labelled FPÖ and only three ÖVP. His ÖVP speeches have a very high score for ÖVP. One of the speeches is about economical espionage and how to protect Austrian economy and innovation. Gerhard Hetzl an MP for FPÖ from 2000 to 2002 held 18 speeches, some of which were strongly labelled as ÖVP. Herbert L. Graf MP for FPÖ 1999 to 2002 was also a board member of the WKO. With these three MPs it could be that their agenda to speak about the economy and business influenced their speech behaviour, particularly because the WKO is ideologically close to ÖVP. Christian Ragger an MP from 2017 to 2018 has only eleven speeches, again with two speeches that were strongly labelled as ÖVP. Christian Schandor an MP from 2017 to 2019 has only eleven speeches, and some were strongly labelled as ÖVP when he was talking about the economy.

A third group of FPÖ MPs misclassified as ÖVP speakers are the ones that switched to the BZÖ. Elmar Lichtenegger was an MP for FPÖ from 2003 to 2006 and later switched to the BZÖ in April 2006. Marialuise Mittermüller was an MP for FPÖ from 2005 to 2006. She has 29 speeches and also joined BZÖ in April 2006. Heinz Grasser left FPÖ in 2003, but as an independent MP remained Minister of Finance nominated by the ÖVP, up until 2007. He distanced himself from both FPÖ and the BZÖ, but the parties also distanced themselves from him.

FPÖ → LIF

Willi Brauneder, who was wrongly classified as LIF, was Third President of Parliament. Ten of his 26 speeches were labelled as LIF, a few others as SPÖ and ÖVP and only eight as FPÖ. He was Dean of the Law Faculty at the University of Vienna.

[6] The WKO, Wirtschaftskammer Österreich, is an institutionalised representation of industry, business and employers. Austria has a so called "chamber system", which has a Workers chamber "Arbeiterkammer (AK)" as well as chambers for doctors, farmers and other parts of society.

6.2.4 GRÜNE

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Berivan Aslan	0.32	0.12	0.09	0.23	0.00	0.04	0.12	0.03	0.05	SPÖ
Bettina Hradecsni	0.33	0.15	0.18	0.26	0.00	0.04	0.02	0.02	0.01	SPÖ
Heidemarie Rest-Hinterseer	0.37	0.08	0.13	0.36	0.00	0.03	0.03	0.01	0.00	SPÖ
Johannes Voggenhuber	0.49	0.09	0.03	0.22	0.17	0.00	0.00	0.00	0.00	SPÖ
Kurt Grünewald	0.28	0.14	0.2	0.25	0.01	0.06	0.02	0.03	0.01	SPÖ
Wolfgang Pirklhuber	0.4	0.08	0.1	0.33	0.01	0.05	0.01	0.01	0.01	SPÖ
Barbara Neuroth	0.24	0.35	0.2	0.02	0.00	0.03	0.00	0.15	0.00	ÖVP

Table 6.4: Party affiliation score of misclassified GRÜNE members

GRÜNE → SPÖ

Table 6.4 shows that six of the seven misclassified GRÜNE MPs were labelled as SPÖ. The six MPs are Berivan Aslan, Bettina Hradecsni, Heidemarie Rest-Hinterseer, Johannes Voggenhuber (who ran for PILZ in the EU-elections 2019), Kurt Grünewald and Wolfgang Priklhuber (who talks about agriculture and food).

GRÜNE → ÖVP

Barbara Neuroth, the only GRÜNE MP misclassified as ÖVP, was in Parliament for only five months with only two speeches. That might be too little to actually predict a party affiliation.

6.2.5 LIF

Table 6.5 shows that LIF had seven misclassified MPs, three of whom were labelled as SPÖ, two as FPÖ and two as GRÜNE.

LIF → SPÖ

Of the three who were misclassified as SPÖ, Friedhelm Frischenschlager was an MP for FPÖ from 1977 to 1993 and Minister for Defence during the SPÖ minority government. He became an MP for LIF later from 1993 to 1996.

Hans Moser was also an MP for FPÖ from 1989 to 1993 and then switched to LIF from 1993 to 1999. He is in the Austrian Military.

Rosa Mlinar was an MP for NEOS/LIF from 2013 to 2014 and a Member of the EU Parliament for NEOS from 2014 to 2019. Her speeches are either labelled NEOS or SPÖ.

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Friedhelm Frischenschlager	0.26	0.11	0.24	0.16	0.23	0	0	0	0	SPÖ
Helmut Moser	0.36	0.16	0.19	0.07	0.2	0.01	0	0	0	SPÖ
Helmut Peter	0.19	0.17	0.29	0.12	0.22	0	0	0	0	FPÖ
Maria Schaffenrath	0.19	0.09	0.24	0.25	0.22	0.01	0	0	0	GRÜNE
Peter Haselsteiner	0.15	0.11	0.31	0.14	0.3	0	0	0	0	FPÖ
Rosa Mlinar	0.29	0.26	0.05	0.03	0.01	0.02	0.28	0.03	0.04	SPÖ
Thomas Barmüller	0.19	0.07	0.19	0.29	0.26	0	0	0	0	GRÜNE

Table 6.5: Party affiliation score of misclassified LIF members

LIF → FPÖ

Of the two MPs that were labelled as FPÖ, Helmut Peter was actually an MP for FPÖ from 1990 to 1993 before joining LIF in 1994 where he stayed until 1999. Peter Haselsteiner was an MP for LIF from 1994 to 1998. He was a businessman who got involved in politics.

LIF → GRÜNE

Thomas Barmüller was an MP for FPÖ from 1990 to 1993 and for LIF from 1993 to 1999. Maria Schaffenrath was an MP for LIF from 1994 to 1999 and later an MP for NEOS.

6.2.6 BZÖ

Of the eight BZÖ MPs shown in Table 6.6 seven were labelled as FPÖ and one as ÖVP

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Ernest Windholz	0.14	0.14	0.4	0.04	0	0.22	0.01	0.04	0	FPÖ
Gerhard Huber	0.11	0.07	0.38	0.03	0	0.34	0	0.06	0	FPÖ
Hubert Gorbach	0.12	0.42	0.39	0.02	0	0.04	0	0	0	ÖVP
Karin Gastinger	0.18	0.32	0.46	0.01	0	0.02	0	0	0	FPÖ
Peter Westenthaler	0.08	0.12	0.4	0.09	0	0.29	0.01	0.01	0	FPÖ
Sigisbert Dolinschek	0.16	0.12	0.49	0.03	0	0.18	0.01	0.01	0	FPÖ
Veit Schalle	0.2	0.16	0.38	0.04	0	0.18	0.01	0.04	0	FPÖ
Wolfgang Spadiut	0.15	0.1	0.49	0.04	0	0.17	0.01	0.03	0	FPÖ

Table 6.6: Party affiliation score of misclassified BZÖ members

BZÖ → ÖVP

The one BZÖ MP who was misclassified as ÖVP was Hubert Gorbach who was Minister for Transport, Technology and Innovation as well as Vice Chancellor from 2003 to 2007. His time in Parliament overlaps mostly with the time of the BZÖ/FPÖ coalition with ÖVP from 2002 to 2006.

BZÖ → FPÖ

Five of the seven BZÖ MPs wrongly labelled as FPÖ had actually switched parties from FPÖ to BZÖ. The MPs with an actual FPÖ background are: Ernest Windholz, Peter Westenthaler, Sigisbert Dolinschek, Wolfgang Spadiut and Christoph Hagen.

Christoph Hagen switched party many times. He was an MP for FPÖ until 2004 when he joined BZÖ. From 2008 to 2012 he became an MP for BZÖ again and left the party to be independent for a short time. After less than a month of being independent he became an MP for STRONACH from 2012 to 2017. Finally, he left STRONACH and became independent again for just three months before he left Parliament.

Gerhard Huber was MP for BZÖ from 2008 to 2009, became an independent MP from 2009 to 2010 and then switched back to BZÖ from 2010 to 2013. He was asked to leave the Parliamentary Group for some of that year because of a case against him.

Karin Gastinger was a Minister of Justice from 2004 to 2007. She was never a member of a party, but was nominated by the BZÖ. Veit Schalle was an MP from 2006 to 2008 when he was already retired. Before he worked for the REWE group and was active against "Kärntner Slowenen". He has connections to right wing activism in Carinthia. There are quotes about the Third Reich's market politics in media coverage.

6.2.7 NEOS

The five NEOS MPs who were misclassified, three as FPÖ, one as SPÖ and one as GRÜNE, are shown in Table 6.7

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Christoph Vavrik	0.16	0.16	0.21	0.17	0.01	0.05	0.13	0.06	0.04	FPÖ
Gerald Loacker	0.18	0.11	0.27	0.12	0.01	0.09	0.16	0.05	0.02	FPÖ
Irmgard Griss	0.28	0.14	0.13	0.24	0.02	0.05	0.09	0.01	0.06	SPÖ
Josef Schellhorn	0.1	0.15	0.32	0.08	0.01	0.05	0.17	0.1	0.02	FPÖ
Nikolaus Scherak	0.15	0.06	0.24	0.24	0	0.07	0.19	0.01	0.04	GRÜNE

Table 6.7: Party affiliation score of misclassified NEOS members

NEOS → SPÖ

Irmgard Griss was an MP for NEOS from 2017 to 2019 even though she was actually never a party member of NEOS. She is a judge and her speech predictions are somewhere between SPÖ/ÖVP and NEOS.

NEOS → FPÖ

Christoph Vavrik was an MP for NEOS from 2013 to 2017 but joined the ÖVP Parliamentary Group for six months in 2017. He was banned from the Parliamentary Group of NEOS because he was accused of making a homophobic post on Facebook, and ÖVP took him in. Gerald Loacker was an MP for NEOS from 2013 to 2024 and active in the fraternity KaV Norica Wien. Before his time in Parliament as a NEOS MP, he was an ÖVP member. Josef Schellhorn was a member of ÖVP until 2013. He then switched to become an MP for NEOS from 2014 to 2021 and also from 2024 to 2025. From 2024 onwards he is Minister for Outer Agendas and European Policies.

NEOS → GRÜNE

Nikolaus Scherak was an MP for NEOS from 2013 to 2014.

6.2.8 STRONACH

As detailed in Table 6.8, nine STRONACH MPs were misclassified, six as FPÖ, two as BZÖ and one as ÖVP.

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Elisabeth Kaufmann-Bruckberger	0.17	0.16	0.43	0.01	0	0.18	0	0.05	0	FPÖ
Erich Tadler	0.2	0.12	0.23	0.02	0	0.26	0	0.16	0	BZÖ
Georg Vetter	0.14	0.26	0.31	0.08	0.01	0.06	0.05	0.1	0.01	FPÖ
Leopold Steinbichler	0.2	0.2	0.33	0.05	0	0.07	0.04	0.1	0.01	FPÖ
Martina Schenk	0.11	0.06	0.24	0.05	0	0.33	0.02	0.17	0.01	BZÖ
Robert Lugar	0.05	0.04	0.3	0.05	0	0.29	0.03	0.23	0.01	FPÖ
Rouven Ertlschweiger	0.21	0.32	0.23	0.04	0.01	0.02	0.05	0.11	0.01	ÖVP
Ulrike Weigerstorfer	0.18	0.09	0.35	0.13	0	0.03	0.04	0.16	0.01	FPÖ
Waltraud Dietrich	0.11	0.19	0.41	0.02	0	0.11	0.02	0.15	0	FPÖ

Table 6.8: Party affiliation score of misclassified STRONACH members

STRONACH → ÖVP

Whilst Rouven Ertlschweiger was an MP for STRONACH from 2014 to 2015, he then switched from STRONACH and became an ÖVP MP from 2015 to 2017.

STRONACH → FPÖ

Elisabeth Kaufmann-Bruckberger was an MP for BZÖ from 2011 to 2012 and was briefly independent before she joined the STRONACH Parliamentary Group from 2012 to 2013. She was an FPÖ member before she joined BZÖ and was politically active on a local political level before and after her time in Parliament.

Leopold Steinbichler was an MP for ÖVP from 1997 to 2003 and reentered Parliament as an MP for STRONACH from 2013 to 2017.

Two of the six misclassified MPs switched parties from STRONACH to ÖVP, but nonetheless were labelled as FPÖ members according to their speeches: Georg Vetter and Marcus Franz were both MPs for STRONACH from 2013 to 2015 and switched to being MPs for ÖVP in 2015 after they were invited by Reinold Lopatka to join ÖVP Parliamentary Group. Franz had to leave, due to a critical statement he made regarding Angelika Merkel and her policies in 2016 and he became independent after that. Robert Lugar was an MP who switched parties multiple times. He was an MP for BZÖ from 2008 to 2011, an independent MP after that and an MP for STRONACH from 2012 to 2017. He then joined the FPÖ Parliamentary Group from 2017 to 2019. So to a degree the FPÖ prediction is not completely wrong. Ulrike Weigerstrofer was an MP for STRONACH from 2013 to 2017. She worked for Stronach's Magna Group.

Waltraud Dietrich's prediction is also partially correct. She was an MP for STRONACH from 2013 to 2017, but before 2012 was an FPÖ member.

STRONACH → BZÖ

Erich Tadler was rightfully predicted as BZÖ. From 2008 to 2010 he was actually a BZÖ MP. However, following a period as an independent MP from 2010 to 2012, he then joined STRONACH as an MP from 2012 to 2013. Likewise, Martina Schenk was an MP for both BZÖ from 2008 to 2013 and then STRONACH from 2013 to 2017.

6.2.9 PILZ

Table 6.9: There were only six MPs in the entire data set period. None of the six were correctly classified as being a PILZ MP. One explanation could be that the roughly 90 speeches were not enough labelled data for training a text classifier for PILZ. Five of the PILZ MPs were misclassified as SPÖ. The other, Alfred Noll, was misclassified as GRÜNE. One of the five, Alma Zadic, was an MP for PILZ from 2017 to 2019, an independent MP from July to October 2019 and then ran for GRÜNE in the following elections and served as a GRÜNE MP from 2019 to 2020. Later, from 2020 to 2025, she was Minister of Justice for GRÜNE.

Speaker	SPÖ	ÖVP	FPÖ	GRÜNE	LIF	BZÖ	NEOS	TS	PILZ	predicted
Alfred J. Noll	0.17	0.1	0.24	0.28	0.01	0.03	0.13	0.02	0.04	GRÜNE
Alma Zadic	0.23	0.14	0.11	0.15	0	0.03	0.17	0.05	0.12	SPÖ
Daniela Holzinger-Vogtenhuber	0.41	0.1	0.08	0.16	0	0.02	0.09	0.08	0.04	SPÖ
Martha Bißmann	0.28	0.14	0.09	0.13	0	0.03	0.13	0.02	0.17	SPÖ
Peter Kolba	0.37	0.17	0.12	0.09	0	0.03	0.07	0.07	0.1	SPÖ
Stephanie Cox	0.31	0.11	0.09	0.2	0	0.01	0.18	0.04	0.05	SPÖ

Table 6.9: Party affiliation score of misclassified PILZ members

Daniela Holzinger-Vogtenhuber was an MP for SPÖ from 2013 to 2017 before she became an independent MP from July 2017 to November 2017 and then switched to PILZ where she was an MP from 2017 to 2019.

Martha Bißmann was an MP for PILZ from 2017 to 2018. She was active in the green student organisation but later went campaigning for Irmgard Griess at the presidential elections. Irmgard Griess went to NEOS.

Peter Kolba was an MP for PILZ from 2017 to 2018. When he left Parliament, Peter Pilz took over his seat. Stephanie Cox was an MP for PILZ from 2017 to 2019. Her speeches are quite spread between SPÖ with 31%, 20% for GRÜNE and 18% for NEOS.

6.3 Learnings from model scores

Three big hypotheses in my work: 1. MPs adapt their speech behaviour to their party. If that is true, then the speech behaviour would change after they switched party. Otherwise, it would stay the same or similar. Mostly adaptation happens later, sometimes before. We discussed that in chapter 6.1.

2. Parties adapt their speech behaviour to their coalition partners, if two parties are in coalition they have more similar scores.

3. Parties adapt their speech behaviour to their oppositional position, if a party is in opposition it moves further from the governing parties.

If that is true, parties will have most misclassified speeches towards their coalition partners.

As seen in the graphs in Figures 6.23 to 6.25 for SPÖ, ÖVP and FPÖ, this is the case.

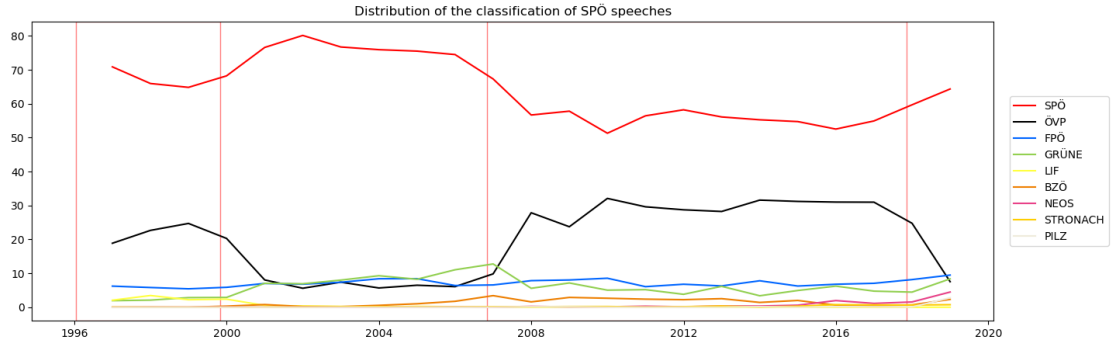

Figure 6.23: Proportion of SPÖ speeches being classified as ÖVP or FPÖ

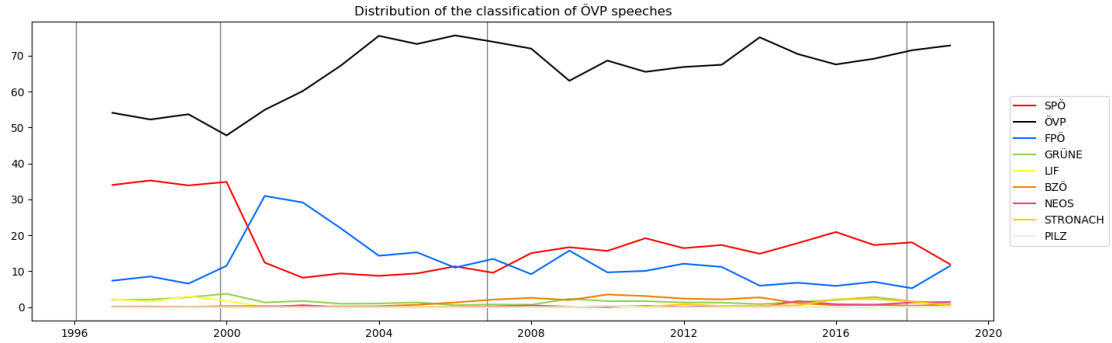

Figure 6.24: Proportion of ÖVP speeches being classified as ÖVP or FPÖ

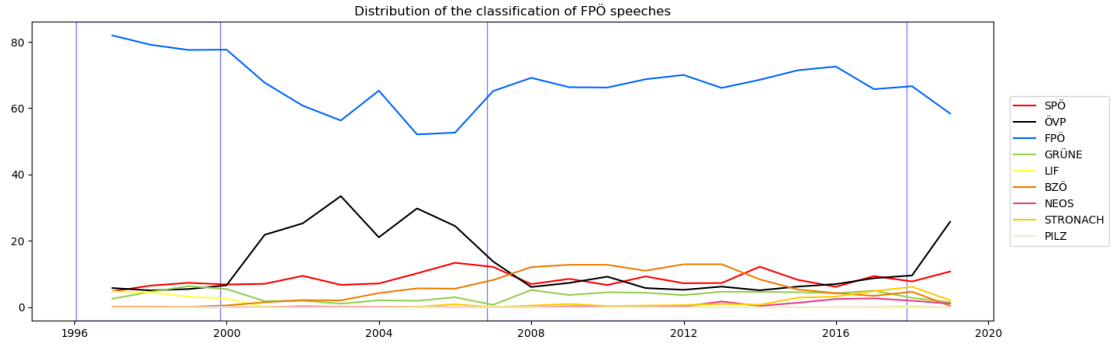

Figure 6.25: Proportion of FPÖ speeches being classified as ÖVP or FPÖ

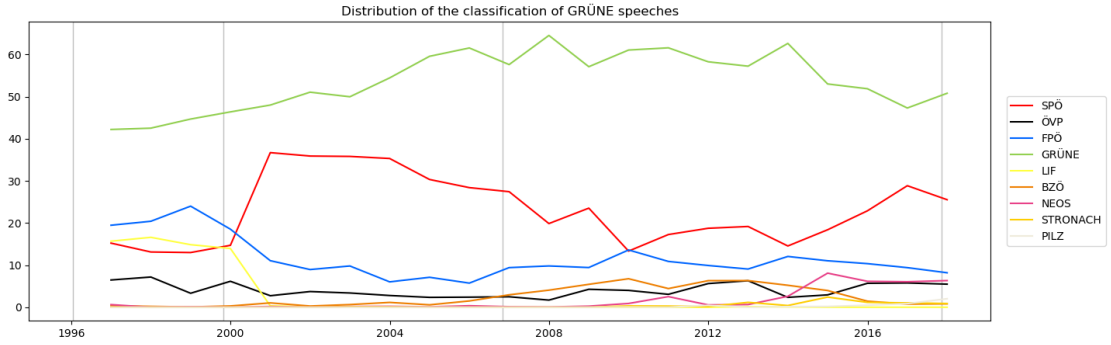

Figure 6.26: Proportion of GRÜNE speeches being classified as ÖVP or FPÖ

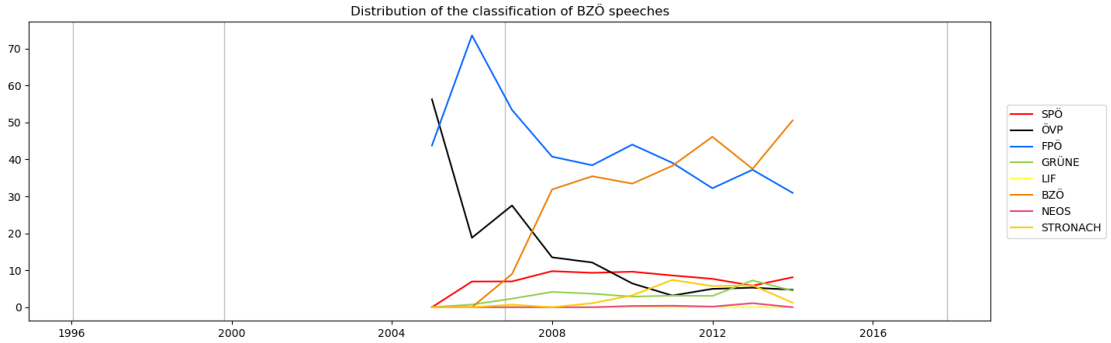

Figure 6.27: Proportion of BZÖ speeches being classified as ÖVP or FPÖ

Deducting from the previous hypothesis, the parties should distance themselves again to the extent they got closer during governing times. Because ÖVP was never in an oppositional position within the covered data it is difficult to say anything for sure. It is still observable, that the misclassified speeches towards SPÖ drastically reduced after being in a coalition with FPÖ.

7 Conclusion

To conclude my thesis, I will first talk about the party affiliation and speech behaviour. This will be followed by insights about the political developments in Austria based on the data set I created. Finally, I will talk about the limitations of my findings and ideas for further work.

Speech activity

The evaluation of my models shows that there are differences in the MPs' speech behaviour according to their party affiliation. Specifically, I found that the number of speeches varies based on party affiliation. The following observations were made:

- Members of Parliamentary Groups generally speak more than independent MPs.
- MPs in opposition speak more than MPs in government.

A possible explanation for why independent MPs speak less than affiliated MPs can be found in the organisation of Parliamentary Work. Independent MPs may have fewer opportunities to make speeches in Parliament due to the division of work within Parliamentary Groups or the comparatively shorter time they spend in Parliament. From a democratic point of view, this observation brings up an interesting discussion. Technically MPs are voted into Parliament as lists of individuals created by parties and not as a party. Therefore, they do not lose their seat when switching parties or becoming independent MPs. Nonetheless in the perception of voters, they vote for a party that creates those lists and sends people to Parliament as MPs. Explanations for why MPs in opposition speak more than MPs in government could be due to motivations and reasons behind giving a speech in Parliament. Perhaps MPs in government are more bound to only present and defend positions, and MPs in opposition are more there to question the government's actions. It could also be purely because the smaller parties are in opposition more often and have fewer MPs and so each MP has to cover more topics and Parliamentary Working Groups. Overall, whether the differences in the number of speeches based on party affiliation are good or bad is an open question for political science and society to assess.

Party affiliation

Before training my models and evaluating them, there was the idea that maybe we could predict which MPs would switch parties. In my models, this was not observed. Most MPs who switched parties adapted their speech behaviour to their new parties. The time of adaptation is not necessarily predictable or systematically observable. Some

thoughts on why the models are not able to predict which MPs would switch parties are the following: Most MPs switched to a newly created party rather than to an established party. Maybe it takes some time for newly founded parties to create a new identity or at least a common communication strategy or shared speech habits. That means that maybe the speech style has not yet evolved, and they could not have been labelled that way. It would be interesting to look at the speech behaviour of Parliamentary Groups in a sociolinguistic way and observe how the intragroup communication of a newly founded party can influence the individual speech patterns. The analysis does not clearly say that the MPs follow the "Klubzwang" in their speeches. Some parties have less variety within themselves compared to other parties. Similar speech behaviour could be a sociolinguistic phenomena. All I can say for certain is that there is a common way of speaking for parties. As we saw in the t-SNE plot at the end of chapter 5, there are clear patterns of speech that group pretty much around party boundaries. Another important observation is that the misclassifications of the model are not "too far away" politically speaking. The left party GRÜNE is more likely to misclassify to SPÖ, the centre-left party. SPÖ will most likely misclassify to GRÜNE or the centre-right party ÖVP. ÖVP will most likely misclassify to the centre-left party SPÖ or the right party FPÖ. The right party FPÖ will either misclassify to another right party, BZÖ or the centre right party, ÖVP. Only LIF and NEOS do not follow this pattern. They misclassify to both parties on the left and right, GRÜNE, SPÖ, FPÖ, ÖVP, because they are socially liberal but economically right. My models struggle to classify them clearly. We cannot predict an MP changing parties just by their speeches. It is rather the other way around. With some exceptions, MPs changed their speech behaviour after they switched. Nonetheless sometimes the misclassifications give important information about the MP's background.

Government versus opposition

Looking at the predictions of all speeches over time revealed that parties behave differently in opposition and government. When governing, they get closer to their coalition partner, whilst in opposition, they move away from the governing parties. This does not apply equally to all governing and oppositional parties at all times, but has a temporal aspect that affects some parties more than others: Most misclassifications of SPÖ to ÖVP happened after the first coalition of FPÖ and ÖVP. That means that there was a difference in speech behaviour the second time SPÖ was in government with ÖVP compared to the first time they governed together in this data set. Most misclassifications from ÖVP to SPÖ happened before the first time FPÖ was in government with ÖVP. After the first grand coalition [1] in my data set, ÖVP moved away from SPÖ and never fully returned to the proximity it had when it was in coalition the first time. Also, the misclassifications of ÖVP to FPÖ happened within or after the first time working with FPÖ.

Overall, MPs change their speeches according to their party affiliation. Further, parties adapt their speeches to their governing position. There is one more observation that

[1] Grand coalition means the coalition between SPÖ and ÖVP, the traditionally biggest parties in Parliament. Today, this might not be a relevant or correct term, as FPÖ has grown to become just as large as the other two parties.

may be of interest for further research in the field of political science. ÖVP has never been in opposition in my data set and according to my models predictions, has moved more to the right. In fact, looking at my models predicted probability distributions of each speech between 1996 and 2018, the Parliament as a whole has moved more to the right as there were no new left parties to take up the space in the political spectrum that ÖVP moved away from. That means that there are fewer SPÖ labelled speeches overall in Parliament by time and no new parties that would give speeches that are classified towards the left, except for PILZ which replaced GRÜNE for one period. There are fewer speeches classified as left-leaning over time in Parliament as a whole if ÖVP moves away from SPÖ, SPÖ slowly moves further to FPÖ and ÖVP and three new parties [2] lean more towards ÖVP and FPÖ in their speech predictions.

Limitations

I will finish with some remarks on what this model is capable of and what it cannot do. As discussed in both chapters 5 and 6, it is fairly good at predicting a speaker's party affiliation. It is however, limited when looking at the individual speech level. Exploring why this is the case would be an interesting project. One idea is that it might not be enough input to actually map someone to a party by just one speech. Usually, a speech can have a certain aim and is limited to a smaller range of topics. Some parties might share positions, or some positions might change over time. The bigger granularity of the aggregated MP level might give more leeway to topical differences. This could help understand why the aggregation of speeches works best.

Another limitation is that we cannot know why exactly the model is behaving that way. I tried, outside the scope of this thesis, to look into feature based explanations models. Naturally, studies in social science will carry out discourse analyses and utilise qualitative methods with speeches. It would have been great to give those research fields more insight into special words or phrases that helped the classifier to identify the speech. Unfortunately, both SHAP [LL17] and DIG [SR21] did not really give any meaningful output for social science. I cannot say if the model works that well because certain parties use more numbers or ask more questions or just have more adverbial variety than others. It is a black box, as with most NLP methods that work with neural networks and backpropagation. There are very strong hints that my models do find political proximity. It certainly provides useful insights by looking at it in a temporal or aggregated way. However, it is not something that can be used the same way as a data set that was tagged and curated by political scientists. This thesis has its strength in the quantities that it covers, rather than the explanations on the feature level it offers. The data set does not give an explanation of which words or phrases caused the party predictions. Nevertheless, by being able to aggregate vast amounts of speeches, we can discover the patterns discussed in this thesis on an empirical level. Some further research in the field of political science could look into the proximity of parties. One could connect the speech behaviour to other aspects, for example the actual work and policies introduced. Further,

[2] BZÖ, STRONACH and NEOS

one could even connect it to surveys of parties at that time to have more data apart from the electoral data. This could be useful to both parties and journalists in writing about certain phenomena and give more insight to the public about democratic processes in Parliament.

Bibliography

[CSM19] Branden Chan, Stefan Schweter, and Timo Möller. German bert (cased). `https://huggingface.co/google-bert/bert-base-german-cased`, 2019.

[CSM20] Branden Chan, Stefan Schweter, and Timo Möller. German's next language model. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6788–6796, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.

[DCLT19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[DEZ26] Edward Strasser/Meine Abgeordneten Verein zur Förderung von Transparenz und Demokratie Dr. Elke Zuckermann. Meine abgeordneten. so tickt Österreichs politik. `https://www.meineabgeordneten.at/`, 2026. Accessed: 2026-02-16.

[Gal91] Michael Gallagher. Proportionality, disproportionality and electoral systems. *Electoral studies*, 10(1):33–51, 1991.

[GMGBGC23] Eduardo C. Garrido-Merchan, Roberto Gozalo-Brizuela, and Santiago Gonzalez-Carvajal. Comparing bert against traditional machine learning models in text classification. *Journal of Computational and Cognitive Engineering*, 2(4):352–356, April 2023.

[Har54] Zellig Harris. Distributional structure. *Word*, 10(2-3):146–162, 1954.

[HC26] Helf Christopher/Verein zur Förderung von digitaler Politik und Künstlicher Intelligenz Helf Christina. parliament.fyi. `https://www.parlamentarian.fyi/`, 2026. Accessed: 2026-02-16.

[Iak25] Gennadii Iakovlev. Elite polarization in european parliamentary speeches: a novel measurement approach using large language models, 2025.

[JM26] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models.* 3rd edition, 2026. Online manuscript released January 6, 2026.

[KS23] Kontilenia Maria Kotsifakou and Dionisios N. Sotiropoulos. Greek political speech classification using bert. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–7, 2023.

[LL17] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017.

[MSC+13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality, 2013.

[MYZ13] Tomas "Mikolov, Wen-tau Yih, and Geoffrey" Zweig. "linguistic regularities in continuous space word representations". In Lucy "Vanderwende, Hal Daumé III, and Katrin" Kirchhoff, editors, *"Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies"*, pages "746–751", "Atlanta, Georgia", jun "2013". "Association for Computational Linguistics".

[nRoA23] 2nd Republic of Austria. Geschäftsordnung des nationalrates (gog-nr). BGBl. Nr. 410/1975 idF BGBl. I Nr. 129/2023, 1975–2023. Parlament Österreich.

[NZY+25] Stephanie Ng, James Zhang, Samson Yu, Asim Bhatti, Kathryn Backholer, and C. P. Lim. Stance classification: a comparative study and use case on australian parliamentary debates. *Journal of Computational Social Science*, 8(2):1–37, May 2025.

[PBS+25] Lukas Pätz, Moritz Beyer, Jannik Späth, Lasse Bohlen, Patrick Zschech, Mathias Kraus, and Julian Rosenberger. Analyzing german parliamentary speeches: A machine learning approach for topic and sentiment classification, 2025.

[RRP24] Ines Reinig, Ines Rehbein, and Simone Paolo Ponzetto. How to do politics with words: Investigating speech acts in parliamentary debates. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8287–8300, Torino, Italia, May 2024. ELRA and ICCL.

[RS20] Christian Rauh and Jan Schwalbach. The parlspeech v2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies, 2020.

[SFT+24] Raphael Scheible, Johann Frei, Fabian Thomczyk, Henry He, Patric Tippmann, Jochen Knaus, Victor Jaravine, Frank Kramer, and Martin Boeker. GottBERT: a pure German language model. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 21237–21250, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[SR21] Soumya Sanyal and Xiang Ren. Discretized integrated gradients for explaining language models, 2021.

[vdTP26] & Kritzinger S. /Institut für Staatswissenschaft. Universität Wien. von der Thannen P., Partheymüller J. Autnes-dashboard. `https://autnesdashboard.univie.ac.at:3838/autnes/`, 2026. Accessed: 2026-02-16.

[VS25] Sowmya Vajjala and Shwetali Shimangaud. Text classification in the llm era – where do we stand?, 2025.

[VSP+23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[WSY25] Franziska Wagner, Dean Schafer, and Mehmet Yavuz. Opposition to government and back: How illiberal parties shape immigration discourse and party competition. *Politics and Governance*, 13(0), 2025.

[ZHL+25] Junyan Zhang, Yiming Huang, Shuliang Liu, Yubo Gao, and Xuming Hu. Do BERT-like bidirectional models still perform better on text classification in the era of LLMs? In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 18980–18989, Suzhou, China, November 2025. Association for Computational Linguistics.

8 Appendix

For my code, experiments and all results visit `https://github.com/Sarah1108/aut_parl_speeches`.