Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Department of Political Science

Text Complexity and Sentiment Analysis: The Case of the Swiss Voting Booklet

Stefan Keel, 23366536

Lecturer: Dr. Martyn Egan

March 7, 2024

[0]Replication files are publicly accessible at github.com/keeleek42/qtaproject.

School of Social Sciences and Philosophy: Assignment Submission Form

Student Name:	Stefan Keel
Student ID Number:	23366536
Programme Title:	MSc. in Applied Social Data Science
Module Title:	Quantitative Text Analysis for Social Scientists
Assessment Title:	Research Project
Lecturer (s):	Dr. Martyn Egan
Date Submitted	March 7, 2024

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://www.tcd.ie/calendar

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write

Signed: Stefan Keel Date: March 7, 2024

Contents

1 Introduction **1**

1.1 Contribution 1

2 Literature **2**

3 Swiss Voting Booklet: Data **3**

4 Methods **4**

4.1 Tokenisation, Collocation and MTTR 4

4.2 Mean Type-Token Ratio 4

4.3 Frequency and Keyness 5

4.4 Complexity Analysis 5

4.5 Sentiment Analysis 5

5 Findings **6**

5.1 Frequency 6

5.1.1 Keyness 7

5.2 MTTR 7

5.3 Complexity 8

5.3.1 SMOG.de Index and Slopes 8

5.4 Sentiment 10

6 Discussion **12**

List of Figures

5.1 Keyness Govt. vs. Initiative and Referenda 7
5.2 MTTR 2003 until 2024 8
5.3 SMOG.de Index over years 9
5.4 Net Sentiment 10

1 Introduction

Swiss citizens can vote regularly on different topics that affect the political landscape. Initiatives are put forward by citizens wanting to amend the constitution whereas referenda come about as the result of a challenge to a law. From voting on initiatives, which are changes to Swiss law to come to vote by popular support, whether employees shall have six weeks of holidays, to whether Switzerland shall become an EU member state, the range of topics cover a wide range of aspects. Voting can be initiated by the government, initiative.

Each voting is accompanied by the official Swiss Voting Booklet that every Swiss citizen above the age of 18, and who is eligible to vote, receives via postal mail including the voting ballots. The Swiss Voting Booklet has explanatory text and advice from the government, on why a voting shall be accepted or declined. Furthermore, it provides the statement of the initiative or referendum committee with their arguments.

Different initiatives or referenda that are launched on the spectrum from right to left on the political landscapes seem to become more pithy. It is of a question whether this seeming shift in the tone is to address and engage more voters in favour of each vote. This study aims to find whether there are significant effects that political texts in voting booklets are becoming less complex and a general shift to more negative sentiments can be observed. This prior subjective observation by the author led to the two hypotheses.

1.1 Contribution

This paper seeks to contribute to the current by answering the question of whether the political argumentation in the Swiss voting booklet is becoming less complex and whether it can be observed that the wording is becoming more negative. This will be done by analysing the arguments of the Swiss Government, Initiative and Referendum committees in the Swiss voting booklets in the timespan from 2004 until 2024.

The main goal of this paper is to accept or reject the hypotheses that political texts are becoming less complex and that there are indicators to more negative sentiments.

2 Literature

Evidence was found that different parties use different levels of complexity in their argumentation and this was also validated with the extreme comparison in the case of Hillary Clinton vs. Donald Trump (Degani, 2016). Leading up to and during the US presidency of Donald Trump, it was observed, that political language has become simpler, populist and easier to comprehend (Degani, 2016). Bischof and Senninger (2018) found, that political texts have become less complex over time to address the broad population in a more accessible way.

The measurements for the complexity of political argumentation are not standardised and prone to the change in language itself. Observing a general trend, there is evidence that sentences are becoming shorter and therefore less complex (Benoit et al., 2019).

Whilst analysing texts, sentiment analysis is of utmost importance. A positive sentiment is not always in line with the stance of the arguer and leaves space for interpretation as well as misunderstanding when analysing in a qualitative approach (Bestvater and Monroe). Nevertheless, Noury and Roland (2020) have shown that since 2008 populist tendencies in politics can be seen in many European countries. Populist rhetoric is associated with (Gerstlé and Nai, 2019) negativity, emotionalising and creating fear.

3 Swiss Voting Booklet: Data

The Swiss Voting Booklets (booklets) are accessible in German, French and Italian language to the public and are digitalised back to the year 1970. For this research, the German language was selected.

As a first step, the available booklets were individually assessed and it was evident that since 1970 a lot of changes to layout, style and size have been made. Therefore, a reliable comparison between Government, Initiative and Referenda since 1970 was not feasible.

A consistent design of the booklets was introduced in the year 2003. Nevertheless, even though visually consistent the underlying format in the PDFs was inconsistent. This resulted in the impossibility of automatically scraping and transforming them. The individual texts were copied by hand and a dataframe was built. A validation of the correctness of the created dataframe was secured by comparing it to a separate CSV file, provided by the Swiss Government, which provided the outcomes for each vote that was merged with the text dataframe. This yielded 288 votes and respective texts.

4 Methods

The methods are based on the Quanteda-Package for R. The applied methods in this analysis included the preparation of the texts with tokenisation, collocation and cleaning to conduct statistical analysis. The sentiment analysis was conducted by using an imported dictionary for political texts in the German language. The complexity of the texts was assessed by applying an index formula (SMOG.de) that is widely used when analysing German language texts.

4.1 Tokenisation, Collocation and MTTR

The texts were cleaned by getting rid of obsolete special signs and spaces. Collocations were built and assessed. It was found that the collocations were in most cases not logical. This is because each vote is on a new topic and only a few collocations are used over different voting topics. These collocations were assigned manually due to subjective knowledge of the political argumentation in Switzerland. With the inbuilt stop-words function from Quanteda (2018), stop-words were removed and expanded further by four important but not inbuilt stop-words. The cleaned tokens and collocation were allocated into a corpus for further analysis.

4.2 Mean Type-Token Ratio

The Mean Type-Token Ratio (MTTR) is a simplified test for assessing the complexity of a text by referring to the average rate at which a new types of tokens are encountered in a sequence of text. A token is a unit of text where a token type refers to a single instance of a token in a text. This method was applied to achieve a first assessment of text complexity and reference for a more elaborate text complexity analysis.

4.3 Frequency and Keyness

The frequency analysis counts the occurrences of words in a text. This indication was used for a deeper understanding of the differences or similarities between the texts of the government, initiative and referendum.

The keyness compares the frequency of a text to a reference. The government texts were chosen as the reference as the government argumentation refers to the initiative and the referendum refers to the government. The keyness analysis provides a deeper understanding for government, initiative and referendum texts.

4.4 Complexity Analysis

The complexity can be analysed with different models. For a German text corpus the most commonly used one is SMOG.de from the Quanteda (2018) package. The SMOG.de applies this formula:

$$\text{SMOG.de Index} = \sqrt{\frac{30 \times \text{Polysyllabic Words}}{\text{Number of Sentences}}} + 3.1291$$

A SMOG.de score of 10 equals the text understanding of a 10th grader, roughly 15 to 16 years old, in Switzerland (SMOG Index).

4.5 Sentiment Analysis

A sentiment analysis compares texts with a pre-defined dictionary and allocates negative and positive scores to a text. Subtracting the negative from the positive score the net-sentiment was calculated.

As the Swiss voting booklet is written in the German language a sentiment dictionary capable of assessing the German language had to be applied. Rauh's German Political Sentiment Dictionary (Rauh C.) was chosen as it is the most elaborated dictionary for German text analysis. Furthermore, it is built as a political language dictionary and can therefore be applied to the Swiss voting booklet.

5 Findings

5.1 Frequency

The frequency for the texts of Government, Initiative and Referenda were analysed and the 30 most frequent terms compared.

Table 5.1: Top 4 Frequency

Frequency	**Government**	**Initiative**	**Referenda**
1.	Bundesrat (Government)	Initiative	Nein (No)
2.	Schweiz (Switzerland)	Schweiz (Switzerland)	Schweiz (Switzerland)
3.	Initiative	mehr (more)	mehr (more)
4.	Parlament (Parliament)	ja (yes)	Franken (Swiss Franc)

Analysing this table it is possible to get an understanding what the messages are. The Government explains the voting and gives the recommendation for Switzerland and backs this up by arguing that the parliament discussed the vote as well. Furthermore, the Government is almost always against the Initiative, so therefore it is feasible that it has such a high frequency. The Initiative and Referenda show a similar picture but put more stress on voting yes or no. The term "more" is feasible as the initiative is against (example) "more immigration" for "more retirement money" whilst the Referenda is mostly against "more taxes" and "more restrictions".

5.1.1 Keyness

The Keyness comparison between Government with Initiative and with Referenda texts portrays a more refined picture from the frequency analysis.

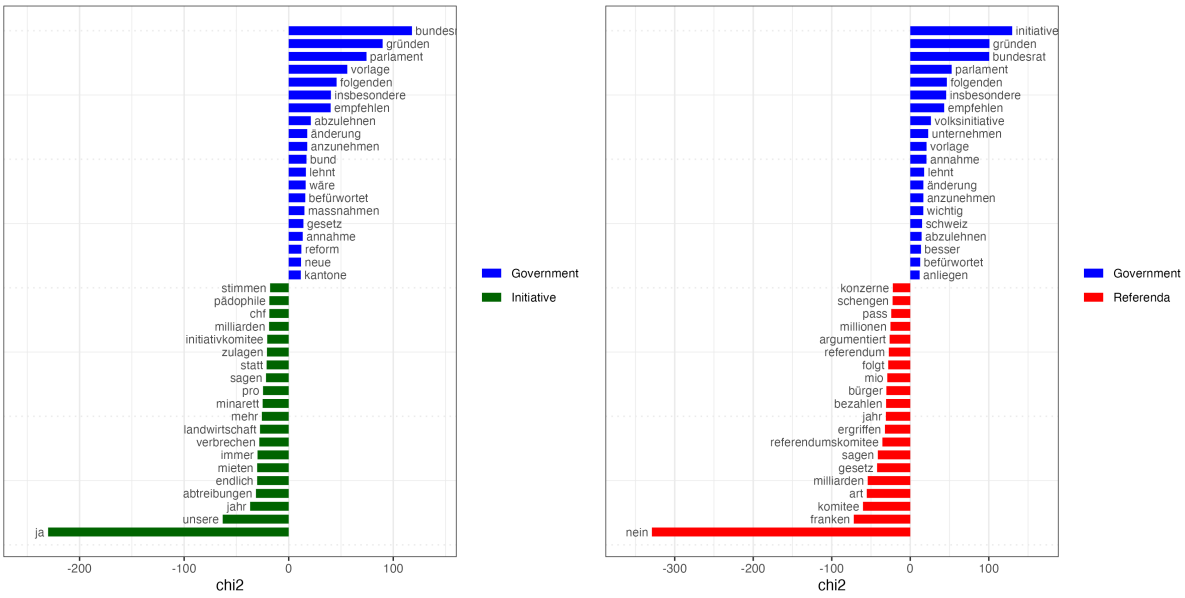

Figure 5.1: Keyness Govt. vs. Initiative and Referenda

Moreover, it indicates that the Government is consistent in their argumentation when it comes to Initiative as well as Referenda as the main keys are almost the same. The comparison to Referenda and Initiative gives us a good picture that "yes" and "no" are the main keys in the text whilst also giving insights into terms that provide a rough picture on the wording style. Therefore, a KWIC-Analysis (Keyword in Context) was applied, showcasing with Initiative keywords like "unsere" (our), "endlich" (finally), "immer" (always) it can be understood that the Initiatives associate an unfavourable state or a better outcome that the initiative is going to create. Same can be interpreted for Referenda, where the focus is mostly on money related terms like "franken" (Swiss Francs), "milliarden" (billions), "bezahlen" (pay).

5.2 MTTR

The corpus consists of 117.170 tokens and 66.626 respective types. The overall MTTR indicates an equilibrate value of 0.58.

The most complex texts are initiatives with a MTTR of 0.61, followed by referenda with 0.60 and the government text are the least complex ones with a MTTR of 0.55.

The MTTR of government texts seem to be lowering which indicates that the complexity could be decreasing whereas initiative complexity seems to be rising. MTTR score of

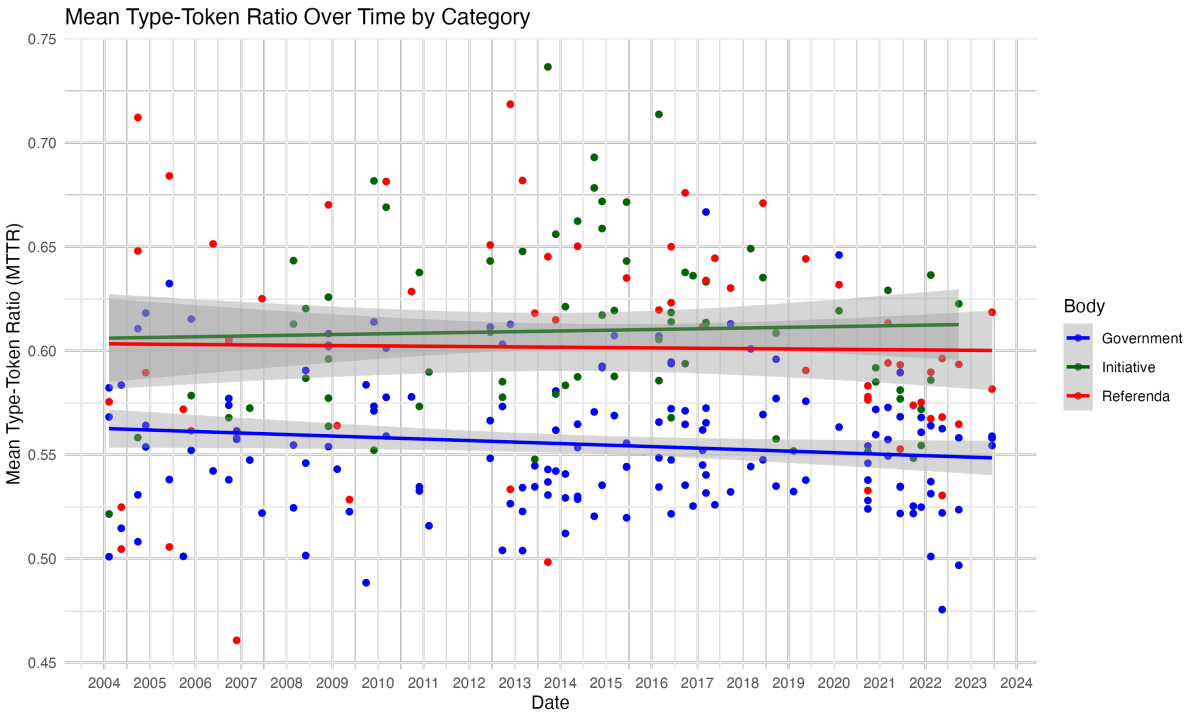

Figure 5.2: MTTR 2003 until 2024

referendum is stagnant.

5.3 Complexity

Interpreting the upper plot we can see that the government texts are the most complex and that there was a slight drop after 2016. referendum and initiative texts are below a SMOG.de index of 10 but both had an increase after 2017 whereas the latter has seen a sharper rise in complexity.

Overall, interpreting the first plot, it seems that the complexity did drop a little for government texts but for initiative and referenda it increased in complexity.

The second plot with the LM-Line supports the interpretation.

5.3.1 SMOG.de Index and Slopes

Analysing the slopes of government, initiative and referenda and their development in the observed period, only the coefficient of the initiative has an almost significant p-value.

- government: coefficient: -3.700e-05 / p-value: 0.418
- initiative: coefficient: 0.0002365 / p-value: 0.05110
- referendum: coefficient: 8.239e-05 / p-value: 0.244

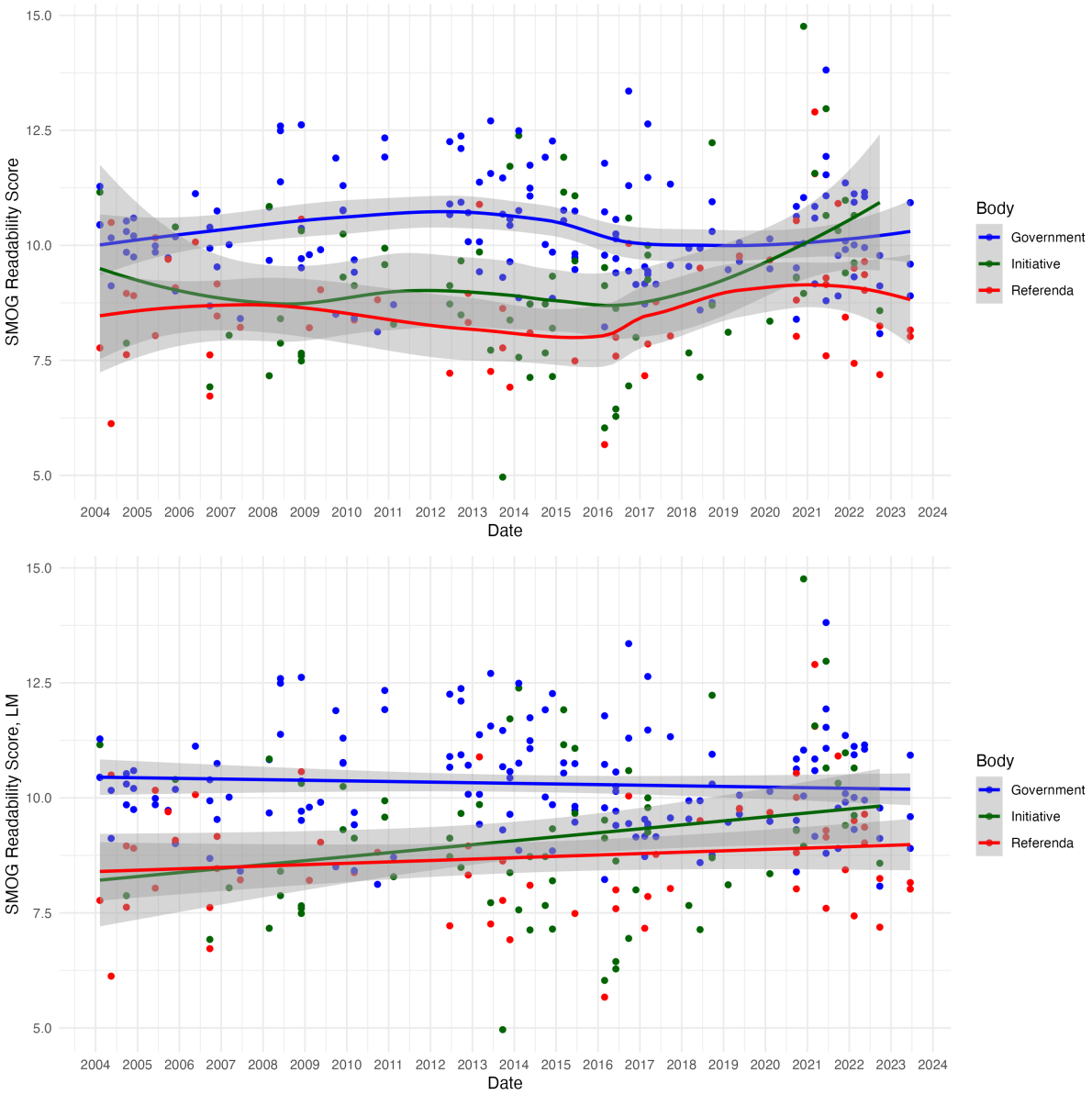

Figure 5.3: SMOG.de Index over years

The government model has, although small and statistically insignificant, a negative coefficient which would indicate the reduction of complexity in their texts.

5.4 Sentiment

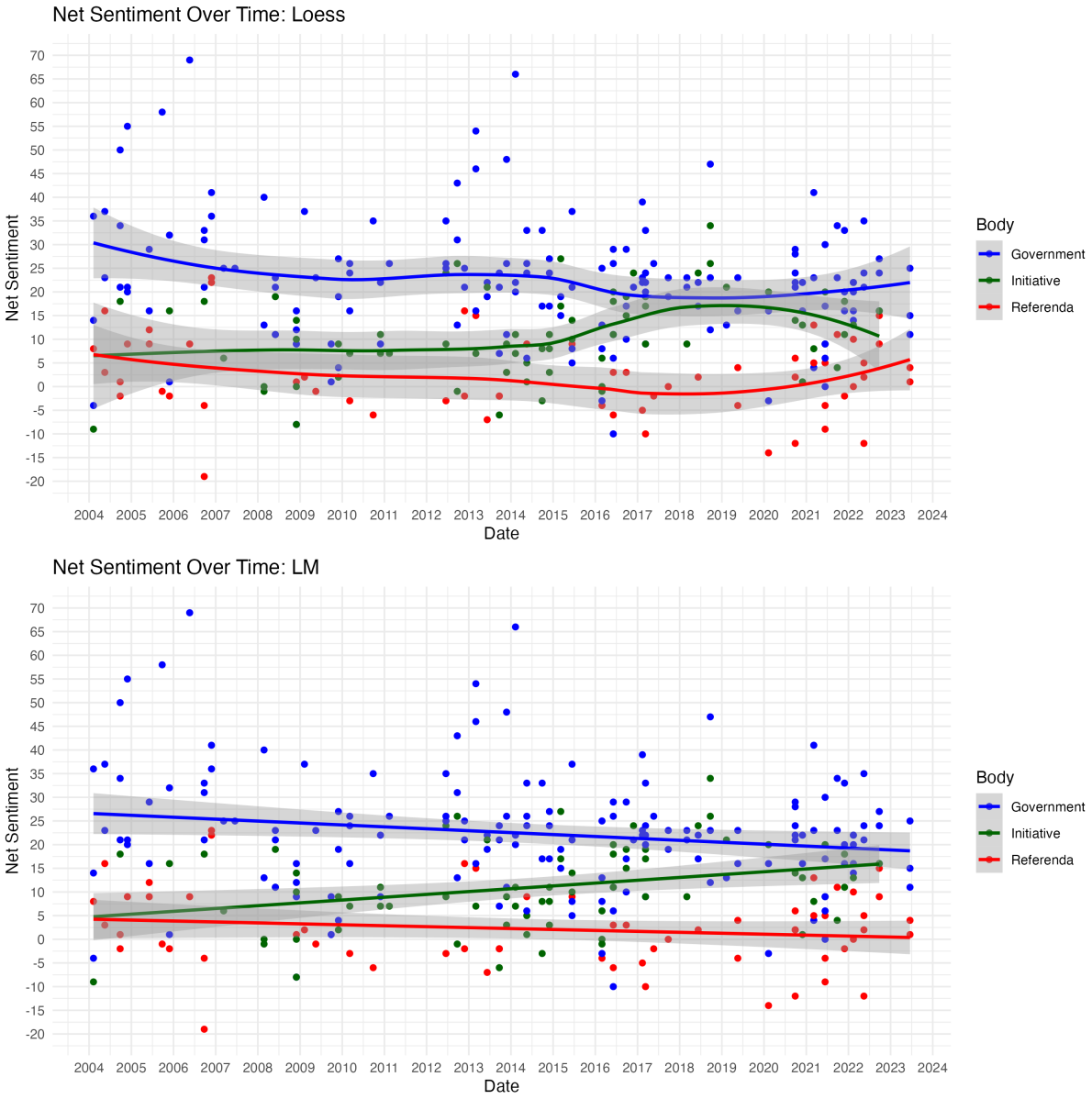

Figure 5.4: Net Sentiment

Interpreting the first plot of the sentiment analysis, it was visible that the government sentiment is in general the most positive. It seems that the government sentiment is becoming less positive whereas the initiative rose strongly towards more positive sentiments from the year 2016 and then declined after 2021. The sentiment for referendum is mostly close to zero and in general negative. There is a small indication towards 2024 that the positivity is accelerating.

Fitting a linear model to government, initiative and referendum significant effects were observed.

- government: coefficient: -0.001114 / p-value: 0.0299
- initiative: coefficient: 1.635e-03 / p-value: 0.00713
- referendum: coefficient: -0.0003526 / p-value: 0.459

Significant effects on the net sentiment over time for the government and initiative texts were found.

6 Discussion

The hypothesis that the texts in the Swiss voting booklet are becoming less complex has to be rejected as not enough evidence could be found to support this hypothesis. This research even found almost significant indications that the complexity of texts for initiative is increasing. The average SMOG.de score between 7.5 and 11 reflects the fact, that Swiss citizens have to attend 11 years of education and are then, expected to be between 17-19 years old, and able to vote.

The sentiment analysis reflected the initial findings in MTTR. Referenda texts are by default negative and initiative positive. The hypothesis that the rise of populism and challenging economic times affect negative sentiments found within the Swiss voting booklet can be somewhat supported. Whereas evidence for more negative sentiments in the government texts was found the trend to more positive sentiments in the initiative texts was observed.

The Swiss voting booklet provides only insights into the last aspect of campaigning for voting in Switzerland.

Further research could assess whether allocating the initiative and referendum committees to their political spectrum as well as a sentiment and complexity analysis of their campaigning could give a better understanding of trends.

Bibliography

K. Benoit, K. Munger, and A. Spirling. Measuring and Explaining Political Sophistication through Textual Complexity. *American Journal of Political Science*, 63(2):491–508, 4 2019. ISSN 15405907. doi: 10.1111/AJPS.12423.

S. E. Bestvater and B. L. Monroe. Sentiment is Not Stance: Target-Aware Opinion Classification for Political Text Analysis. doi: 10.1017/pan.2022.10. URL https://doi.org/10.1017/pan.2022.10.

D. Bischof and R. Senninger. Simple politics for the people? Complexity in campaign messages and political knowledge. *European Journal of Political Research*, 57(2):473–495, 5 2018. ISSN 1475-6765. doi: 10.1111/1475-6765.12235. URL https://onlinelibrary.wiley.com/doi/full/10.1111/1475-6765.12235https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6765.12235https://ejpr.onlinelibrary.wiley.com/doi/10.1111/1475-6765.12235.

M. Degani. Endangered Intellect: A Case Study of Clinton Vs Trump Campaign Discourse. *Iperstoria*, (8), 12 2016. ISSN 2281-4582. doi: 10.13136/2281-4582/2016.I8.392. URL https://iperstoria.it/article/view/392.

J. Gerstlé and A. Nai. Negativity, emotionality and populist rhetoric in election campaigns worldwide, and their effects on media attention and electoral success*. *European Journal of Communication*, 34(4):410–444, 8 2019. ISSN 14603705. doi: 10.1177/0267323119861875/ASSET/IMAGES/LARGE/10.1177{_}0267323119861875-FIG9.JPEG. URL https://journals.sagepub.com/doi/full/10.1177/0267323119861875.

A. Noury and G. Roland. Identity Politics and Populism in Europe. *https://doi.org/10.1146/annurev-polisci-050718-033542*, 23:421–439, 5 2020. ISSN 10942939. doi: 10.1146/ANNUREV-POLISCI-050718-033542. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-polisci-050718-033542.

Quanteda. quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774, 10 2018. doi: 10.21105/JOSS.00774.

Rauh C. data_dictionary_Rauh: Rauh's German Political Sentiment Dictionary in quanteda/quanteda.sentiment: Sentiment Analysis using 'quanteda'. URL https://rdrr.io/github/quanteda/quanteda.sentiment/man/data_dictionary_Rauh.html.

SMOG Index. What Is the SMOG Index? | Free Readability Test Tool. URL https://www.webfx.com/tools/read-able/smog-index/.