

Emotions in the parliament: Lexical emotion analysis of parliamentary speech transcriptions

Egon Werlen¹
egon.werlen@ffhs.ch

Christof Imhof¹
christof.imhof@ffhs.ch

Per Bergamin¹
per.bergamin@ffhs.ch

¹Swiss Distance University of Applied Sciences (FFHS)

Abstract

Politics is emotional. So far, relatively few studies investigated the emotional content in parliamentary speeches. In this study, we analysed emotional valence and arousal of German and French speeches of a Swiss cantonal parliament and whether we can use them to predict the membership of parliamentarians to one of two groups: those who won more of the votings than others. The emotional text analysis showed that these speeches are indeed emotional. However, the results regarding the predictions were mixed. Arousal and language showed no effects and valence was only partially successful as a predictor.

1 Introduction

Ever since it came into existence, politics has exerted influence on the daily life of humans all over the world. For a long time, the idea prevailed that politics has to be rational rather than emotional. However, it is not surprising that many political issues are emotional at their core. This leads to debates about very emotional topics, which are not always handled as rationally as one might assume. Audible and visible evidence is provided by debates on the Internet and on television: parliamentarians cheer, yell, throw things, and even have fistfights on rare occasions. Moreover, political campaigns often aim at emotionally relevant aspects of political topics rather than the actual ramifications of the topic at hand (e.g. Widmann, 2021; Erisen and Villalobos, 2014). Thus, politics is very emotional. Today, in modern parliaments (e.g. Switzerland, Germany, France, the UK, or the European parliament) verbatim protocols as well as videos are recorded and used for tracking and archiving. The advantage

of videos is that a large part of the observable verbal and non-verbal signals of emotional states in political speeches (e.g. posture, gestures, facial expressions, phonology, speaking style) can be traced. In transcribed form, these signals are no longer represented to the same extent. However, some emotional characteristics remain. These are primarily the emotional potential of words and other linguistic features like phonemes, accent, number of syllables and letters, and word frequency. This is where we come in with the present exploratory study. We want to find out whether it is possible to estimate emotional states on the two dimensions valence and arousal in literal transcripts of parliamentary sessions with a rather simple lexical method. Further, we intend to find out whether we can predict the parliamentary groups that lose more votings than the average of lost votings in the parliament by the emotional content of the speeches. We use this subdivision into 'vote winners' and 'vote losers' as an analogy to the more common contrast between ruling party and opposition found in many other countries. This differentiation is found in many studies on parliamentary speeches. In Switzerland however, there is no classical division into governing and opposing party. Instead, the parliaments of Switzerland and its 26 cantons are built on consensus, which is why another approach was needed to differentiate between parliamentary groups.

2 Theoretical background

2.1 Emotions

Roughly classified, there are three basic paradigms in emotion research (Holodyski and Friedlmeier, 2012). The first one is the structural emotion paradigm (Izard, 1991; Panksepp, 1998; Ekman, 1999) in which emotions are defined as specific mental states. In the second one, the functional paradigm, emotions are viewed as a set of specific mental functions, defined as changes in the disposition to act and help the individuals to ad-

just their motives and intentions to the changes (for example: Frijda, 1986; Lazarus, 1991; Scherer, 1999). Under the third paradigm, the contextual paradigm, emotions are defined as socially and culturally constructed psychological functions resulting from interpersonal interactions (for example: Lutz and White, 1986; Matsumoto et al., 2008). In general, it can be observed that many political studies follow the functional paradigm (e.g. Lara et al., 2016).

From a different perspective, according to which emotional feelings are sometimes expressed as emotional colouring, further fundamental distinctions of theoretical approaches can be found. There are theories that start from different distinct emotions (e.g. joy, fear, anger, surprise). Well-known approaches include the basic emotion theory of Ekman (1999) or the process component theory of emotions of Scherer (2010). The latter assumes that every emotion consists of five components (cognitive, physiological, motivational, motor expression, subjective feeling). Other theories assume that emotions are based on two or three dimensions with high and low emotional levels. For example, Bradley and Lang (1994) postulate three dimensions: emotional value, emotional arousal and emotional dominance. Another very prominent representative of this approach is Bertrand Russell with his collaborators (Barrett and Russell, 1999). In the Circumplex model two emotional dimensions are postulated, namely the emotional valence and the emotional arousal. Valence refers to the experience of one's own positive or negative feelings. Arousal refers to the experience of the intensity, the activation level of one's own feelings. Both dimensions form the "core affect", as "the most elementary, consciously accessible affective feelings, which do not have to be directed at anything" (p. 806).

2.2 Emotions in politics

In formal discourses, such as parliamentary speeches, one assumes that fewer emotions are expressed, compared to everyday conversations. Day-to-day conversations seem to offer more immediacy and closeness and thus stimulate the expression of emotions (Lara et al., 2016). Historically, emotions have been part of public and political life as in the case of the Greeks, Machiavelli or Hume. Throughout the 20th century, however, emotions were not considered important in politics and social life. This changed in the 1990s, when interest in human emotions grew in various disciplines such as psychology, neuroscience, sociology and philosophy. This led to the rediscovery of emotions in political science (Hoggett and Thompson, 2012) and the systematic use of emotions in democratic systems, for example, by politicians in election pro-

cesses, debates and written texts (Freeden, 2013). Political science often looks at things from the functional paradigm perspective. Barbalet (1998) and Freeden (2013) assume that emotions are common everyday processes. They influence political thinking through three syntactic functions in that they (1) emphasise concepts by reinforcing morphological structuring, (2) relativise meanings by classifying importance, or (3) reduce or reinforce connections. In their qualitative research, Lara et al. (2016) form functional categories in parliamentary discourses by assuming that emotions are used to "emphasise the speaker's argumentation", "attack the opponent", "express proximity and create a distinctive 'identity' with respect to the rest of the group", and emotions are also "used as an argument itself" (p. 155).

2.3 Emotion analyses methods for texts

In order to measure emotions in speech and text, an analytical framework is first needed that helps to reduce the number of categories (Cowie and Cornelius, 2003). In the present study we have chosen to describe emotions based on the circumplex model of Barrett and Russell (1999) with its two-dimensional classification of emotions (valence and arousal). Furthermore, we use a lexical approach based on individual words. From a technical point of view, word-based lexical analysis can be classified as a semantic approach to sentimental analysis, but it does not necessarily implement machine learning. This type of approach is historically based on early work by Freud (1891) and Bühler (1934), who assumed that spoken or written words have the potential to elicit both overt or covert sensu-motoric or affective reactions. From this point of view, words can evoke both basic and induced emotions (Jacobs et al., 2015).

Lexical analysis usually relies on word lists, consisting of thousands of words whose values (e.g. valence, arousal, dominance etc.) were previously validated as the result of rating procedures. Examples of such lists are the Affective Norms for English Words (ANEW; Bradley and Lang, 1999), the Warriner list of norms for valence, arousal and dominance for English lemmas (Warriner et al., 2013), the NRC-VAD lexicon (National Research Council Canada - Valence, Arousal, Dominance; Mohammad, 2018), the Berlin Affective Word List (BAWL-R; Vö et al., 2009), the Semantic Lexicon of Emotion (SLE; Leleu, 1987) or the French interlingual metanorm for the emotional analysis of texts (EMONORM; Leveau et al., 2012). In many cases, the emotional valence and arousal of texts is calculated by averaging the values for valence and arousal of all words contained within. However, values can also be derived for smaller units such as sentences or paragraphs. Such a procedure

has been used in the context of political studies in the analysis of "emotional conversations" by [Lara et al. \(2016\)](#) or the analyses of emotional words by [Koschut \(2020\)](#), to name two examples.

The BAWL-R is the largest German emotional word list and has been utilised for the analyses of different text forms: poems ([Aryani et al., 2016](#); [Ullrich et al., 2017](#)), E.T.A. Hoffmann's black-romantic story "The Sandman" ([Lehne et al., 2015](#)), passages of Harry Potter novels ([Hsu et al., 2015](#)), Shakespeare's sonnets ([Jacobs et al., 2017](#)), and short stories ([Werlen et al., 2018, 2019](#)). In all these studies, the mean of the affective values of the individual words correlated with the whole text ratings. Studies implementing the BAWL-R to predict subjective emotional states of short texts ([Hsu et al., 2015](#)) and poems ([Ullrich et al., 2017](#)) found correlations for lexical valence with subjective valence of $r = .53$ and $r = .65$, and for lexical arousal with subjective arousal of $r = .59$ and $r = .54$. The SLE was validated by [Leleu \(1987\)](#) and was implemented in experimental studies (e.g. [Degner et al., 2012](#); [Jhean-Larose et al., 2014](#)). Despite there not being a similar comparison between lexical and subjective values as in the case of the BAWL-R, the SLE is relevant to this study because it is the only French word list we are aware of that includes words rated on both emotional valence and arousal.

An alternative could be the NRC-VAD lexicon by [Mohammad \(2018\)](#). This lexicon contains 20,007 annotated words in 103 languages. The English words were annotated with the help of Amazon MTurk for valence, arousal, and dominance using the best-worst scaling method. The translation of the English words into the other languages was accomplished by using Google Translator. The values for valence, arousal and dominance were taken from the English version on the assumption that the values are stable for different languages. In an unpublished study, we compared the NRC-VAD with the BAWL-R in an emotional text analysis of 62 short stories in German and their English translation. The English version of the NRC-VAD correlated with the human ratings of the English texts to a similar extent as the BAWL-R correlated with the human ratings of the German texts. However, in the German version of the NRC-VAD, the correlation values with the human ratings of the German texts were considerably lower than with the BAWL-R. Consequently, the German translation of the NRC-VAD lost some of its predictive power. For this reason, we decided not to use this large lexicon, even though it contains both languages of interest to us.

2.4 Studies of emotions of transcribed parliamentary speeches

The number of studies that analyse parliamentary speeches for their emotional content is growing but still limited. [Abercrombie and Batista-Navarro \(2020\)](#) reviewed 61 studies, 28 were looking for sentiment polarity and three for emotions; 16 worked with dictionary based methods. The same goes for studies establishing a relationship between expressed emotions in the speeches and the role of the parliamentary group (governing or in opposition). [Abercrombie and Batista-Navarro \(2020\)](#) found 14 studies predicting some form of party affiliation.

One example is a study by [Riabinin \(2009\)](#), who classified politicians in the Canadian Parliament based on the dimension Liberal vs. Conservative with a Support Vector Machine using the categories of the Linguistic Inquiry and Word Count (LIWC) by [Pennebaker et al. \(2015\)](#). The authors used the Canadian Hansard, which includes the English and French House of Commons debates. One might assume that the expression of positive (empathy) or negative emotions (contempt) was connected to these specific political ideologies (see [Freedden, 2013](#)), which appeared to be the case in this study, at least at face value: the authors found that in the speeches of the 36th Parliament, the Liberals generally used positive language, while the Conservatives used more negative words. However, they suppose that this difference is not due to party affiliation, but rather the fact that the Liberals were the governing party and the Conservatives were in opposition. [Hirst et al. \(2014\)](#) conducted the same analysis with the speeches of the 36th Parliament, but added the ones from the 39th Parliaments as well, where the roles were switched. In both cases, the respective opposition showed more negative emotions in its speeches than the governing party, which the authors concluded was due to a "language of attack and defence" (p. 93). The differences due to political ideology or party affiliation were thus negligible, confirming the assumption by [Riabinin \(2009\)](#). In this context, it should be noted that the authors of both studies used partially translated speeches, as the Canadian Parliament is bilingual. The French speeches were first translated into English before the analysis. The bilingualism of the speeches and the subsequent translation may therefore have had an influence on the results.

Another example is a study by [Rheault et al. \(2016\)](#), where the British Hansard was used, which includes the transcripts of all parliamentary debates of the British House of Commons between 1909 and 2013. To analyse emotional polarity as a standardised measure from -1 (negative) to +1 (positive), they created a domain-specific lexicon based on the affective content of expressions to ob-

tain an indicator of emotional words in the British Parliament. The mood of politicians of the British parliament was found to have become more positive during the last decades, and the valence of the politicians' speeches fluctuated in accordance with economic business cycles (e.g. indicator of recession, and indicator of labour conflicts).

To our awareness, there are no studies on emotional arousal in parliamentary speeches.

2.5 Research questions and hypotheses

The overall goal of this study is to replicate the results of the studies analysing speeches of the Canadian and British parliaments and to extend them.

As shown in the abovementioned studies, it is possible to estimate emotions in parliamentary speeches. All of them estimated positive-negative emotional states that generally correspond to the emotional valence of the circumplex model (Barrett and Russell, 1999). We intend to extend these results by measuring not only emotional valence, but emotional arousal as well, the second dimension of the circumplex model of emotions. Therefore, the first research question concerns our ability to estimate emotional valence and emotional arousal in parliamentary speeches with our emotional text analysis approach.

The transcribed speeches we analysed stem from a cantonal parliament in Switzerland. The political systems of Switzerland and its cantons do not have a typical government - opposition structure. On first glance, this poses a problem for our replication in light of the results presented above: the prediction of party affiliation or ideology by emotions in the speeches of parliamentarians is, as the study of Hirst et al. (2014) shows, confounded with the division into government and opposition rather than political ideology. In the parliament we analysed, there is no true opposition since most parliamentary groups are represented in the government. Therefore, the definition of an opposition cannot refer to the parliamentary groups alone. Thus, we chose a different operationalisation approach: We examined the proportion of lost votings during the three session weeks that we analysed and the groups that lost more votings were thus defined as the oppositional groups. According to Riabinin (2009), in a parliament with a real opposition, one would assume that the opposition would show more negative emotions in their speeches. Since more negative emotions are usually associated with higher arousal (Kuppens et al., 2013), the opposition would also show more arousal in their speeches. We assume that these correlations are also present with our operationalisation of the opposition as groups with more lost votes.

Research questions

1. Do parliamentary speeches contain emotional information (valence, arousal)?
2. Are there differences in the emotional state of speeches between parliamentarian groups that lost more votings compared to groups that lost fewer votings?

Hypotheses

1. Speeches by members of parliamentary groups with fewer lost votings indicate more positive emotional states than speeches from members of parliamentary groups that lost more votings.
2. Speeches by members of parliamentary groups with fewer lost votings indicate less arousal than speeches from members of parliamentary groups that lost more votings.

3 Methods

3.1 Samples and measurements

For the analyses and the testing of the hypotheses, we used all the transcribed speeches from three sessions, which each occurred within a week in the month of June, September, and November 2019 of a Swiss cantonal parliament (Valais). The parliament includes 130 parliamentarians and 130 substitutes. The speeches of the government representatives (i.e. the five members of the cantonal council) and the president of the parliament were not included in the analyses. The president of the parliament leads and moderates the debates but does not usually contribute to their content and the cantonal council members are not part of the parliament. The parliamentary speeches are automatically transcribed by the company recapp IT AG (<https://recapp.ch>) using AI algorithms. The transcripts are checked by the administration, corrected and formatted, including the insertion of the agenda items and other notes such as information about beginning and end of each session. The literal minutes are published in the original language on the cantonal website (https://parlement.vs.ch/app/de/search/result?object_type=ParlSession).

In order to categorise the parliamentary groups, we first calculated the percentage of won and lost votings of all groups during the three sessions, consisting of 20 half days. The parliament voted 196 times, without counting the issues that were uncontroversial and did not lead to a vote. The 257 parliamentarians - present at least at one voting - cast a total of 22'963 individual votes.

Since speeches are usually given in the mother tongue of the speaker, in this case German or

French, we opted for analysing the original speech contents with language-specific word lists. For the German speeches, we used the revised form of the Berlin Affective Word List (BAWL-R; Vö et al., 2009), while the Semantic Lexicon of Emotions (SLE; Leleu, 1987) served as the word list for the French speeches. In total, we analysed the speeches of 179 parliamentarians from all nine parliamentary groups. Within the three sessions, the parliamentarians held a total of 345 speeches, each lasting up to five minutes. The speeches contained 329'031 words and 16'630 sentences. In German, 72'092 words in 6462 sentences were counted, of which 7443 words (10%) were included in the annotated word list. In French, 256'939 words in 10'168 sentences were counted, of which 24'535 words (10%) were contained in the word list. On average, each speech consisted of 911 words, of which an average of 89 words were represented in the annotated word lists.

The BAWL-R is a large German word list containing almost 3000 words (nouns, verbs, and adjectives) from the CELEX database (Baayen et al., 1996). Each word of the list was rated on valence, arousal, and imageability indicating the feeling when reading each word. The list also includes psycholinguistic factors (e.g. number of letters, phonemes, word frequency, accent). It is free for download¹. The BAWL-R enables estimations of the emotional potential of single words but also extrapolations for sentences and whole texts. In the BAWL-R (Vö et al., 2009), valence had been rated with the Subjective Assessment Manikin (SAM; Bradley and Lang, 1994) on a 7-point scale (-3 very negative through 0 neutral to +3 very positive), and arousal on a 5-point SAM-scale (1 low arousal to 5 high arousal). The split-half reliabilities of the original BAWL-R data can no longer be calculated. According to oral communication with Jana Lüdtke (Free University of Berlin), the split-half reliability with data from a new rating of 466 words resulted in a value of .97 for valence and .92 for arousal. The Semantic Lexicon of Emotion (SLE; Leleu, 1987) is part of an unpublished master thesis that was integrated in the interlingual metanorm for emotional analysis of texts (EMONORM; Leveau et al., 2012). We used the 3000 values for valence and arousal published by Leveau et al. (2012) that were transformed into the interval 1- to +1.

3.2 Analyses

After selecting the specific sessions we were interested in, we downloaded the list of votings and merged them into one data frame in order

¹<https://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/allgpsy/Download/BAWL/index.html> accessed May 2019; To open the file a password must be requested.

to calculate the percentage of lost votings and agreement with the parliamentary group using R (Core Team, 2017). We downloaded the PDF files containing the speeches from the file sever of the canton with a custom Python script, which also served the purpose of immediately splitting the text body based on individual speeches. The resulting files were subsequently further processed in R, where the speeches were first split into chunks with regex functions. Using the R-package `cldr` (McCandless et al., 2013), we identified the language of the text in each chunk (i.e. either French or German) and split the data frame in two based on that information. We then implemented `spacyr` (Benoit and Matsuo, 2019) separately on both subsets in order to tokenise and lemmatise their contents, which were subsequently matched with one of two data bases, again separated by language. For the German transcripts, the semantic lexical analysis was conducted with the BAWL-R (Vö et al., 2009). The French transcripts were analysed with the SLE (Leleu, 1987). After the removal of duplicate entries from the database with rules based on functions from the package `RecordLinkage` (Borg and Sariyar, 2019) and adjusting the scales in the French database to match the German ones, the subsets were reunited and further analysed. In addition to the packages mentioned above, we used `brms` (Bürkner, 2018), `tidybayes` (Kay, 2020), `ggplot2` (Wickham, 2016), `plotly` (Sievert, 2020) and `tidyverse` (Wickham, 2017). For each speech, we averaged the valence and arousal of all the words in that speech represented in the BAWL-R for German speeches and the SLE for French speeches. To answer the two research questions, the mean variance and mean arousal of all speeches of each parliamentarian was calculated for each session week. Neglecting the fact that a parliamentarian can have speeches with positive and negative emotional content, or negative or positive emotional content within a single speech. We have not included the variation of values within the speeches of individual parliamentarians in our analyses.

4 Results

Across all parliamentary groups, parliamentarians lost 22% of the votings. We found four parliamentary groups that lost about a third of the votings (34%) with values from 32% to 35%. The remaining five groups lost 14% of the votings within the three session weeks. Depending on the parliamentary group, the value was between 11% and 18% (see table 1). The parliamentarians voted mostly in agreement with their respective groups. Only 3% of the votes were cast in disagreement with the group.

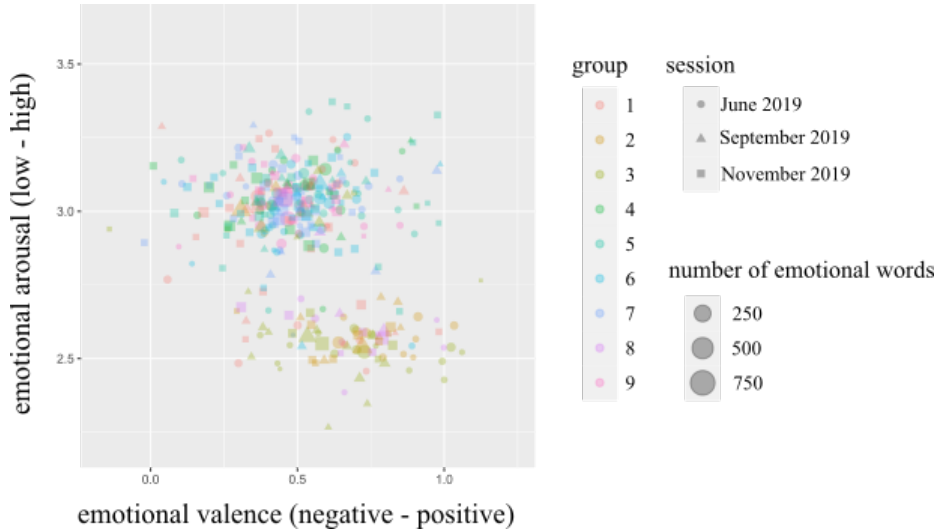


Figure 1: Distribution of emotional valence and emotional arousal in parliament speeches.

| Parl. group | Lost votes | Valence Mean | Valence SD | Arousal Mean | Arousal SD |
|---------------|------------|--------------|------------|--------------|------------|
| <i>Group1</i> | 11% | 0.49 | 1.06 | 3.06 | 0.73 |
| <i>Group2</i> | 12% | 0.52 | 1.04 | 3.02 | 0.70 |
| <i>Group3</i> | 13% | 0.62 | 1.11 | 2.71 | 0.60 |
| <i>Group4</i> | 18% | 0.48 | 1.10 | 3.03 | 0.71 |
| <i>Group5</i> | 18% | 0.64 | 1.07 | 2.62 | 0.54 |
| <i>VWin</i> | 14% | 0.55 | 1.08 | 2.88 | 0.68 |
| <i>Group6</i> | 32% | 0.46 | 1.07 | 3.04 | 0.71 |
| <i>Group7</i> | 33% | 0.46 | 1.10 | 2.96 | 0.68 |
| <i>Group8</i> | 35% | 0.48 | 1.08 | 3.04 | 0.69 |
| <i>Group9</i> | 35% | 0.67 | 1.09 | 2.64 | 0.50 |
| <i>VLose</i> | 34% | 0.49 | 1.09 | 2.96 | 0.69 |
| <i>All</i> | 22% | 0.52 | 1.08 | 2.92 | 0.69 |

Note. Parl. group=Parliamentarian group; VWin=vote winners; VLose=vote losers; SD=standard deviation

Table 1: Emotional valence and emotional arousal in the speeches of the parliamentarian groups

The emotional text analysis confirms the first research question. In the transcribed speeches, emotional states, specifically valence and arousal, can be estimated with a sufficiently large variance. In the last line of table 1, the means and standard deviations of emotional valence and emotional arousal of the total sample are listed. The mean of emotional valence is 0.52 with a standard deviation of 1.08 (absolute range: from -0.90 to 1.40). The mean of emotional arousal is 2.92 with a standard deviation of 0.69 (absolute range: from 2.25 to 3.37). The ranges of the values for valence and arousal are rather narrow. But they are still twice as large as the corresponding values of an analysis of 62 emotional short text with a range of 1.15 points (0.02 to 1.17) for valence and the range for arousal (2.34 to 2.92; 0.58 points; Werlen et al., 2019). Other studies that analysed different text

types show comparable value ranges to the values of the present study for valence (Hsu et al., 2015; Jacobs et al., 2017; Jacobs and Lüdtke, 2017) and arousal Jacobs and Lüdtke (2017). To be able to classify this result, it is helpful to know the values of emotionally neutral or non-emotional speech. From a purely theoretical point of view, a neutral text has a valence close to 0 and an arousal around 2.5. Three short stories included in the analysis of Werlen et al. (2019) that were deliberately written in an emotionally neutral way have valences close to 0.5 and an arousal close to 2.5. In comparison, the transcribed speeches of our study have values ranging from neutral to significantly stronger emotional arousal. The same is true for valence compared to a theoretical neutral valence. Compared to the emotionally neutral texts, the valence of the parliamentary speeches varies in both directions, negative and positive.

Figure 1 shows the distribution of emotional valence (x-axis) and emotional arousal (y-axis) across all speakers. The different colours represent the nine parliamentary groups. The range of values for single words for valence is -3 to +3, for arousal 1 to 5. Due to the aggregation of single words values into values for each speech, the possible value span got narrower. We estimate the actually possible value span in the speeches for valence and arousal to lie within two standard deviations, i.e. between -1.5 and 2.5 for valence, and between 2.2 and 3.6 for arousal. The scaling in table 1 is adjusted accordingly. Generally, the illustration shows that valence has a wider distribution than arousal. Arousal is divided in two sections: The section with a higher arousal contains mostly speeches in French, the lower arousal section speeches in German.

Table 1 shows also the percentages of lost votes, and the means and standard deviations for

emotional valence and emotional arousal of the nine parliamentarian groups. The vote winners (v win) have a more positive average valence, with a mean value of 0.55 (standard deviation: 1.08) than the vote losers (v lose) with a mean value of 0.49 and standard deviation of 1.09. With regard to arousal, the vote winners have a lower emotional arousal (mean value: 2.88, standard deviation: 0.68) than the vote losers (mean value: 2.96, standard deviation: 0.69). However, the differences in valence and arousal between vote winners and vote losers are very small.

In order to address the second research question, i.e. whether emotional valence and arousal are able to predict the membership of parliamentarians in one of two groups (fewer lost votings vs. more lost votings), we calculated several Bayesian regression models. Since parliamentarians spoke multiple times across the three different sessions, resulting in repeated measures, we decided to calculate multilevel models using brms (Bürkner, 2018) with session as the grouping factor. Model 0 was an intercept-only model, model 1 added the speeches' *valence* and *arousal* values as predictors plus the *session* as a categorical predictor, and model 2 added *language* as a fourth predictor. In order to reflect the nested structure of our data, models 1 and 2 were each calculated twice, once with fixed effects and once with additional random effects, allowing the relation between the variables to be moderated by the grouping factor *session*. In order to inspect the role the word lists may play, we conducted the analysis twice, once for each of the two French word lists (SLE and translated BAWL-R²). The German word list remained constant.

As an example of how the models were specified, the design formula for model 1 is shown here:

$$\begin{aligned}
 L_i &\sim \text{Binomial}(1, p_i) \quad [\text{likelihood}] \\
 \text{logit}(p_i) &= \alpha + \beta_v P_i + \beta_a P_i \quad [\text{linear model}] \\
 \alpha &\sim \text{Normal}(0, 10) \quad [\alpha \text{ prior}] \\
 \beta_v &\sim \text{Normal}(0, 10) \quad [\beta_v \text{ prior}] \\
 \beta_a &\sim \text{Normal}(0, 10) \quad [\beta_a \text{ prior}]
 \end{aligned}$$

First, we calculated the models with the SLE word list for the French speeches. The R-hat diagnostic with all R-hat values below 1.02 indicated good convergence for all estimated parameters in the models. However, emotional valence and arousal did not yield fixed effects in any of the models (see Table 2), and neither did language, as the credible intervals of these predictors always included 0. Random effects of valence and arousal were found in both random effects models, implying the relationship between the predictors and the group membership depends on the session (valence:

²Regarding the translation of a word list, see our remarks in the discussion section

$E_{model1RE}=.42, [.01, 1.76]$; $E_{model2RE}=.46, [.02, 1.94]$, arousal: $E_{model1RE}=.62, [.03, 2.09]$; $E_{model2RE}=.67, [.02, 2.69]$). These results did not confirm the two hypotheses that emotional valence and arousal of parliamentarians' speeches predicts the membership to parliamentarian groups with different percentages of lost votings. Therefore, we have to reject both of them. A comparison of the models with the Bayesian ELPD_LOO-criterion (theoretical Expected Log Pointwise Predictive Density - Leave One Out) showed that model 1 with random effects had the best fit, however the ranking is very unreliable due to the high standard errors, which are larger than their respective ELPD difference, with two exception (see table 4).

| Predictor | Estimate | Est.Error | l-CI | u-CI |
|------------------|----------|-----------|------|------|
| <i>Intercept</i> | -.48 | .18 | -.84 | -.14 |
| <i>Valence</i> | -.20 | .12 | -.44 | .05 |
| <i>Arousal</i> | .22 | .13 | -.03 | .47 |
| <i>Nov2019</i> | -.10 | .31 | -.70 | .51 |
| <i>Dez2019</i> | -.07 | .25 | -.56 | .43 |

Note. l-CI=lower lower limit credible interval; u-CI=upper limit credible Interval

Table 2: Prediction of political affiliation (vote winners vs. vote losers); Model 1 fixed effects with SLE (French) and BAWL-R (German)

Next, we calculated all of the models again, this time with the translated BAWL-R word list for the French speeches. All of the models converged again, as indicated by the low R-hat values. This time, a fixed effect emerged for valence in models 1 and 2 ($E_{model1FE}=-.34, [-.57, -.11]$; $E_{model2FE}=-.33, [-.63, -.03]$). Arousal and language again showed no effects (see Table 3). As before, random effects of valence and arousal emerged in both models (valence: $E_{model1RE}=.51, [.01, 1.99]$; $E_{model2RE}=.51, [.01, 2.02]$, arousal: $E_{model1RE}=.69, [.02, 2.41]$; $E_{model2RE}=.77, [.03, 3.08]$).

| Predictor | Estimate | Est.Error | l-CI | u-CI |
|------------------|----------|-----------|------|------|
| <i>Intercept</i> | -.46 | .18 | -.81 | -.12 |
| <i>Valence</i> | -.34 | .12 | -.57 | -.11 |
| <i>Arousal</i> | .23 | .12 | -.01 | .46 |
| <i>Nov2019</i> | -.11 | .31 | -.73 | .50 |
| <i>Dez2019</i> | -.11 | .26 | -.63 | .40 |

Note. l-CI=lower lower limit credible interval; u-CI=upper limit credible Interval

Table 3: Prediction of political affiliation (vote winners vs. vote losers); Model 1 fixed effects with BAWL-R for French and German

Figure 2 visualises the effects of both predictors using the BAWL-R for German and French speeches in model 1 (fixed effects). The figure

| Model | Diff ELPD | se | ELPD LOO |
|---------------------|--------------|------|-------------|
| <i>Leleu – BAWL</i> | | | |
| <i>Model1RE</i> | 0.00 | 0.00 | -229.50 |
| <i>Model0FE</i> | -0.01 | 3.12 | -229.50 |
| <i>Model1FE</i> | -0.05 | 1.17 | -229.54 |
| <i>Model2RE</i> | -1.04 | 1.20 | -230.53 |
| <i>Model2FE</i> | -1.16 | 0.26 | -230.65 |
| <i>BAWL – BAWL</i> | | | |
| <i>Model1RE</i> | 0.00 | 0.00 | -227.25 |
| <i>Model1FE</i> | -0.28 | 1.71 | -227.53 |
| <i>Model2FE</i> | -1.19 | 1.73 | -228.44 |
| <i>Model2RE</i> | -1.19 | 0.23 | -228.44 |
| <i>Model0FE</i> | -2.33 | 3.91 | -229.57 |

Note. FE=fixed effects; RE=random effects; ELPD=Expected Log Pointwise Predictive Density; LOO=Leave One Out; se=standard error

Table 4: Model fits: Model comparisons

shows the slope (blue line) with its 95% grey-shaded credible interval. Arousal has a large credible interval that includes 0, indicating no effect. The effect of valence is visualised with the narrower credible interval.

5 Discussion

The goal of this study was to find out if parliamentary speeches in a Swiss canton feature emotional content (valence and arousal) and whether that content is able to predict the membership of the speakers in one of two groups (one with fewer lost votings than the other, as an approximation of the more common divide between governing party and opposition). In line with our research question, we were able to estimate emotional states (valence, arousal) in the parliamentary speeches we analysed, with a rather narrow range of values for valence and even a narrower range of values for arousal. Nonetheless, these ranges were larger as the corresponding ranges in Werlen et al. (2019), where 62 emotional short stories were analysed in the same manner, or had a comparable range to other studies that analysed different text types (Hsu et al., 2015; Jacobs et al., 2017; Jacobs and Lüdtke, 2017). This indicates that assessing emotions by text analysis with annotated word lists works well in parliamentary speeches with a sufficiently large variance. In Figure 1, it is noticeable that the relationship between valence and arousal does not have the typical u-shape often found in the literature. But as Kuppens et al. (2013) show, depending on the origin of the data and the context of the study, the relationship between valence and arousal may take other forms. Interestingly, in our study, we have found two clusters that primarily concern the difference in arousal. The German-

language speeches have a lower arousal. This effect disappears when the French translation of the BAWL-R or the NRC-VAD of Mohammad (2018) is used, which indicates a problem with the word list of Leleu (1987).

Predicting the membership of speakers in parliamentary groups with fewer or more lost votings yielded ambivalent results, depending on the French word lists. The Semantic Lexicon of Emotions by Leleu (1987) resulted in no effects. The alternative - a French translation of the BAWL-R - showed a weak effect for valence. Language not producing an effect was surprising, given that we found that German speeches displayed higher valence and lower arousal compared to their French counterparts. Despite this difference, the predictor *language* was not able to predict the membership to parliamentary groups. The authors of one of the studies we intended to replicate, Hirst et al. (2014), encountered a similar issue. In comparison to English transcriptions, they found a lower accuracy for French transcriptions of speeches of the Canadian parliament. It is unclear whether the discrepancies in both studies were due to the different word lists or linguistic and cultural influences. We suspect that this lack of fixed effects may indeed be a result of the different word lists used for our analyses. The values of common words in SLE and BAWL-R show correlations of $r=.89$ for valence (457 common words) and $r=.31$ for arousal (501 common words). This suggests that the SLE measures at least arousal differently than the BAWL-R does. In other studies, it was also found that arousal, in contrast to valence, has a weaker correlation between different instruments and usually has a lower inter-rater correlation (e.g. Kaakinen et al., prep). As mentioned in the chapter on measuring emotions in texts, we did not employ the German translated NRC-VAD from (Mohammad, 2018) as an alternative word list due to the expected loss of predictive power, as indicated by the lower correlations between human ratings and the valence and arousal values of the German translation of the NRC-VAD compared to the original English version. An analysis of our data with the German NRC-VAD confirmed this; the correlations with the values of the BAWL-R and the SLE were indeed very low.

The results of the random effects models indicate that the relationships between the predictors and the outcome are influenced by the sessions themselves. However, we do not know why exactly sessions exert an influence. A plausible explanation could be the topics that were discussed within the individual sessions. Since not every topic is equally emotional, this is likely to be reflected in the respective speeches. In order to examine this, future studies would need to quantify and categorise the

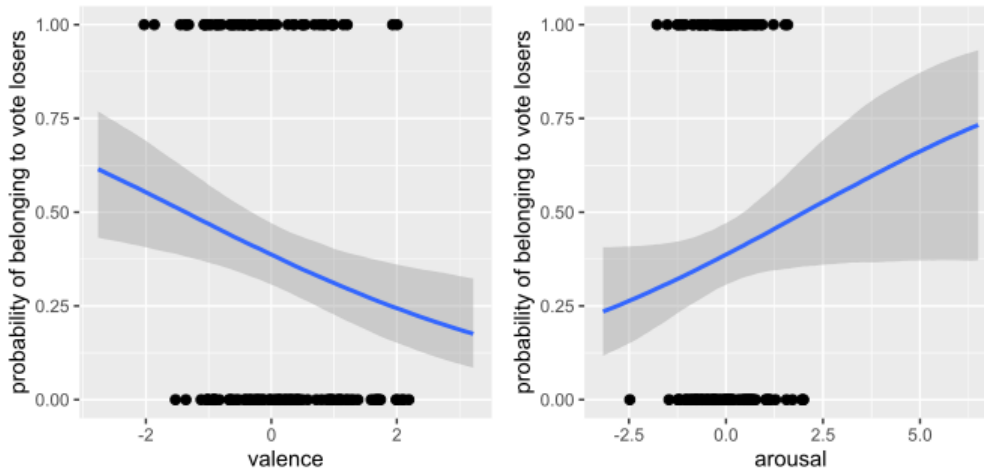


Figure 2: Slopes of valence and arousal.

contents of the sessions, which would also require more sessions and legislatures to be included in the analysis.

Overall, there are multiple possible reasons that could explain the weak effects we found when predicting the affiliation with specific parliamentary groups. Rheault et al. (2016) mentions that different parliaments have their own expressions with specific meanings. Consequently, Salah et al. (2013) proposes that "dedicated political lexicons might need to be built to improve overall accuracy" (p. 128). Furthermore, Rheault et al. (2016) lists other commonly known linguistic features that cannot be captured with a simple text analysis based on word lists. These include sarcasm, irony, and hyperbole. In addition, there are other factors besides valence and arousal that can be used to predict affiliation to parliamentary groups. For instance, reason, logic, and culture were used in another context, namely the effects of speeches in parliament (Freedeen, 2013).

Finally, the strength of the prediction effects also was not equally consistent in the studies that analysed the transcribed speeches of the Canadian parliament (Riabinin, 2009, Hirst et al., 2014). Hirst et al. (2014) reported better results for the 36th government compared to the much less clear results from the 39th government, indicating that inconsistent effects may be expected in this type of study.

6 Conclusions

In conclusion, emotional valence and emotional arousal of parliamentary speeches can be assessed with a lexical approach of emotional text analysis. Depending on the word list used, the valence of parliamentary speeches is able to predict whether parliamentarians belong to groups that lost fewer

or more votings (as an analogy to governing party or opposition), replicating the results of previous studies. In comparison, arousal and language were far less successful. Future studies need to take additional predictors into account, particularly attributes of the parliamentary sessions (e.g. the discussed topics and their affective potency) or non-emotional ones.

Acknowledgments

Many thanks to the three anonymous reviewers for their suggestions for corrections and their valuable comments, most of which we were able to incorporate into the paper.

References

- Abercrombie, G. and Batista-Navarro, R. (2020). Sentiment and position-taking analysis of parliamentary debates: A systematic literature review. *Journal of Computational Social Science*, pages 1–26.
- Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A., and Conrad, M. (2016). Measuring the basic affective tone in poetry using phonological iconicity and subsyllabic salience. *Psychol. Aesthet. Creat. Arts*, 10(2):191–204.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1996). The celex lexical database (cd-rom).
- Barbalet, J. M. (1998). *Emotion, Social theory, and social structure: A macrosociological approach*. Cambridge University Press.
- Barrett, L. F. and Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current directions in psychological science*, 8(1):10–14.

- Benoit, K. and Matsuo, A. (2019). *spacyr: Wrapper to the 'spaCy' 'NLP' Library*. R package version 1.2.
- Borg, A. and Sariyar, M. (2019). *RecordLinkage: Record linkage in R*. R package version 0.4-11.2.
- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.
- Bradley, M. M. and Lang, P. J. (1999). Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida.
- Bühler, K. (1934). Sprachtheorie (language theory). *Stuttgart: G. Fischer*.
- Bürkner, P.-C. (2018). Advanced Bayesian multi-level modeling with the R package brms. *The R Journal*, 10(1):395–411.
- Core Team, R. (2017). R: A language and environment for statistical computing. r foundation for statistical computing. *Vienna, Austria: URL <https://www.R-project.org/>. [Google Scholar]*.
- Cowie, R. and Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech communication*, 40(1-2):5–32.
- Degner, J., Doycheva, C., and Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition*, 15(1):181–189.
- Ekman, P. (1999). Basic emotions. In Dalglish, T. and Power, M. J., editors, *Handbook of Cognition and Emotion*, pages 45–60. John Wiley & Sons.
- Erisen, C. and Villalobos, J. D. (2014). Exploring the invocation of emotion in presidential speeches. *Contemporary Politics*, 20(4):469–488.
- Freeden, M. (2013). Editorial: Emotions, ideology and politics.
- Freud, S. (1891). *Zur Auffassung der Aphasien: Eine kritische Studie*. F. Deuticke.
- Frijda, N. H. (1986). *The emotions*. Cambridge University Press.
- Hirst, G., Riabinin, Y., Graham, J., Boizot-Roche, M., and Morris, C. (2014). Text to ideology or text to party status? In Kaal, B., Maks, I., and van Elfrinkhof, A., editors, *From text to political positions: Text analysis across disciplines*, pages 93–116. John Benjamins.
- Hoggett, P. and Thompson, S. (2012). Introduction. In Hoggett, P. and Thompson, S., editors, *Politics and the emotions: The affective turn in contemporary political studies*, pages 1–19. Continuum Books.
- Holodynski, M. and Friedlmeier, W. (2012). *Emotionale Entwicklung: Funktion, Regulation und soziokultureller Kontext von Emotionen*. Springer.
- Hsu, C.-T., Jacobs, A. M., Citron, F. M., and Conrad, M. (2015). The emotion potential of words and passages in reading harry potter—an fmri study. *Brain and language*, 142:96–114.
- Izard, C. E. (1991). *The psychology of emotions*. Springer Science & Business Media.
- Jacobs, A. M. and Lüdtke, J. (2017). Immersion into narrative and poetic worlds. a neurocognitive poetics perspective. *Narrative absorption*, 27:69–96.
- Jacobs, A. M., Schuster, S., Xue, S., and Lüdtke, J. (2017). What’s in the brain that ink may character. . . .: A quantitative narrative analysis of shakespeare’s 154 sonnets for use in (neuro-) cognitive poetics. *Scientific Study of Literature*, 7(1):4–51.
- Jacobs, A. M., Vö, M. L.-H., Briesemeister, B. B., Conrad, M., Hofmann, M. J., Kuchinke, L., Lüdtke, J., and Braun, M. (2015). 10 years of bawling into affective and aesthetic processes in reading: what are the echoes? *Frontiers in psychology*, 6:714.
- Jhean-Larose, S., Leveau, N., and Denhière, G. (2014). Influence of emotional valence and arousal on the spread of activation in memory. *Cognitive processing*, 15(4):515–522.
- Kaakinen, J. K., Werlen, E., Kammerer, Y., Acartürk, C., Aparicio, X., Baccino, T., Balenghein, U., Bergamin, P., Castells, N., Costa, A., Falé, I., Megalakaki, O., and Ruiz-Fernandez, S. (prep). Idest: International database of emotional short texts. *Behavior Research Methods*.
- Kay, M. (2020). *tidybayes: Tidy data and geoms for Bayesian models*. R package version 2.3.1.
- Koschut, S. (2020). Emotion, discourse, and power in world politics. In Koschut, S., editor, *Emotion, discourse, and power in world politics*, pages 3–27. Taylor & Francis.
- Kuppens, P., Tuerlinckx, F., Russell, J. A., and Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, 139(4):917–940.
- Lara, E. A., Márquez, A. C., and Fuentes-Rodríguez, C. (2016). Emotional argumentation in political discourse. *A Gender-based Approach to Parliamentary Discourse*, 68:129–159.
- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

- Lehne, M., Engel, P., Rohrmeier, M., Menninghaus, W., Jacobs, A. M., and Koelsch, S. (2015). Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLOS One*, 10(5):e0124550.
- Leleu, S. (1987). Un atlas sémantique de concepts d’émotion: Normes et validation [a semantic lexicon of emotion: Norms and validation]. *Unpublished MA Thesis, Psychology, Catholic University of Louvain*.
- Leveau, N., Jhean-Larose, S., Denhière, G., and Nguyen, B.-L. (2012). Validating an interlingual metanorm for emotional analysis of texts. *Behavior Research Methods*, 44(4):1007–1014.
- Lutz, C. and White, G. M. (1986). The anthropology of emotions. *Annual review of anthropology*, 15(1):405–436.
- Matsumoto, D., Yoo, S. H., and Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology*, 94(6):925.
- McCandless, M., Sanford, M., and Firat, A. (2013). *cldr: Language identifier based on CLD library*. R package version 1.1.0.
- Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184.
- Panksepp, J. (1998). *The foundations of human and animal emotions*. Oxford University Press.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of liwc2015. Technical report, University of Texas at Austin, Austin, Texas.
- Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary debates with automated textual analysis. *PLoS one*, 11(12).
- Riabinin, Y. (2009). Computational identification of ideology in text: A study of canadian parliamentary debates. *MSc paper, Department of Computer Science, University of Toronto*.
- Salah, Z., Coenen, F., and Grossi, D. (2013). Extracting debate graphs from parliamentary transcripts: A study directed at uk house of commons debates. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, pages 121–130.
- Scherer, K. R. (1999). Appraisal theory. In Dalglish, T. and Power, M. J., editors, *Handbook of cognition and emotion*, pages 637–663. John Wiley & Sons.
- Scherer, K. R. (2010). The component process model: Architecture for a comprehensive computational model of emergent emotion. *Blueprint for affective computing: A sourcebook*, pages 47–70.
- Sievert, C. (2020). *Interactive web-based data visualization with R, plotly, and shiny*. Chapman and Hall/CRC.
- Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., and Conrad, M. (2017). On the relation between the general affective meaning and the basic sublexical, lexical, and inter-lexical features of poetic texts—a case study using 57 poems of hm enzensberger. *Frontiers in Psychology*, 7:2073.
- Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., and Jacobs, A. M. (2009). The berlin affective word list reloaded (bawl-r). *Behavior Research Methods*, 41(2):534–538.
- Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.
- Werlen, E., Imhof, C., Benites, F., and Bergamin, P. B. (2019). The reader’s feeling and text-based emotions: the relationship between subjective self-reports, lexical ratings, and sentiment analysis. In *SwissText 2019, Winterthur, 18-19 June 2019*. CEUR-WS.org.
- Werlen, M., Moser, I., Imhof, C., and Bergamin, P. (2018). Is reading mirrored in the face? a comparison of linguistic parameters and emotional facial expressions. In *SwissText 2018, Winterthur, 12-13 June 2018*. CEUR-WS.org.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wickham, H. (2017). *tidyverse: Easily install and load the 'Tidyverse'*. R package version 1.2.1.
- Widmann, T. (2021). How emotional are populists really? factors explaining emotional appeals in the communication of political parties. *Political Psychology*, 42(1):163–181.