“Shame on you, honourable colleague!” Detecting Affect in Political Speech

Ivo Bantel
University of Zurich
bantel@ipz.uzh.ch

Matthias Roesti
University of St. Gallen
matthias.roesti@unisg.ch

25 June 2021

Abstract

This project encompasses a new metric for negative emotions expressed in political speech. It aims to detect differences in emotional content in speeches directed at (i) individuals, (ii) parties, or (iii) other types of groups/entities. We apply our method to parliamentary speeches from the UK House of Commons. Our preliminary findings suggest that traditional sentiment classifiers can serve as a first indication of the general tone of a politician's statement, but fail to detect the nuances associated with the rhetorical intricacies of parliamentary debates. In response, we implement our own procedure trained on labelled emotion data and present a pilot analysis focusing on expressions of shame in parliamentary speech leading up to the Brexit withdrawal agreement.

I. Background and research question

Although ideological polarization is less pronounced than it seems (Enders and Armaly, 2019), it is in fact quite substantial and has come into the focus of scholarly debate in recent years (e.g. Fiorina and Abrams, 2008; McCarty, 2019), not least due to the electoral rise of populist and radical right parties (Mudde and Rovira Kaltwasser, 2018; Akkerman et al., 2016). Additionally, a notedly *affective* dimension has been added to political competition and political polarization (Wagner, 2021; Reiljan, 2019), which is distinct from that of ideology (Iyengar et al., 2012). In line with this, partisan alignment, at least in the United States, increasingly is not structured in support of one party but *against* the other (Abramowitz and Webster, 2018, "negative partisanship"). Two aspects are of particular relevance: first, political competition – and indeed, *polarization* – is a notedly emotional in addition to the ideological dimension. Second, little is known about the linkages of the increasingly *emotional* dimension of political polarization and the links between elite- and mass-level.

For this reason, we put forward the question: *which actor characteristics explain variation in the emotionality of parliamentary debates?* To address this issue, we aim to trace the prevalence of several negative emotions (e.g. anger, shame or disgust) in parliamentary debates, as well as characteristics of the parliamentarians uttering the respective statements and the target of these statements. Exploring this question has relevant implications in two domains: first, policy-making depends on political elites compromising. While ideological disparities are necessary and not harmful per se, increasingly negative emotions between political elites makes cooperation less likely. Second, elite interactions provide a norm for societal interactions between political opponents. If the former deteriorate (e.g. becoming dominated by disgust), this can have severe consequences for interactions between citizens of competing parties. In short, emotion in parliamentary debates can serve as an indicator of elites' affective polarization (i.e. dislike of political out-groups), which in turn has important implications.

II. Data

Using *parliamentary speeches* to measure polarization in elite affect has one important advantage (and a crucial challenge which we will discuss later): desirably, the records range back in time and thus provide us with an uninterrupted record that is superior to sparse survey data only capturing mass sentiment.

Our analysis draws on two main sources of data: first, for training, we draw on the

ISEARs dataset (International Survey on Emotion Antecedents and Reactions), which contains descriptions of situations in which individuals felt seven particular emotions (anger, disgust, fear, guilt, joy, sadness, and shame; Scherer and Wallbott, 1994). The data set contains ca. 7,600 cases, equally distributed across the aforementioned emotions.

Second, emotions in political speech are measured in our main corpus of all speeches from the UK House of Commons from the *ParlSpeech* data set (Rauh and Schwalbach, 2020). Our current analysis focuses on speeches from September 2018 to December 2019, but the data also cover previous decades going back as early as the 1980's, and future iterations of the project will extend to this time period. This corpus is rather detailed, containing the transcribed speeches as well as information about the speaker and the agenda item, allowing a clear identification of utterances' origin and context from meta data.

III. Method

Methodologically, our project proceeds in two steps: first, we establish a baseline reference for the affective dimension in political speech using sentiment analysis. Second, we use more elaborate Natural Language Processing (NLP) to detect emotions in parliamentary speeches.

Sentiment analysis is a widely used way to extract whether speech has a positive, neutral, or negative tone (e.g. Liu, 2015) and has been applied to political speech as well (Proksch et al., 2019). As a consequences, pre-trained (general purpose) models for sentiment detection exist, which we draw on: we implement both a dictionary-based sentiment classifier ("Vader" dictionary sentiment classifier from python's `nltk` library), a more sophisticated modelling-based classifier ("TextBlob" from python's `textblob` library), and finally a neural net classifier ("flair", Akbik et al., 2019). The three classifiers' results were manually validated on a small sample, yielding mediocre results (dictionary-based), acceptable (modelling-based), and good results (neural net-based), respectively. As a consequence, the sentiment classifications produced by the "flair" classifier form the basis for the (preliminary) results of the sentiment analysis.

The analysis of sentiments yields a score indicating how negative or positive (or neutral) the sentence is. Additionally, the "TextBlob" classifier provides an indicator of "subjectivity" indicating how objective vs. subjective a sentence is.[1] However, existing classifiers cannot unpack the affective structure of negative sentiment.[2] Thus, we detail our

[1]This metric appeared of acceptable quality upon the validation described previously.

[2]One may be tempted to employ classifiers trained to detect hate speech in this context, but our intuition

suggested approach to remedy the shortcomings of existing classifiers in our dimension of interest – negative emotions.

In the second step, we thus detect emotions in the previously described corpus drawing on emotion detection from text (Seal et al., 2020; Abercrombie and Batista-Navarro, 2020), applied to a subset of sentences. For this, we identified those sentences that were more subjective (subjectivity score > 0.3, excluding ca. 40% of the $635,000$ sentences), and performed emotion recognition on them. For this, we draw on the previously described *ISEARs* data set to train a bi-directional neural net (see Table 1 in the Appendix for details on the structure) to predict the emotions in the labelled data set. Naturally, the main challenge associated with deriving emotions from text records is reliability. The trained model performs relatively well on unseen test data (accuracy $> .8$); however, the data genre differs, and since parliamentary speech is unlabelled, we currently have no benchmark for quantitatively validating performance. However, the preliminary descriptive results give us some confidence that the classifier is picking up true tendencies in sentiment, although we will need to work on robustness (see section "Outlook" for more details on this).

IV. Preliminary results

A. Off the shelf sentiment analyzers can help, but not excite

As a baseline reference, we look how the average sentiment of political speeches towards the Prime Minister varies across party lines. We include the last 9 months of Theresa May's tenure, as well as the first 6 months of Boris Johnson, who took office on 24 July 2019. We classify sentiment using a pretrained neural net provided by the Python library *flair* (Akbik et al., 2019).

As we can see in Figures 1 and 2, statements from Conservative MPs referring to Boris Johnson tend to be more positive than those of their Labour colleagues, with a marked spike right around the time when internal elections for the succession of Theresa May were held (the black vertical line marks May's announcement of resignation, the red vertical line is when she left office). Turning to Theresa May on the other hand, we can see that the Conservative sentiment towards her tends to decrease up until she announces her resignation, after which she recovers somewhat. The graph stops in the third quarter of 2019 because there were virtually no mentions of her (by name) anymore.

is that they will perform poorly since politicians are well versed in communicating negative sentiment without directly resorting to inflammatory language.

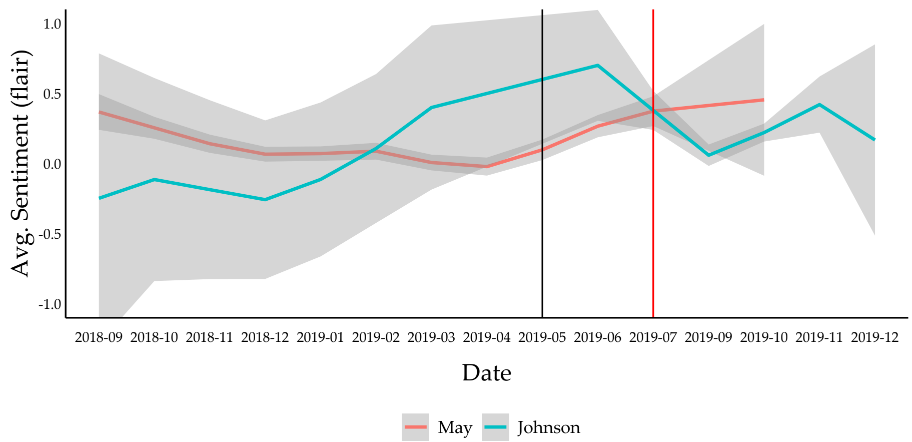

Figure 1: *Conservative sentiment towards Boris Johnson and Theresa May*

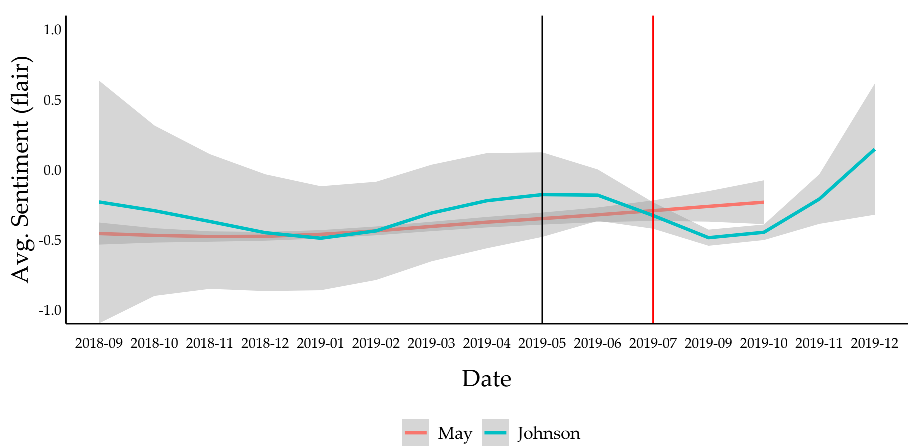

Figure 2: *Labour sentiment towards Boris Johnson and Theresa May*

B. Our emotion classifier – a shameful affair?

As a first preview of how our own model performs, we plot the extent to which both major parties' speeches reflected the emotion of *shame*. For both parties – but especially Labour – we see a rise in *shame* towards the end of 2019, when the uncertainty surrounding the Brexit deal were at their peak (Parliament approved the Brexit deal on 20 December 2019).

However, at this stage we are mindful not to overinterpret the results from our own model, given that we have not yet spent enough time validating/assessing the model predictions.

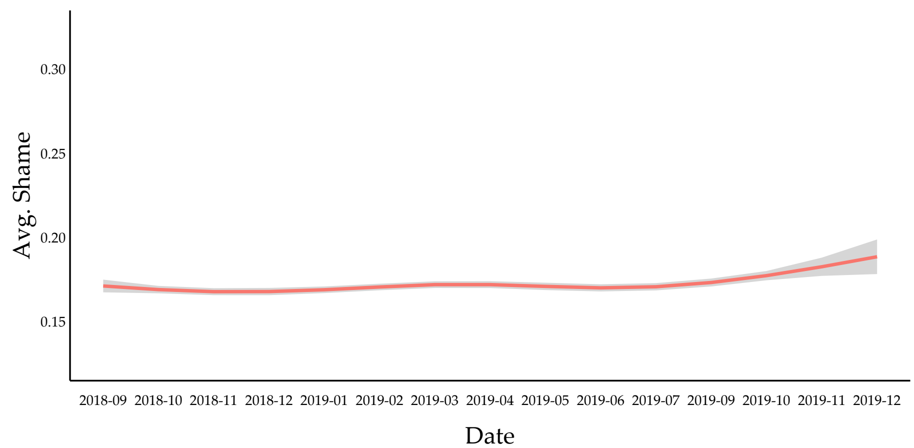

Figure 3: *Conservative extent of 'shame' across time*

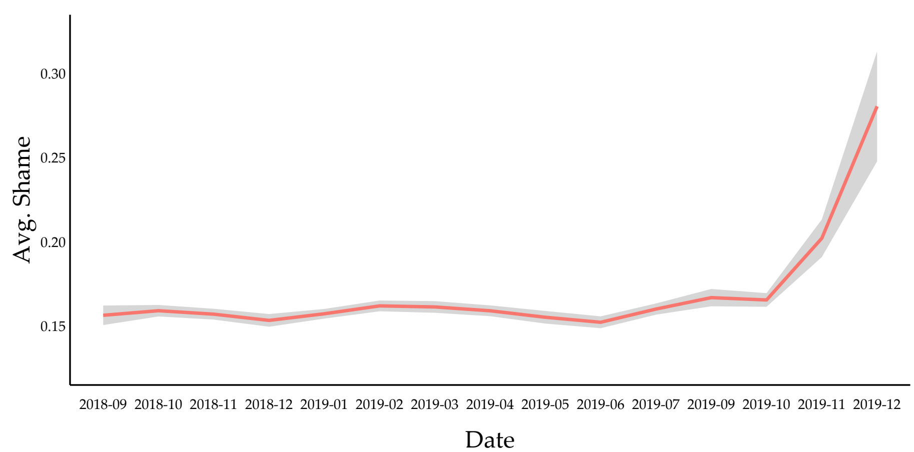

Figure 4: *Labour extent of 'shame' across time*

V. Conclusions & Outlook

With this project, we have shown a rudimentary approach to capture emotions in parliamentary speech. The emotion detector's results must be interpreted cautiously, however, and further improvement is needed (see below). Nevertheless, we believe that this is an important proof of concept, and that further research is both necessary and worthwhile.

We have shown that (changes in) emotion expressed in political speech mirror political events to some degree; in combination with the meta data and results from the *target* of the emotions, we are able to show not only who evokes more negative sentiment but also *which negative emotions* are directed at whom. Findings in this regard will have important implications for the study of polarization, in particular elite-level affective polarization.

Naturally, much work remains to be done. For this project to advance in the desired direction, we see two main avenues that need to be further explored: (i) methodology, and (ii) application.

On the methodological end, more fine tuned classifiers (along with more extensive preprocessing of text) will be necessary to bring performance to a level at which meaningful analysis can be attempted. It is likely that parliamentary speech structurally differs from the typical text corpora (e.g. movie an restaurant reviews) on which sentiment classifiers are trained. In particular, two remedies are conceivable: first, improving the model by including pre-trained embeddings (e.g. using the "BERT" language represntation model; Devlin et al., 2019). Secondly, by improving the data: in the short time available, we were unable to manually annotate sentences from the corpus consisting of the goal genre of parliamentary speech. This will be a crucial next step to improve the performance of the model. To make the most out of the annotations, an active learning approach (e.g. Schröder and Niekler, 2020) will help us to efficiently use the smallest number of annotations while continuously evaluating the performance and thus optimally bridging the gap between genre applications.

With regards to application, further iterations on correctly identifying the targets and the context of parliamentary speeches are on our agenda. The metadata for the speeches are very detailed as far as the topic of debate are concerned, but there are still thousands of categories that need to be discretized to a more tractable number of dimensions (e.g. Brexit, minority/poverty issues, immigration, finance/taxes, health, environment). Having these dimensions, combined with a better labelling of groups (e.g. more targeted minorities, social class related entities) will allow us to explore heterogeneity of emotive speech patterns across these dimensions.

References

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

Abramowitz, A. I. and S. W. Webster (2018): "Negative Partisanship: Why Americans Dislike Parties But Behave Like Rabid Partisans," *Political Psychology*, 39, 119–135.

Akbik, A., T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf (2019): "FLAIR: An easy-to-use framework for state-of-the-art NLP," in *NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, 54–59.

Akkerman, T., S. L. d. Lange, and M. Rooduijn, eds. (2016): *Radical right-wing populist parties in Western Europe: Into the mainstream?*, Extremism and democracy, London and New York, NY: Routledge.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *http://arxiv.org/pdf/1810.04805v2*.

Enders, A. M. and M. T. Armaly (2019): "The Differential Effects of Actual and Perceived Polarization," *Political Behavior*, 41, 815–839.

Fiorina, M. P. and S. J. Abrams (2008): "Political Polarization in the American Public," *Annual Review of Political Science*, 11, 563–588.

Iyengar, S., G. Sood, and Y. Lelkes (2012): "Affect, Not Ideology," *Public Opinion Quarterly*, 76, 405–431.

Liu, B. (2015): *Sentiment analysis: Mining opinions, sentiments, and emotions*, New York NY: Cambridge University Press.

McCarty, N. (2019): *Polarization: What everyone needs to know*, New York, NY: Oxford University Press.

Mudde, C. and C. Rovira Kaltwasser (2018): "Studying Populism in Comparative Perspective: Reflections on the Contemporary and Future Research Agenda," *Comparative Political Studies*, 51, 1667–1693.

PROKSCH, S.-O., W. LOWE, J. WÄCKERLE, AND S. SOROKA (2019): "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches," *Legislative Studies Quarterly*, 44, 97–131.

RAUH, C. AND J. SCHWALBACH (2020): "The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies," *Harvard Dataverse, https://doi.org/10.7910/DVN/L4OAKN*.

REILJAN, A. (2019): "'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems," *European Journal of Political Research*, 59, 376–396.

SCHERER, K. R. AND H. G. WALLBOTT (1994): "Evidence for universality and cultural variation of differential emotion response patterning: Correction," *Journal of personality and social psychology*, 67, 55.

SCHRÖDER, C. AND A. NIEKLER (2020): "A Survey of Active Learning for Text Classification using Deep Neural Networks," *http://arxiv.org/pdf/2008.07267v1*.

SEAL, D., U. K. ROY, AND R. BASAK (2020): "Sentence-Level Emotion Detection from Text Based on Semantic Rules," in *Information and Communication Technology for Sustainable Development*, ed. by M. Tuba, S. Akashe, and A. Joshi, Singapore: Springer Singapore, vol. 933 of *Advances in Intelligent Systems and Computing*, 423–430.

WAGNER, M. (2021): "Affective polarization in multiparty systems." *Electoral Studies*, 69, 1–13.

VI. Appendix

The neural net is constructed as follows:

Layer (type)	**Output Shape**	**Param #**	**Connected to**
Input Layer	[(None, 100)]	0	
Embedding	(None, 100, 500)	4626000	Input Layer
SpatialDropout	(None, 100, 500)	0	Embedding
Bidirectional	(None, 100, 256)	644096	SpatialDropout
Conv1D	(None, 98, 64)	49216	Bidirectional
Global Avg. Pooling 1D	(None, 64)	0	Conv1D
Global Max Pooling 1D	(None, 64)	0	Conv1D
Concatenate	(None, 128)	0	Global Avg. P& Max Pooling 1D
Dense	(None, 7)	903	Concatenate

Total params: 5,320,215

Trainable params: 5,320,215

Non-trainable params: 0

Table 1: *Architecture of neural net.*