

The Role of Negation in Political Text: British and German Electoral Campaigns

Tom Paskhalis *London School of Economics*
Christian Müller *London School of Economics*

May 12, 2017

Abstract

Most quantitative text analysis approaches rely on 'bag-of-words' model, which ignores word order. This might raise a number of issues if the analysed texts extensively use negation at the sentence level. In this work we investigate the implications of this simplifying assumption for the analysis of different types of political texts. Drawing on natural language processing tools, we consider party manifestos and electoral debates in Britain and Germany. Using grammatical sentence as the unit of analysis, we show that opposition parties are more prone to frame their campaigns in negative terms than governing parties. Among opposition we also observe the variation between centrist and more radical and fringe parties. These findings advance our understanding of negative campaigning in European party systems. We further compare our results with traditional, dictionary-based approach. Overall, sentence-level analysis based on automatic, machine-learnt segmentation rules allows both replicable and transparent research design, together with conceptually easier and linguistically sound interpretation of the results.

1 Introduction

Text-as-data approach has revolutionised our understanding of political documents. By allowing to process large quantities of data at very high speed and low cost, scientists have been able to gain insights into policy movements of political parties over time (Proksch and Slapin, 2008), study treatment effects in open-ended surveys (Roberts et al., 2014) and analyse mechanisms behind online censorship in China (King et al., 2013). In their core these advancements were all based on the analysis of texts transformed into quantitative information. Or, to put it crudely, into matrices with word counts. These, so-called 'document term-frequency' matrices typically have documents as rows and words as columns with the number of times a given word occurs in a given document on in

the cells. The principal assumption underlying this transformation is that the order in which words appear in the document does not matter in many cases and can be safely discarded as a source of information. Although even a casual observer can say that this assumption is clearly wrong, in that there is a considerable difference between ‘renewable energy subsidies are a no go’ and ‘no, go for renewable energy subsidies’, this approach has performed remarkably well in many practical applications. There are several possible reasons for why this could be the case. First, this difference is task specific. Using the toy example above we can learn that the phrase is about renewable energy and environmental policies, if we want to be more general. In other words, if we are only interested in the subject or topic of this phrase, than it does not really matter whether it is even ‘energy, go for no renewable subsidies’, both a human coder and a computer program can correctly infer what the text is about. As the whole body of literature on saliency theory (Budge et al., 1987; Lowe, 2013) has shown, issue saliency is what we often get and what we, perhaps, less often are interested in. Second, with large volumes of text it is plausible that subtleties like an individual speaker’s position on renewable energy work both ways and cancel each other out, producing an unbiased estimate on aggregates. And, third, it is also possible that political actors do not use wordplays and subtle meanings, keeping the language simple and straightforward. The emerging literature on the complexity of political speeches over time Spirling (2015) suggests that with the transition to addresses delivered orally, rather than in writing, as well as the expansion of the electoral base from educated elite to mass voters, resulted in less sophisticated texts, shorter and easier to understand. Although this does not necessarily eliminate the possibility for an occasional pun, this figure of speech is too uncommon for today’s politicians to make a difference for quantitative text analysis.

As the proponents of the ‘bag of words’ approach would argue, despite the model being clearly wrong, its usefulness is demonstrated by a long list of intuitively meaningful results. However, restricting ourselves to a single quantitative perspective of the document might not be justified for a diverse set of tasks that political scientists undertake.

The purpose of this paper is threefold. First, we would like to zoom in from a broad question about the good performance of the ‘bag-of-words’ model, given its completely unrealistic assumptions, to one specific hypothesis listed above. Namely, that linguistic instruments used by politicians are just a restricted subset of those offered by language. In other words, certain features that are often encountered in routine conversations simply never or almost never occur in political texts. The one that can potentially have most significant implications for word-based language models is negation. When doing quantitative analysis of text we do not want to confuse negative reference to immigration with a neutral statement that contains the same words. Thus, negation in political texts is an important phenomenon in itself. With the literature only giving it only a cursory treatment, it is something that we should know more about when adopting text-as-data approach. And, third, we are introducing new tools borrowed from natural language processing that can formalise our approach to such aspects of the analysis as unitisation and tagging. By relying on syntactic parsing software, we are removing potential sources of unreliability that can arise when texts are prepared for the analysis.

2 Negation in Political Texts

It is worth starting the discussion of negation in political text by drawing a line between the use of the word ‘negative’ in some political science studies and the one we adopt here. While many, following the landmark study by Ansolabehere and Iyengar (1995) use negative in its broad sense, would refer to negative campaigning as the one blighted by vitriolic attacks on the opponent, we take a purely linguistic approach, where negation is simply a statement of type ‘X is not Y’. Although in practice the two can overlap, we will try avoiding any normative implications and focus on describing the cases where it can have change our understanding of political mechanisms. One of the widely studied questions in comparative politics is party competition. The largest research undertaking to conceptualise it both theoretically and methodologically is the Manifesto Project, also

known as MARPOR (Manifesto Research on Political Representation), which analysed party competition as expressed in pre-election manifestos. Containing over 4'000 manually coded manifestos at the time of writing, this is, perhaps, the most frequently used source for party positions and their movements over time. The theoretical basis of this work is saliency theory (Budge et al., 1987), which states that parties do not engage in direct confrontation and argument, but rather put differential emphasis on policy areas. Evidently, emphasis, as opposed to rejection, is, by far, more stable across different word orders and, thus, is more amenable to word frequency analysis (Laver and Garry, 2000; Laver et al., 2003). The absence of confrontation in party competition has been recently questioned by Dolezal et al. (2016). Analysing Austrian manifestos, the authors argue that negative references are more common than is assumed by the saliency theory.

While adopting a more linguistic operationalisation of negation in this article, we can still test the hypotheses from the literature on negative campaigning. Typically, we would expect opposition parties to be more likely to use negative language in their campaigns (Lau and Rovner, 2009). The potential gains they can get from criticising the ruling party could potentially outweigh those from stressing their own competence. Another expectation concerns extreme parties that are assumed to be more critical and, thus, more negative. This stems from the extremity of their ideology that puts them further away from rest of the political spectrum (Elmelund-Præstekær, 2010; Dolezal et al., 2016).

3 Unitisation

Unitisation, or segmentation of the document into basic units for analysis, forms the first step in text analysis. Unsurprisingly, the selection of the appropriate unit of analysis has been a prominent topic of research work throughout the years of MARPOR existence. In the original empirical study that paved the way for the project, Robertson (1976) used natural sentences as the units of policy expression. Later on, Budge et al. (1987) have

argued that a single sentence, especially a long one, can contain references to multiple policy propositions. This, in authors' view, necessitated splitting up some sentences into shorted word sequences, that were jointly termed 'quasi-sentences'. However, the actual segmentation was left to the discretion of human coders without any reported checks of agreement and without possibility of calculating those ex post, due to having only single coder to complete the task. Däubler et al. (2012) have criticised this approach of endogenously defining units of analysis as highly subjective and prone to producing unreliable results. They also showed that a natural sentence containing multiple quasi-sentence is relatively uncommon and, thus, splitting them up does not have any added benefits to exogenously defined natural sentences. The most recent approach has been offered by Dolezal et al. (2016), who made an attempt to give a solid linguistic justification to the choice of the basic unit. Building upon Chomsky's phrase structure grammar (Chomsky, 1957), they derive 'kernel sentences' or the shortest possible, yet grammatically correct, sentences. Essentially, in linguistic terms this method derives noun phrase (NP) and the corresponding verb phrase (VP), together with the other words that in the parse tree are attached to the same NP or VP. Despite this approach being linguistically well-defined, there are still inherent methodological problems in using human coders to manually parse the sentences. We believe that this task can be well handled by the modern natural language processing (NLP) tools, without jeopardising reliability and replicability of the study. Modern NLP libraries have the same linguistic foundation, can process large quantities of text in many languages of interest and, thus, are well suited for the unitisation of sentences in text analysis.

3.1 Data Sources

The texts analysed in this study include British and German electoral manifestos, as well as transcripts of televised debates, where those are available. To make comparisons across different competition platforms, we focus specifically on the years since 2000, as the leadership debates have been introduced only recently in these countries. Specifically,

Britain saw first live televised debate between part leaders only prior to 2010 General Election, although the idea has been in the air since the arrival of television half a century ago. Germany, on the contrary, first introduced TV debates in 1972. However, due to then chancellor Helmut Kohl refusing to participate in them, there is a gap, which spans the period between 1990 and 2002 (Anstead, 2015).

4 Dictionary Analysis

We focus at this stage on party manifestos, as they are easier to obtain and provide an easy way to test the key assumptions, underlying future work. The most straightforward way of sanity check for the presence of negation in party manifestos is applying dictionary and calculating the relative frequency of negative words within each of them. Dictionary approach in its modern form dates back to the seminal work by Stone et al. (1966). Here we apply a very primitive dictionary that consists of about dozen words, indicating negation¹ to the party manifestos in two countries starting from the year 2000.

party	2001	2005	2010	2015
Conservative	0.56	0.45	0.41	0.29
Green	NA	0.33	0.43	0.34
Labour	0.33	0.45	0.38	0.42
Liberal Democrat	0.39	0.51	0.20	0.28
Plaid Cymru	NA	NA	0.42	0.34
SNP	0.39	0.27	0.35	0.19
UKIP	0.63	0.69	0.36	0.64

Table 1: Percentages of negations in UK party manifestos from 2001 to 2015 (NA are missing data points)

The results for Britain are presented in Table 1 and in Table 2 for Germany. Both hypotheses about the fringe and opposition parties cannot be rejected from this cursory analysis. Most far right parties (UKIP in Britain and AfD in Germany) have the highest percentage of negations in their manifestos, reaching more than 1% of the total word

¹The precise number is 10 for English (*no, not, n't, neither, never, nobody, none, nor, nothing, nowhere*) and 14 for German (*kein, keiner, nein, nicht, nichts, nie, niemals, niemand, nirgends, nirgendwo, nirgenwoher, nirgendwohin, keinesfalls, keineswegs*)

party	2002	2005	2009	2013
AfD	–	–	–	1.10
CDU	0.56	0.54	0.33	0.31
FDP	0.57	0.66	0.60	0.50
Green	0.45	0.77	0.70	0.61
Linke	–	–	0.55	0.62
Pirate	–	–	NA	0.56
SPD	0.33	0.60	0.53	0.55

Table 2: Percentages of negations in German party manifestos from 2002 to 2013 (NA are missing data points, – are electoral years in which parties did not run)

count in the case of AfD. Also, less radical, but still further from the centre, Greens in both countries, Linke and Pirate Party in Germany use more negative language. Perhaps, the most striking observation is the switch from more to fewer negations when a party enters the government. This pattern holds for Conservatives (in opposition from 2001 to 2010) and CDU (in opposition from 1994 to 2005), as well as SPD, except for 2005 electoral campaign, when it was still the incumbent together with the Greens. Labour shows a more complex pattern, with a slight increase towards the last General Election, but with considerable variation even when it was in power.

Overall, due to language differences the raw percentages cannot be compared between countries, but within-country patterns of variation seem to reflect some mechanism of party competition, which deserves further investigation.

5 Syntactic Parsing

After showing that negation is indeed an existing phenomenon in political campaigns, the next step is to learn more about what is being negated. At this step we will rely upon NLP library to extract syntactic structure of the sentences and look at parts of the parse tree that follow negation. The critical part here is that NLP tools allow us to retain word order during the analysis. After extracting the negated parts, we will be able to run topic models to look at the topics that parties prefer to talk about in negative.

References

- Ansolabehere, S. and S. Iyengar (1995). *Going Negative: How Political Advertisements Shrink and Polarize the Electorate*. New York: The Free Press.
- Anstead, N. (2015). Media Policy Brief 13: Televised Debates in Parliamentary Democracies. Technical report, LSE Media Policy Project.
- Budge, I., D. Robertson, and D. Hearl (1987). *Ideology, Strategy, and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*. Cambridge: Cambridge University Press.
- Chomsky, N. (1957). *Syntactic Structure*. The Hague: Mouton & Co.
- Däubler, T., K. Benoit, S. Mikhaylov, and M. Laver (2012). Natural Sentences as Valid Units for Coded Political Texts. *British Journal of Political Science* 42(4), 937–951.
- Dolezal, M., L. Ennser-Jedenastik, W. C. Müller, K. Praprotnik, and A. K. Winkler (2016). Beyond salience and position taking: How political parties communicate through their manifestos. *Party Politics* (June), 1–13.
- Dolezal, M., L. Ennser-Jedenastik, W. C. Müller, and A. K. Winkler (2016). Analyzing Manifestos in their Electoral Context A New Approach Applied to Austria, 20022008. *Political Science Research and Methods* 4(3), 641–650.
- Elmelund-Præstekær, C. (2010). Beyond American negativity: toward a general understanding of the determinants of negative campaigning. *European Political Science Review* 2(01), 137.
- King, G., J. Pan, and M. Roberts (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review* 107(917), 326–343.
- Lau, R. R. and I. B. Rovner (2009). Negative Campaigning. *Annual Review of Political Science* 12, 285–306.
- Laver, M., K. Benoit, and J. Garry (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review* 97(2), 311–331.
- Laver, M. and J. Garry (2000). Estimating Policy Positions from Political Texts. *American Journal of Political Science* 44(3), 619–634.
- Lowe, W. (2013). There’s (basically) only one way to do. Working Paper.
- Proksch, S.-O. and J. B. Slapin (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science* 52(3), 705–722.
- Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 58(4), 1064–1082.

Robertson, D. (1976). *A Theory of Party Competition*. New York: Wiley.

Spirling, A. (2015). Democratization and Linguistic Complexity: The Effect of Franchise Extension on Parliamentary Discourse, 1832-1915. *The Journal of Politics* 78(1), 235–248.

Stone, P. J., D. C. Dunphy, and M. S. Smith (1966). *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge: The MIT Press.