

The We-Narrative in Public Communications of German Politics - A NLP Analysis

Term Paper^{*}

Johannes Renz[†]

October 3, 2023

Abstract

This paper conducts an empirical analysis into use of we-narratives in groups. Specifically, tweets from German elected party members are analysed. The aim is to check if in coalition politicians use we-narratives more. Identification of we-narratives is done with pretrained language models. The resulting finding is that group members do use we-narratives about 5% more.

Keywords: Narratives, Natural Language Processing, Communication, Collective Action.

^{*}This paper is written as a term paper in the course Collective Action, taught by Prof. Roy.

[†]6th semester, BA Philosophy & Economics, Universität Bayreuth. Student No.: 1786400. Address: Rupprechtstraße 11, 95444 Bayreuth. Tel.: +177 4908085. Email: johannes.renz@uni-bayreuth.de.

Contents

1	Introduction	1
2	Literature	1
3	Empirical Analysis	2
3.1	Data	2
3.2	Method	3
3.3	Regression Analysis	5
4	Discussion	7
5	Conclusion	7
	Bibliography	8

1 Introduction

We-narratives are central to groups from the perspective of collective action. However, even with some work from the philosophical/theoretical side, there seems to be a lack of empirical studies. Novel natural language processing techniques allow identification of narratives in general and we-narratives in specific. The aim of this paper is to identify if individuals and parties in politics use we-narratives more often when they are in a group. Specifically, I will use German coalitions' public communications on X (formerly Twitter).

I will proceed by briefly reviewing some literature on we-narratives in Section 2. Following this, I will present my empirical analysis in Section 3 starting with discussing my data, then my identification method of we-narratives and finally a regression analysis of the data. Following this, I will discuss my method in Section 4 and conclude in Section 5.

2 Literature

While there is a large volume of academic work discussing political narrative, I will discuss narratives from the angle of collective action. This is because political narrative are discussed more as a form of storytelling and related less to group dynamics in politics (see for example Shenhav (2006)). Within the larger subject of collective action, narratives are seen from a more philosophical angle. They have complicated philosophical problems to address, for example, the infinite regress that Schmid (2014) poses and the process of forming a collective on the level of mental states. I will be mainly drawing on work by Gallagher and Tollefsen (2017). The account places narratives as central for building group identity, joint agency and stability. They argue that for groups we-narratives, which are defined generally as just first person plural narratives, are especially central. As Carr puts it, "a community exists where a narrative account exists of a *we*" (Carr (1986), p. 130). We-narratives help groups to create a sense of stability, identity and goals.

I want to apply this concept to a practical group building situation, and evaluate if this can be observed empirically. However data on these classical social situations does to my knowledge not exist, which is why I chose a special setting with more recorded statements: Political coalitions. Furthermore it also seems this setting is interesting because the results can be interpreted in answer to political questions.

Coming back to the account of Gallagher and Tollefsen (2017), my application of coalitions centrally involves what they call D-Intentions.¹ Within this kind of long term collective action, Gallagher and Tollefsen find "no problem of speaking about a collectively formed intention and a reflective attribution of a we" (Gallagher and Tollefsen (2017), p. 213). The reflective attribution of the we mostly occurs in communicating (joint) actions. This attribution of the we is what my empirical analysis aims to identify.

3 Empirical Analysis

With recent great advances in Natural Language Processing (NLP), quantitative analysis of text data has become more common, especially in economics. Narratives have also begun to be a target of analysis (see Ash and Hansen (2023) for a general overview). Specially written tools, notably the *relatio* package by Ash, Gauthier, and Widmer (2022), have proven to identify narratives well. First applications in the scientific discipline have been with promising results (Andre, Haaland, Roth, and Wohlfart (2021)). Sadly, these packages are mainly available in English.² My empirical exercise uses a German dataset and can therefore not take advantage of these novel algorithms. Instead, I develop a small algorithm myself, which does not try to extract overall narratives, rather just the information if a text contains a we-narrative or not. I use this algorithm to look at tweets by politicians. The aim is to identify if parties that are in the process of building a coalition utilise we-narratives to communicate a sense of collective identity and joint action. In this section, I will proceed by describing the data in Section 3.1, my method in Section 3.2 and characterise my results in Section 3.3.

3.1 Data

The data used consists of tweets by members of the Bundestag from 2012 until 2019. Specifically, all members of the 2013 Bundestag are recorded, with the addition of 2017 Bundestag AfD politicians.³ For them, all posted tweets within this period are in the dataset. Sadly, the FDP is not included in this dataset. The tweets totally amount to about 1.6 million tweets. Since one focus of this paper is to examine coalition periods, it is noteworthy that these are

¹They define D-intention as follows: "D-intentions are best understood in terms of a narrative timescale extending over variable periods between time of decision and time of action." (Gallagher and Tollefsen (2017), p. 212)

²Note that *relatio* is also available in french.

³Note that the tweets for the MdBs (*Mitglied des Bundestags*, Member of German Parliament) are also recorded before they enter the Bundestag.

recorded for the 2013 Bundestag and the 2017 Bundestag.

Optimally, I would analyse group-internal communications and look for we-narratives in this context. This is because identity building mainly occurs in this format (Gallagher and Tollefsen (2017)). However, since these kinds of communications are not available as data, I resort to using the public communications I previously described. Still, for the context of political coalitions, I argue that when the coalition builds a group identity, this would also be conveyed in their communication to the public.

For the method of analysis the tweets are split into sentences, resulting in nearly 3 million sentences. These are then preprocessed further by removing @ mentions and #, since they are mostly not integrated in sentence syntax. Links and Retweets are also removed.

3.2 Method

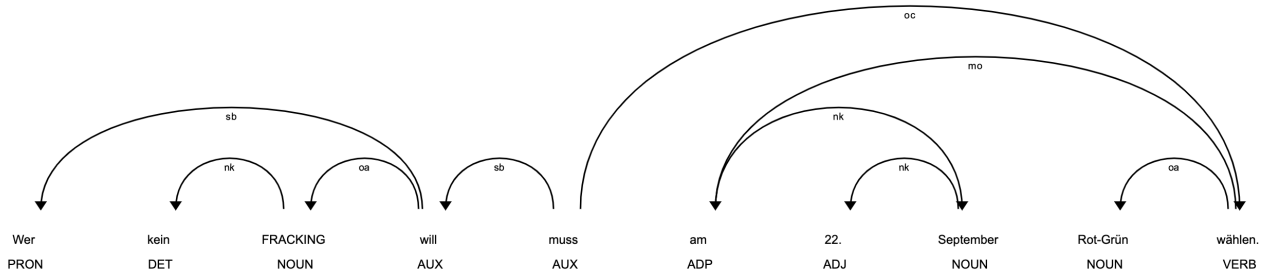
To identify We-narratives, I employ two different methods, which yield a very similar result. The main method uses verbs to identify we-narratives. To evaluate robustness, I also check for occurrence of "wir" which is the word for we in German. My main method utilises the spaCy library by Honnibal, Montani, Van Landeghem, and Boyd (2020), which provides industrial-strength Natural Language Processing models in python. The models have three functions/pipelines I will utilise: Firstly, the model can identify what word corresponds to which function in a sentence, which is called POS-tagging (Part of Speech tagging). This can identify which word in a sentence is the verb. The second function is dependency parsing, which can detect relations between different parts of the sentence. These two functions are illustrated in Figure 1. The arrows illustrate dependency relationships, while the tags below the words describe the functions within the sentence.

The third function I take advantage of is the morphologizer. This can identify what grammatical form the verb is in. For a given verb, the model can output the person, number and other information like tense, aspect and mood. I apply these three features in the following algorithm:

After preprocessing, which is described in Section 3.1, all sentences in the dataset are passed to the spaCy model.⁴ The model then labels the sentence's words with POS-tags. Following this, all words that are labelled as a verb (POS-tag "VERB") are added to a list containing all verbs.

⁴Specifically, I use the model `de_core_news_lg`, which is a large model trained on a labelled German news corpus dataset.

Figure 1: POS-tagging and dependency parsing using spaCy with an exemplary tweet from the dataset.



If no verb is found in the sentence, then words labelled auxiliary verbs (POS-tag "AUX") are added to the same list. After creating the list with all verbs and auxiliary verbs, the morphologizer is run to detect number and person of the verb. If none is found, the dependency tree is searched for any other auxiliary verbs.⁵ If a word is found that is the auxiliary verb of the main verb, then the morphologizer is again asked to detect number and person. Using this, we then output a list of all verbs and a list of the person and number these verbs (sometimes identified via their auxiliary verbs).

After recovering this information, I reformat the data and output to a data-frame, which contains all the information collected in this procedure. The sentences are then labelled as we-narratives if they contain at least one verb in first person plural form. To avoid any issues due to different numbers of sentences in tweets, the sentences are aggregated on the tweet level. This means that tweets with multiple sentences just get labelled as we-narratives if at least one sentence within the tweet contains a we-narrative.

For robustness, I compare results with a much simpler method: Labelling a sentence as a we-narrative if it contains a mention of the word "wir". For this, I get qualitatively very similar results, with the two methods correlating to about 95% in the sentence level dataset.

For the results analysis, the tweets are then aggregated by party and date, so that each day has one value per party. This value corresponds to we-narrative tweets per total tweets and is my main measure.

⁵This is important for the following case: The sentence contains a verb, but this verb cannot be identified with both number and person, since it uses an auxiliary verb. The morphologizer can then identify number and person of the verb via the auxiliary verb.

3.3 Regression Analysis

To quantitatively determine if parties that are in a coalition express themselves more in we-narratives, I utilise a panel regression, specifically a two way fixed effects model. This is because the method can take advantage of the whole panel data (2013-2019) that is created as described in Section 3.

Briefly spoken, the fixed effect panel regression calculates the effect of an event on an outcome variable across time and individuals. One major advantage of the method is that time-invariant characteristics of individuals and time-specific shocks can be controlled for. This means that the trend over time and a party effect can be excluded from the resulting calculations. Therefore, the estimator should calculate the effect of a party going into coalition on the use of we-narratives. (For explanation of the method, see Wooldridge (2012))

For the method, I need to define when a party is in a coalition. At first this seems obvious, as I can just include the coalition parties that are SPD, CDU and CSU over nearly the whole period. However, to avoid changing narratives during election, I will leave out six months before the election. Furthermore, I will include the Jamaica negotiation period, where the Grüne, CDU, CSU and FDP negotiated a coalition. During this time, the SPD also excluded a coalition with the CDU and CSU (for further reference, see Decker (2019)). Thus, the treated periods are:

Table 1: Periods where Parties are in coalition

	Period 1 2013-12-17 -2017-02-07	Period 2 2017-09-27 -2017-11-19	Period 3 2017-11-20 -End of observation
AfD			
CDU	X	X	X
CSU	X	X	X
Grüne		X	
Linke			
SPD	X		X

Note: Periods 1 and 3 are from signing of the coalition contract until 6 months before the election. Period 2 is from beginning of Jamaica negotiations until their end.

Admittedly, this is not perfect, since the parties in coalition do not change very much, which would strengthen identification.

The regression is specified as follows, with subscript i denoting the party, t the time and Y_{it} our outcome variable (We-narrative tweets / total number of tweets). β is our coefficient of

interest and $Coalition_{it}$ the dummy variable that is 1 when a party is in a coalition or coalition building phase and 0 otherwise (see Table 1). γ_i is the party fixed effect, δ_t the time fixed effect and ϵ_{it} the random error term.

$$Y_{it} = \beta Coalitio_{it} + \gamma_i + \delta_t + \epsilon_{it}$$

Table 2 presents the results of this regression.

Table 2: Results of two-way fixed effects regression

	Percentage of Tweets using We-Narratives	
	(1) Full Sample	(2) Contract-Coalitions
<i>Coalition</i>	0.0034** (0.0016)	0.0040** (0.0016)
Time Fixed Effects	Yes	Yes
Party Fixed Effects	Yes	Yes
<i>N</i>	15220	15220
Parties	5	5
Time Periods	2557	2557

Notes: Standard errors are in round brackets. Full Sample includes Jamaica Coalition building period, while Contract-Coalitions does not. The Regression uses Time and Party Fixed effects. ** denotes significance at 5 % level.

Results listed in Column (1) show that parties that are in a coalition use We-narratives more than parties that are not in a coalition. The magnitude of the parameter might seem small, but since the sample overall average use of We-narratives is at 0.06, this constitutes more than a 5% increase. Since the Jamaica Period does not constitute a classical coalition, Column (2) reports results for just the two "GroKo" periods (This excludes Period 2 from Table 1). Here, the effect is even larger.

My results should however not to be interpreted as causal evidence. Instead, I am just measuring a correlation, where some other effect reasons are excluded. For this to be causal evidence, a deeper analysis would be necessary.⁶

⁶For example with difference-in-differences, the groups need to be comparable pre-treatment and not be affected differently over the time period by other events to have causal effect estimation. This seems hard to argue since control (opposition parties) and treatment groups (coalition parties) would have to be comparable.

4 Discussion

Admittedly, my identification of what corresponds to a we-narrative is not perfect. Works like Bekhta (2017) distinct different strengths of we-narratives. She argues that only "narratives in which collective subjectivity defines the dominant mode of narration" (Bekhta (2017), p. 3) should be considered a we-narrative. Within a sentence or tweet, I argue that if the verb is in we-form, the dominant mode of narration is the we. However, across multiple tweets which may be connected, the dominant mode of narration is harder to identify. Ideally, an optimal method could do this. So while my presented method and regression analysis are using state of the art techniques to identify narratives, there is still significant headroom.

5 Conclusion

In this paper, I have demonstrated a method of identifying we-narratives in tweets. I have applied this method to the setting of German politics. In this setting, I took coalitions as an example for a social group. Using this setup, I demonstrated that politicians use we-narratives about 5% more when they are in a coalition. This serves as an example of groups using we narratives to communicate goals, building identity and stability. My results support theory on collective action and use of we-narratives.

References

- ANDRE, P., I. HAALAND, C. ROTH, AND J. WOHLFART (2021): “Narratives about the Macroeconomy,” ECONtribute Discussion Papers Series 127, University of Bonn and University of Cologne, Germany.
- ASH, E., G. GAUTHIER, AND P. WIDMER (2022): “RELATIO: Text Semantics Capture Political and Economic Narratives,” .
- ASH, E., AND S. HANSEN (2023): “Text Algorithms in Economics,” *Annual Review of Economics*, 15(1), 659–688.
- BEKHTA, N. (2017): “We-Narratives: The Distinctiveness of Collective Narration,” *Narrative*, 25(2), 164–181.
- CARR, D. (1986): “Narrative and the Real World: An Argument for Continuity,” *History and Theory*, 25(2), 117–131.
- DECKER, F. (2019): *Über Jamaika zur Fortsetzung der Großen Koalition. Die Entwicklung des Parteiensystems vor und nach der Bundestagswahl 2017*pp. 201–224. Springer Fachmedien Wiesbaden, Wiesbaden.
- GALLAGHER, S., AND D. TOLLEFSEN (2017): “Advancing the ‘We’ Through Narrative,” *Topoi*, 38(1), 211–219.
- HONNIBAL, M., I. MONTANI, S. VAN LANDEGHEM, AND A. BOYD (2020): “spaCy: Industrial-strength Natural Language Processing in Python,” .
- SCHMID, H. B. (2014): “Plural Self-Awareness,” *Phenomenology and the Cognitive Sciences*, 13(1), 7–24.
- SHENHAV, S. R. (2006): “Political Narratives and Political Reality,” *International Political Science Review*, 27(3), 245–262.
- WOOLDRIDGE, J. (2012): *Introductory Econometrics*. CENGAGE Learning Custom Publishing, Mason, OH, 5 edn.