

# Same Language, Different Contexts: Populist Rhetoric in Austria, Germany, and Switzerland

Julian Bernauer, Ines Rehbein, and Simone Paolo Ponzetto

2020-04-07 11:30:12

*Paper Prepared for the Conference “The Language of Polarization: Methodological Challenges and Innovative Approaches in Quantitative Text Analysis”, Düsseldorf, July 2-3, 2020*

## Abstract

While populism is an abundant phenomenon in politics, it’s quite hard to grasp in political text. In other words, it is a latent concept and needs an operationalisation. To this end, this paper seeks patterns of populist rhetoric in text, keeping in mind that these are more than a count of for instance the word “people”. We are looking for textual traces of populism with the help of tools of natural language processing. The data used are 66 national election party manifestos from the three democracies Austria, Germany, and Switzerland, covering the period 2000 to (early) 2020. While these are all of German language (not considering some regional Swiss manifestos written in Italian or French only), national context matters. Populism can have quite varying connotations and vocabularies in the three countries, as well as over time. Take, for instance, the word “Volk”, which is associated with the old rule of the people and direct democracy in Switzerland, and has a more nationalist or right-extremist flavour in Germany. Hence, methods are used that not only extract populist frames or degrees of populism from political text, but also allow variation across countries and time.

## Introduction

Populism is unlike other political phenomena. Is it an issue? An ideology? Or rhetoric? Following a trend towards definitional convergence in the literature we understand populism as an attribute of a message which equals an “anti-elite discourse in the name of the sovereign people” [Aslanidis2018: 1242]. Such rhetoric is often linked to political issues such as immigration or corruption, but these do not define populism.

Detecting sparse and complex elements of this kind in political text is not easy. Fortunately, tools of natural language processing are available which are able to identify similar rhetorical patterns beyond the use of predefined, bag-of-words dictionaries. To this end, we rely on word embeddings (REF). These allow the detection of rhetoric patterns resembling populism. Also, framing theory [?] provides a lens through which populism can be observed. It argues that populist rhetoric is reflected in frames that [...]

Populism has been measured from political text before by a number of scholars (Aslanidis 2018; Bernhard et al. 2015; Ernst et al. 2017; Pauwels 2011; Rooduijn et al. 2014 and Rooduijn and Pauwels 2011). Most of the efforts rely on a sophisticated dictionary approach in combination with hand coding. A prominent example featured in the press (Guardian) is the holistic coding of leader speeches by Hawkins (REF). The notable exceptions are Pauwels (2011) and Rooduijn and Pauwels (2011), who apply a semi-automated method of counting words from a (translated) dictionary of populist words and compare the results (shares of populist words) to their own hand-coding effort. The major challenge for further research is that populism is most

purely captured by the simultaneous occurrence of references to the people, the elites and their interrelation, as for instance by the approach of clause-based semantic analysis advocated by Aslanidis (2018). We seek to advance the detection of such semantic structures indicating populism.

## The Disconnect – A Conceptual Framework

This paper aims at an exploration of the semantic structure of populism expressed in political text, we need adequate methodology able to find descriptions of relationships between “the people” and “the elites”.

Context matters

Built-In context consideration in hand coding?

Hawkins holistic grading

But does it accommodate relative populism? Context?

-> Trump is much more a breach of taboo compared to Chavez? Should take into account level of democracy? It's the provocation.

## Methods

For the exploration of populism in political manifestos, a twofold methodological approach is chosen. First, to explore the structure of populist rhetoric, an array of simpler analytical tools is deployed. These include a populism dictionary provided by @rooduijn as well as tools for the assessment of keyword context, sentiment and language complexity from the R package `quanteda`.

Second, after getting a feel for what we are able to be looking for, a more advanced method of quantitative text analysis is used. Drawing on work in natural language processing, we propose a variant of unsupervised scaling based on word embeddings [following @glavavs2017; see also @mikolov2013]. The analytical steps after preprocessing the texts into single sentences are:

1. Identify some “populist text” manually to start with. This can be done by either identifying prototypical populist text from external manual coding or prior knowledge or based on tools such as a populism dictionary.
2. Extract “populist” triples or other semantic constructs (dependence parsing) describing a relationship between the “evil” elites and the “pure” people being “robbed” of their sovereignty or acts of “betrayal” committed by the elites. An example from the 2017 German AfD manifesto is the sentence “Das Volk muss wieder zum Souverän werden - Heimlicher Souverän in Deutschland ist eine kleine, machtvolle politische Oligarchie, die sich in den bestehenden politischen Parteien ausgebildet hat.” This is populism in its purest form. For our purposes, we would extract semantic triples or quadruples such as “volk souverän oligarchie”, “volk souverän partei\*”, “deutschland kleine oligarchie”, or “volk souverän heimlich parteien” from the sentence.
3. From this “seed”, further word combinations are learned using word embeddings of the German language space. The texts on which the embeddings are applied on are pre-processed into possible word combinations.
4. The resulting, partially manually defined and partially machine-learned prototypical dictionary of populist triples is then used for another round of machine learning looking for the semantic similarity with the word combinations extracted from the texts of interest.
5. A final step is the aggregation at the text level, using those triples above a certain threshold of similarity to scale texts according to their degree of populist rhetoric using a suitable method such as harmonic label propagation [@glavavs2017]

An obvious problem with the method can be the selection of the seed. The results are influenced by its selection. Several countermeasures exist to alleviate this problem, though. The seed is taken from different party manifestos, and not only from those of parties already expected to be populist. They are chosen based on theoretical considerations regarding the definition of populism only. Furthermore, as we are learning further instances of populism via word embeddings, potential bias towards certain formulations of populism is reduced. The method is able to find statements with similar meaning, but a different wording by exploring the semantic space of language.

A further challenge is the consideration of context.

Classification - Casthano Silva

TopFish

SemScale

babelfish

## babelfish

The model described is put into an R package called **babelfish**.

The full code is available on the GitHub repository of the paper, see <https://github.com/julianbernauer/duesseldorf>.

## Data

For the demonstration of the approach selected, a small corpus of manifestos from Austria, Germany and Switzerland is used (source: [polidoc.net](http://polidoc.net)). We include all elections between the years 2000 and (early) 2020. This results in a sample of 66 manifestos from XX unique parties. Most of them are observed XX times, which is the maximum number of elections covered.

Some descriptive statistics on the manifestos included and the respective text length are displayed in Figure XX.

## Warm-Up with a Dictionary

Dictionaries are a way of retrieving information from text when the concept of interest is well-defined. In an early attempt of measuring populism quantitatively, @rooduijn provide a study centred around a populism dictionary. For several languages (English, German, Dutch and Italian), they present a number of populist keywords, each slightly amended to the context they refer in. For instance, political elites in Italy are at times critically referred to as the “partitocrazia”, and the word is included in the Italian dictionary only. The quantity of interest derived in the end is the share of *words* from the self-devised populism dictionaries in parties’ election manifestos. The content of the German dictionary, turned into three sub-categories using the *quanteda* package in R, is:

```
pop_dict_de
```

```
## Dictionary object with 3 key entries.
## - [elites]:
##   - elit*, politiker*, establishm*, *herrsch*
## - [corrupt]:
##   - korrupt*, propagand*, täusch*, betrüg*, betrug*, *verrat*, scham*, schäm*, skandal*, wahrheit*, v
```

```
## - [sovereignty]:
##   - konsens*, undemokratisch*, referend*
```

We apply the dictionary to the set of 67 manifestos.

```
## Corpus consisting of 66 documents, showing 6 documents:
##
##           Text Types Tokens Sentences partyyear ctry
## 41113.000.2009.1.1.txt 7787 39332      2199 Grüne 2009 DE
## 41113.000.2013.1.1.txt 13722 94065      5134 Grüne 2013 DE
## 41113.000.2017.1.1.txt 10772 69371      3642 Grüne 2017 DE
## 41223.000.2009.1.1.txt 5721 22495        820 Linke 2009 DE
## 41223.000.2013.1.1.txt 8370 43695      1852 Linke 2013 DE
## 41223.000.2017.1.1.txt 10553 64056      3837 Linke 2017 DE
##
## Source: M:/duesseldorf/* on x86-64 by Dr. J
## Created: Tue Apr 07 11:30:19 2020
## Notes:

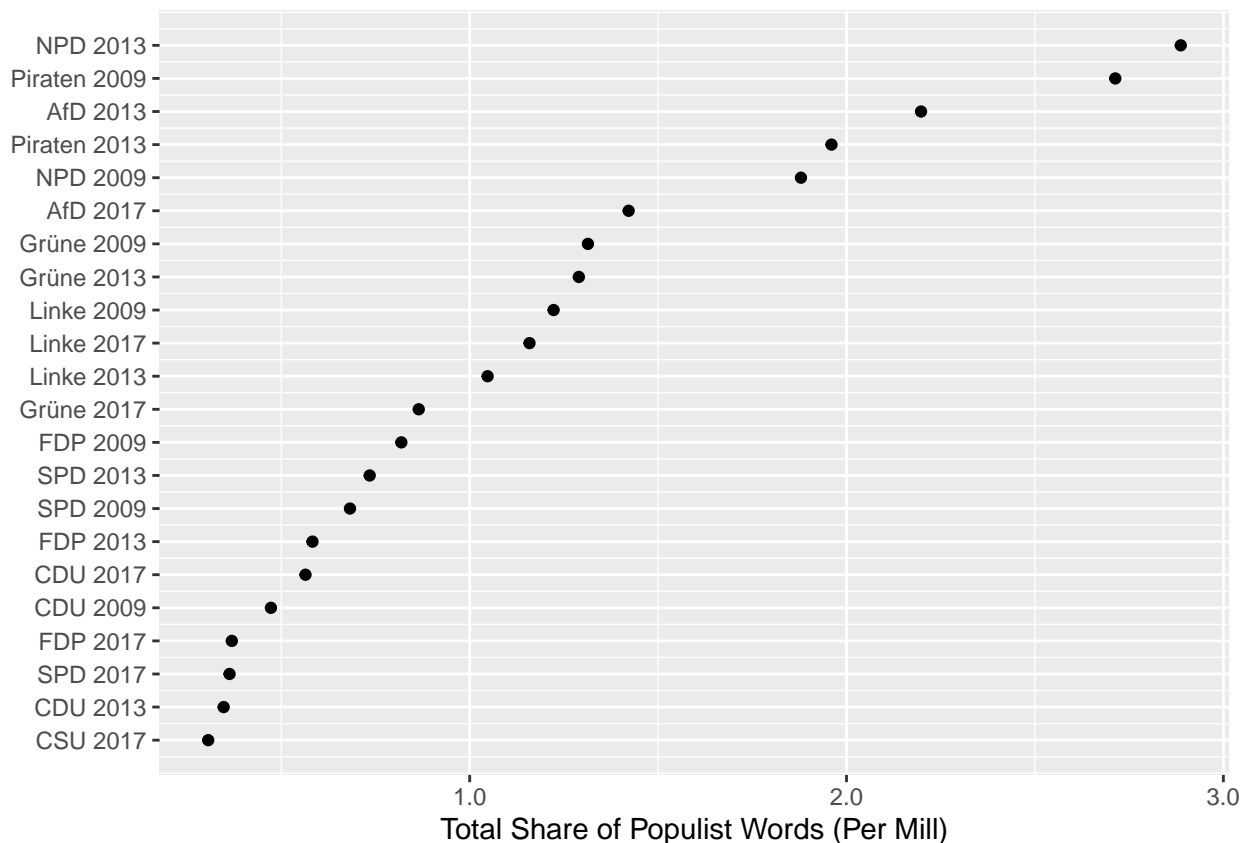
## Document-feature matrix of: 6 documents, 3 features (5.56% sparse).
## 6 x 3 sparse Matrix of class "dfm"
##           features
## docs      elites corrupt sovereignty
## 41113.000.2009.1.1.txt      8      14          3
## 41113.000.2013.1.1.txt     16      36          7
## 41113.000.2017.1.1.txt      3      25          1
## 41223.000.2009.1.1.txt      9       5          0
## 41223.000.2013.1.1.txt     12       8          2
## 41223.000.2017.1.1.txt     18      11          7
```

The share of allegedly populist content in the manifestos corresponds to the share of populist words in the total number of non-stopwords. As “populism” is the only feature for which a word count exists, we can use quanteda’s `ntoken()` function to get the number of populist references per manifesto as well as the total number of tokens in per text from the document-feature matrix “dfm\_uk”, where stopwords have been removed.

per mill

More generally, the count of single features in a document-feature matrix can be extracted in the same fashion as any other matrix, using for instance `dfm_uk_pop[, "populism"]`. A base R data frame suffices to create the starting point of a plot of the relative amount of populist rethoric in the manifestos studied. We also add a party variable to label the data in the plot to come.

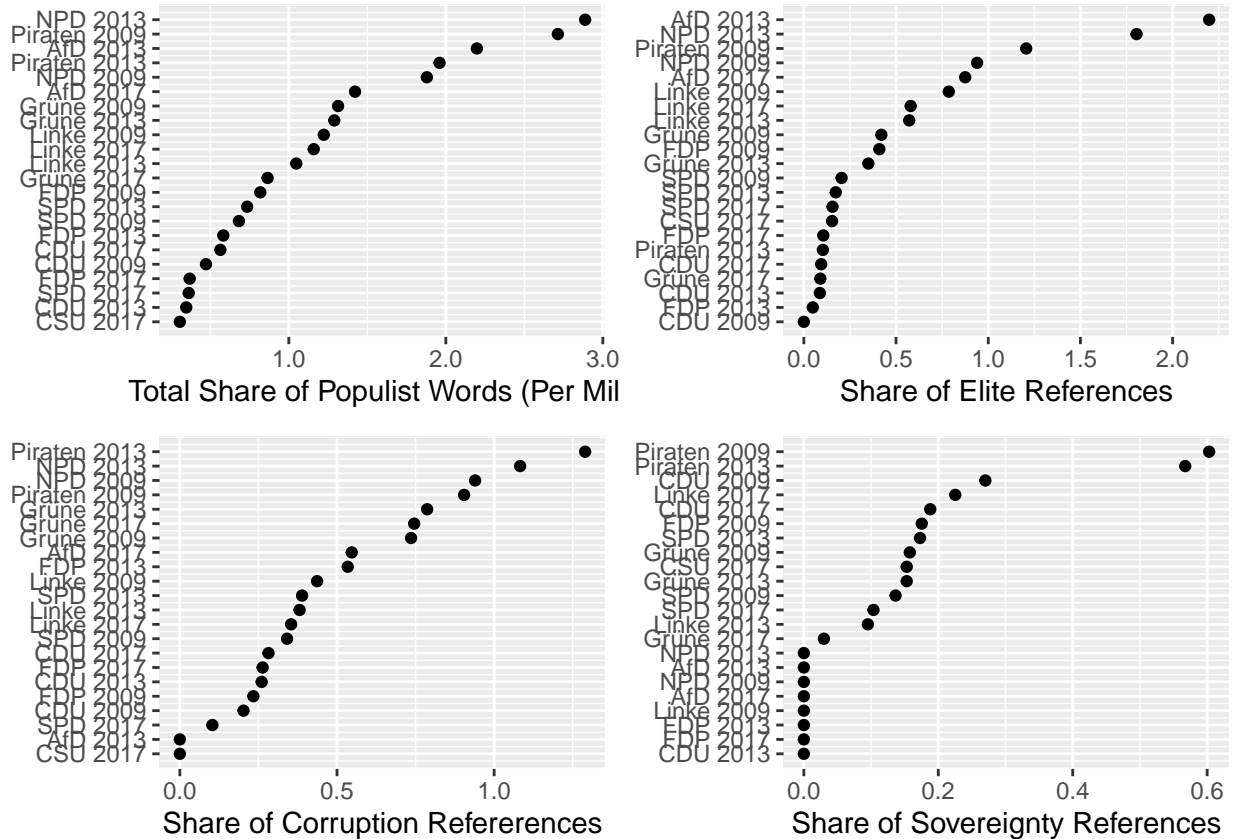
Using `ggplot()`, an aesthetic plot of the total share of populist words in the manifestos is generated:



Accordingly, and considering the length of the manifestos, the NPD in 2013 has the “highest” share of “populism”. As an extremist right party, . The value is still just above 10 words per 1000.

still only a mere 0.2 per cent of the total words. For the Northern Irish Democratic Unionist Party, the number is as low as 0.07 per cent. Other regionalist parties display a relatively high share of populist rhetoric in their manifestos, while the two main parties as well as the liberal and green contenders are placed in the middle.

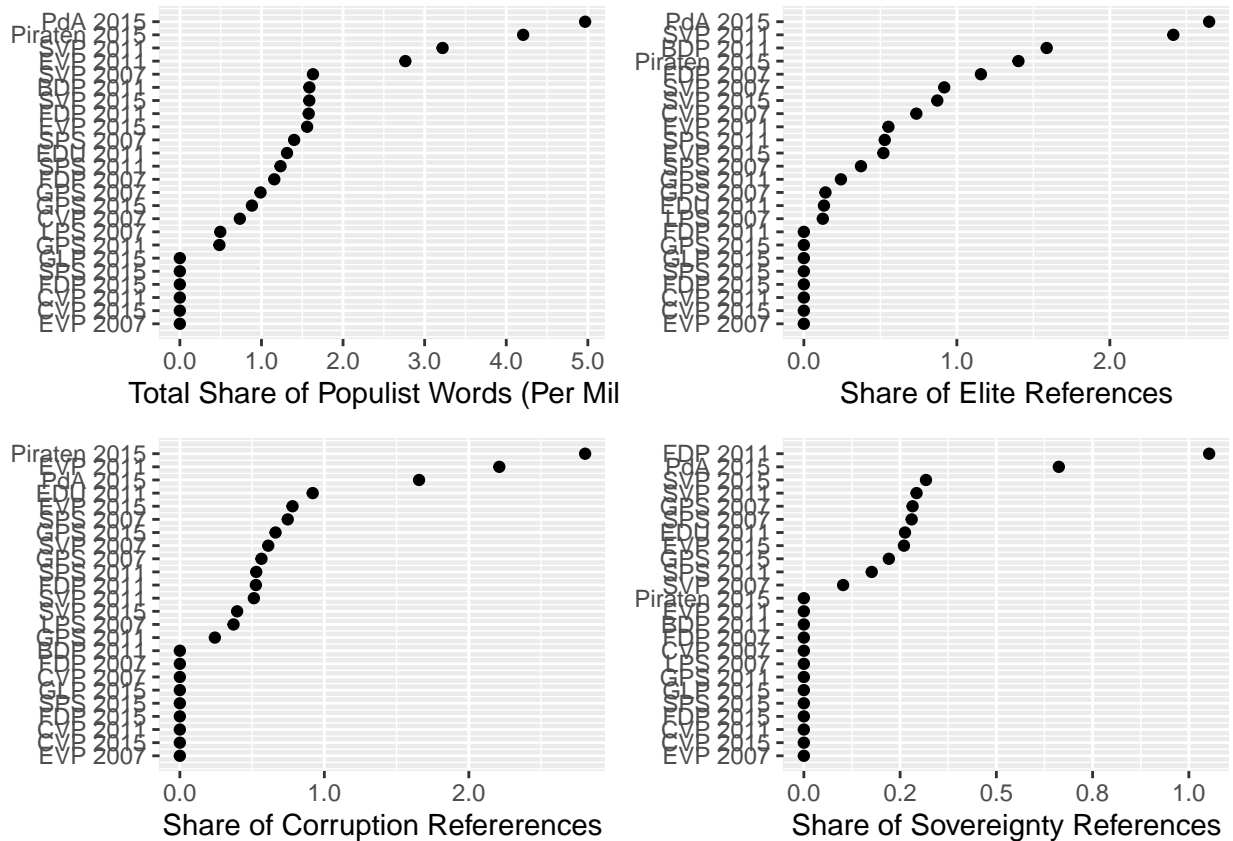
In a further step, we might also want to know whether UKIP and others are appealing more to the people or are critical towards the elites. This is where the alternative dictionary with subcategories (“elites”, “corrupt”, and “sovereignty”) comes in. The full analysis adding this variant is carried out with this code, integration the four categories of all populist words as well as the share of elite, corruption and sovereignty references are integrated into a single plot in the end:



The AfD 2013 manifesto is ranked first in elite references despite the fact that only two of them exist as the document only features 951 words (without stopwords). Hence, the share is over 2 per mill.

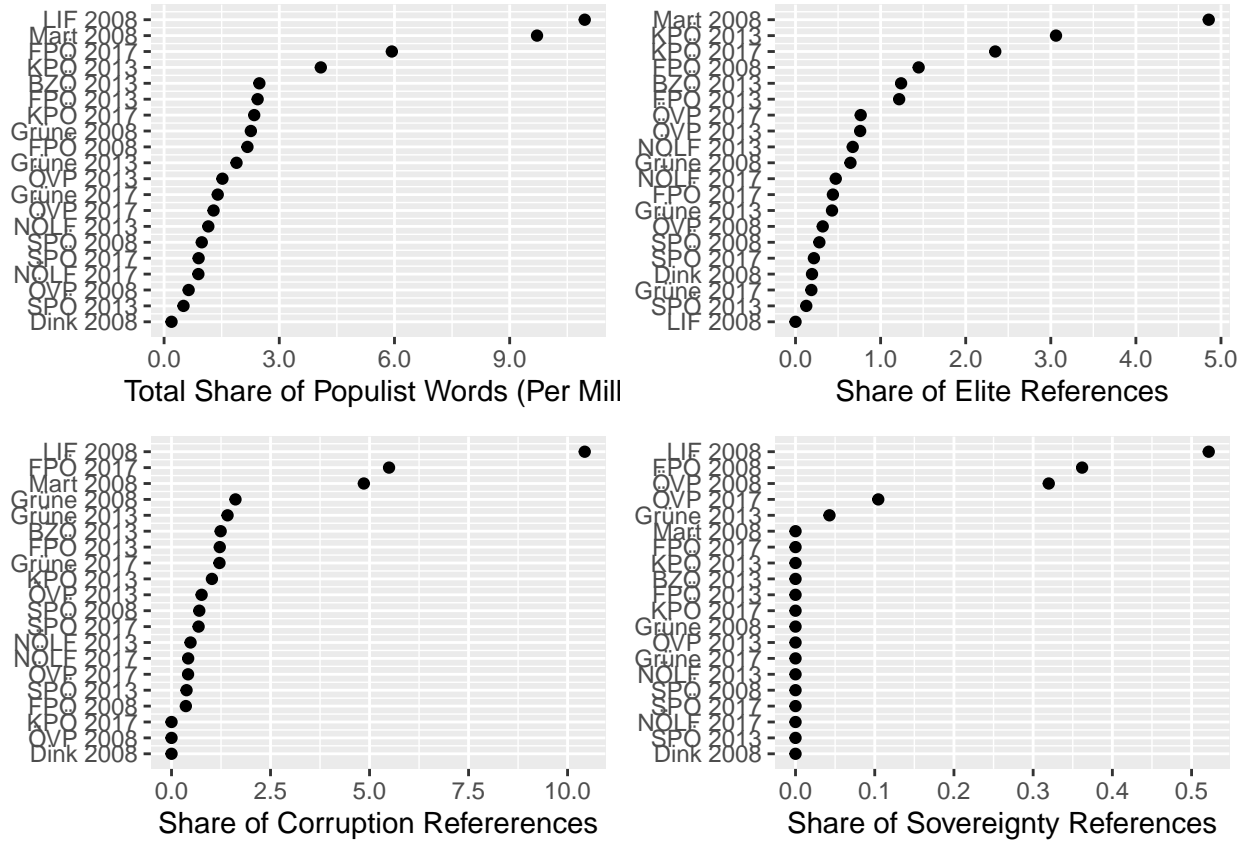
From this overview, we learn that the CDU 2009's leading rank in the overall populism score is driven by its relatively high share of "populist" words in the subcategories "corruption references" and "sovereignty references". The manifesto features no elite references. The AfD, Green and Linke manifestos of 2017 lead the way in the elite category, giving some evidence that elite references might be more suitable for capturing populism than the other categories. Along these lines, human interpretation is still crucial. We investigate the context of some "corruption" references in the CDU 2009 manifesto by displaying it's context.

```
##
## [41953.000.2017.1.1.txt, 1058]
## [41953.000.2017.1.1.txt, 2792]
## [41953.000.2017.1.1.txt, 8211]
##
## die Jahre erodieren lassen und zu einer überbordenden Staatsgewalt geführt: Ehemalige
## diese Entwicklung nicht klaglos hinnehmen und sich zu Recht gegen die politischen
## Loyalität zu unserem Staat. Das Minarett lehnt die AfD als islamisches
##
## | Politiker |
## | Eliten |
## | Herrschaftszeichen |
##
## auf Richterstühlen, Abgeordnete die zugleich Kanzler, Ministerpräsident oder Minister sind
## auflehnen. Die suprastaatliche Rettungspolitik verletzt geltendes Recht, darunter das Verbot
## ebenso ab wie den Muezzin-Ruf, nach dem es außer dem islamischen
```



```
##
## [43220.000.2015.1.1.txt, 79]
## [43220.000.2015.1.1.txt, 143]
## [43220.000.2015.1.1.txt, 3590]
##
## die für sie wichtigen politischen Entscheidungsprozesse. Das Vergeben von Mandaten an
## Prozent- ein deutlicher Anstieg, welcher die zunehmende direkte Einbindung der
## Stelle fest, dass das Privateigentum an Produktionsmitteln die Grundlage der kapitalistischen
##
## | PolitikerInnen |
## | Politiker |
## | Herrschaft |
##
## ist ein wichtiger Weg, um auf politische Prozesse Einfluss zu nehmen
## belegt. Nur eine ganz kleine Anzahl dieser Mandate vertritt die Interessen
## ist. Reaktionäre Antworten Angesichts der Massenproteste und des steigenden Widerstands benutzt
```

Among these three mentions of politicians or the ruling order, one is actually a praise of representative democracy, one resembles populism, and one is a MARxist critique of the economic order. Hence, the accuracy of the populism dictionary is questionable here.



```
##
## [42421.000.2008.1.1.txt, 190]
## [42421.000.2008.1.1.txt, 1113]
## [42421.000.2008.1.1.txt, 3316]
##
## :„ Faire Chancen am Start" Kinder | Unfair |
## Bereich attraktiv machen Wissenschaft, Forschung& Entwicklung | Unfair |
## Mediensprecherin:„ Demokratie braucht Öffentlichkeit" Europa | Unfair |
##
## ist:- dass die Chancen von Kindern
## ist:- dass Kinder mit sozial schwachem
## ist:- dass durch einen rücksichtslosen Anti-Europa-Populismus
```

The example shows that the high score of the 2009 CDU manifestos stems from a list of statements all starting with the rhetoric pattern “Unfair ist:”. This might capture a policy positions aimed at the removal of social inequalities, but is clearly off in terms of measuring populism. Indeed, one of the “unfair” things raised by the manifesto is populism, as the third of the three instances (out of 20 in total) displayed shows.

## Detecting Populism using Machine Learning

Obviously, the populism dictionary is off at times, mainly due to the disconnect between the different categories defined. One intriguing way of detecting semantic structures are word embeddings.



## Holistic grading

available for x of x texts

- add holistic grading via Hiwis ??

## Looking for populist frames

## LIWC emotions ?

## Sentiment

## Language Structure

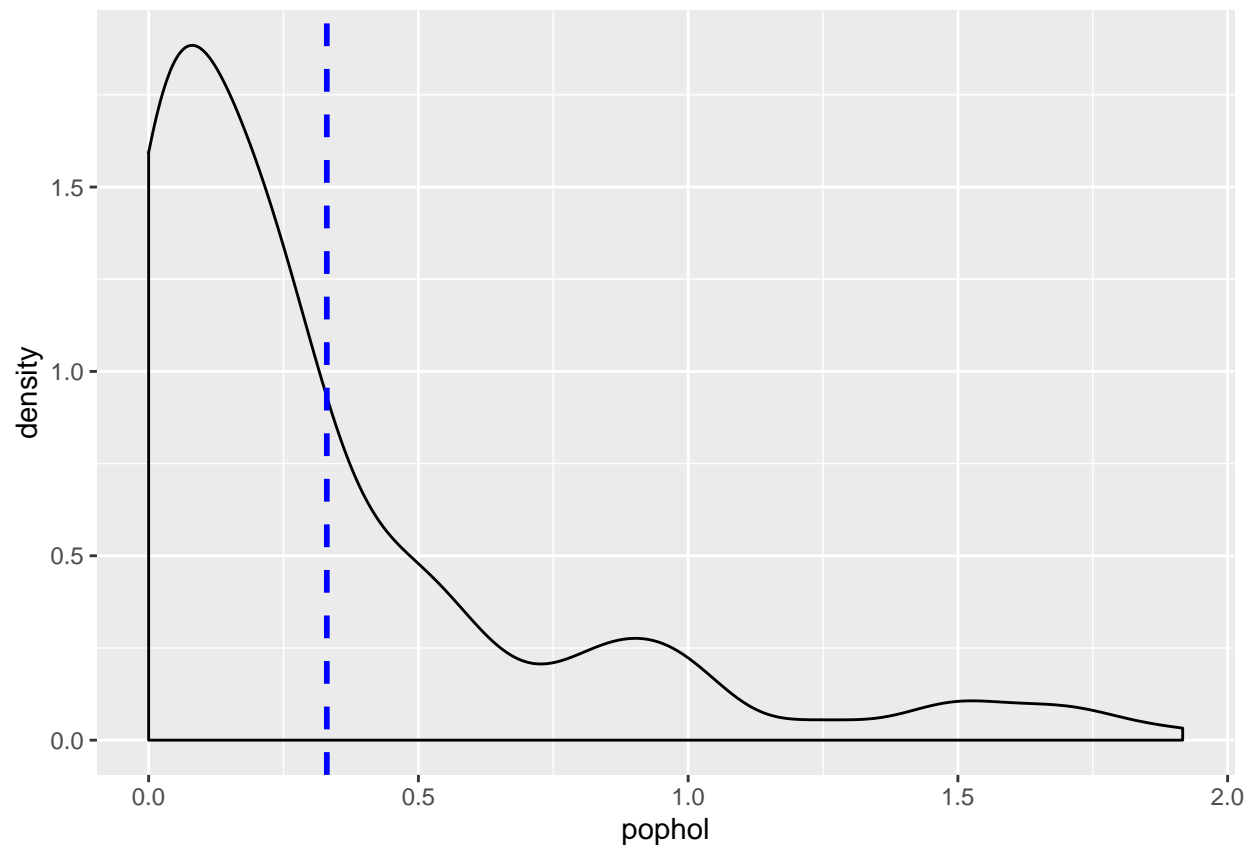
## Contextual Remedy

- single-country analyses of frames
- multilingual embeddings
- 

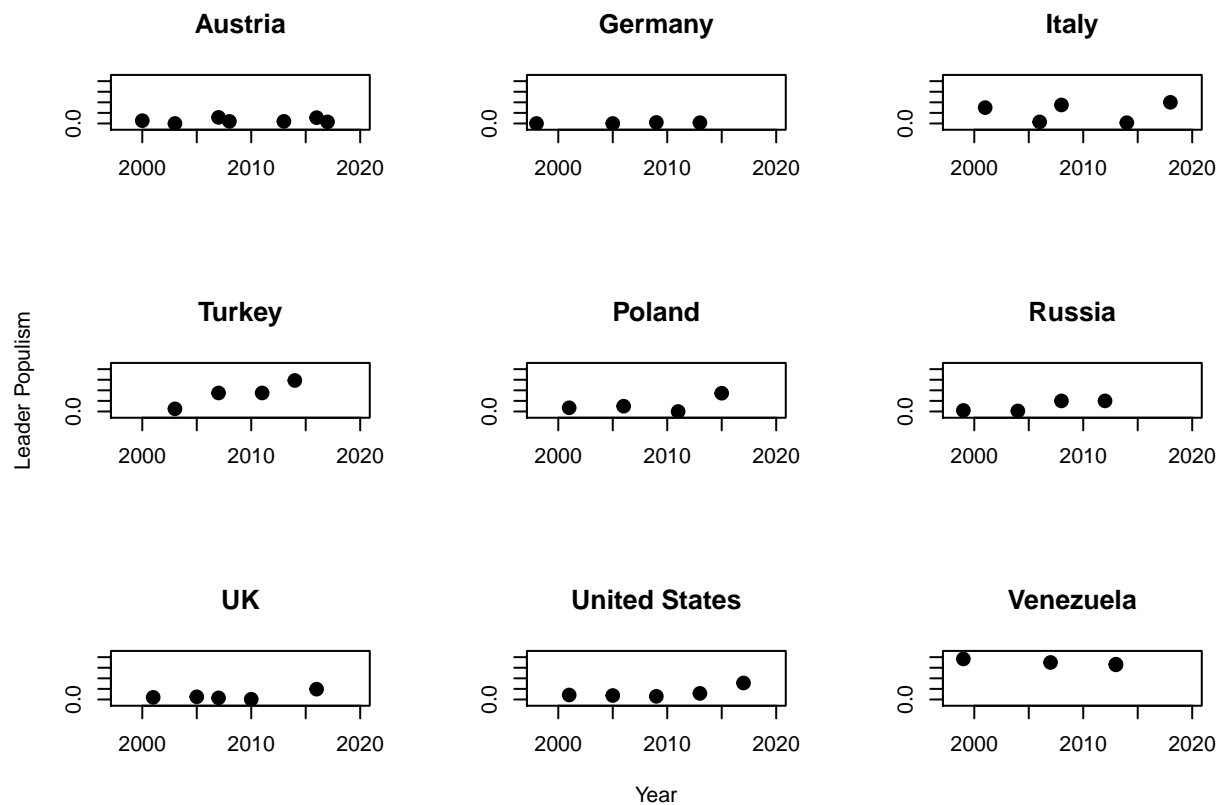
## Contextual Variation

### Across Countries and Languages

The motivation of this research is the expectation that the meaning of populism varies across contexts. A first indication of this would be if we find suspicious patterns in existing data. To explore the structure of existing populism measures, we make use of the Hawkins et al. 2019 Global Populism Database, which features speech data from 182 leader-terms across 40 countries between 2000 and 2018. The text has been manually coded using holistic grading by multiple raters. The resulting scores are quasi-continuous, as the categorical ratings as not populist (0), somewhat populist (1), and very populist (2) are averaged across coders.



The holistic populism scores are heavily skewed to the right and truncated at zero. This already needs to be considered in the modelling of the data. A beta distribution suggests itself, as it allows modelling skewed, bounded distributions via two shape parameters. Given that the beta distribution ranges from 0 to 1, and the holistic populism scores range from 0 to 2, a workaround is needed in rescaling the range to 0 to 1.



```
## Linear mixed model fit by REML ['lmerMod']
## Formula: pophol ~ 1 + (1 | country)
## Data: hawk
## REML criterion at convergence: 149.0017
## Random effects:
## Groups Name Std.Dev.
## country (Intercept) 0.2840
## Residual 0.3031
## Number of obs: 182, groups: country, 40
## Fixed Effects:
## (Intercept)
## 0.3292
```

```
## [1] 0.4837336
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: pophol ~ 1 + (1 | indi)
## Data: hawk
## REML criterion at convergence: 139.4891
## Random effects:
## Groups Name Std.Dev.
## indi (Intercept) 0.3113
## Residual 0.2160
## Number of obs: 182, groups: indi, 140
## Fixed Effects:
```

```
## (Intercept)
##          0.2936

## [1] 0.590366
```

The distribution of the scores is:

The country-wise levels are:

From the visual inspection, we already see that populism varies a lot between and less within countries.

Different- and Same-Language Texts

Using the Hawkins Data

## Modified Wordfish

EPSA 2018 - naja, maybe show model

The figure shows an application of the method to 66 German language party manifestos from Austria, German, and Switzerland. Apparently, some parties known to be populist such as the Swiss SVP, the German Linke and the Austrian FPÖ receive high populism scores, but also less suspicious ones such as the German CDU. Comparing the resulting “populist rhetoric” scores to an existing manual coding of the populism in manifestos which exists for a subset of 17 texts [Hawkins2018], the correlation stands at .49, which is moderate.

10 prototypical sentences instead of triplet?

## Adding context

see semscale

harmonic function propagation

add clusters?

mark german from switzerland as different from other german?

## Frames

## Conclusions

The method is able to identify populism in party manifestos and scale texts according to their level of populism to a certain degree. But the results are mixed, and obvious problems include “false positives” such as the German CDU or Greens as well as some clustering by country, indicating the existence of context effects beyond what can be captured by word embeddings.

## References