

Party differences in minority-related discourse in German election manifestos: A text analysis approach

Ria Timm

January 22, 2026

Abstract

This paper examines how major German parties differ in the themes and keywords they use when discussing minorities, equality, discrimination and societal norms. Using annotated manifesto segments from the Manifesto Project ¹, the study combines descriptive trends over time with TF-IDF-based keyword profiling and topic modelling via Non-negative Matrix Factorization.

1 Introduction

This project investigates whether German parties use systematically different language when addressing minorities and equality. The guiding research question is: *How do parties in Germany differ in the keywords and themes they use when they speak about minorities, discrimination, and societal norms?*

2 Data and Corpus Construction

The dataset is drawn from the Manifesto Project API¹. Because the platform does not provide a single download that includes both the full texts and all metadata in one step, the corpus is built in stages. First, the German subset of the core manifesto table is downloaded. Party and election identifiers are then used to request the corresponding metadata. Only manifestos for which full text is available are kept. The annotated text is then downloaded and filtered again, keeping only the quasi-sentence segments that belong to four policy categories used by the Manifesto Project coding scheme:

- per503: Equality (Positive)
- per705: Underprivileged Minority Groups

¹<https://manifesto-project.wzb.eu/>

- **per603**: Traditional Morality (Positive)
- **per604**: Traditional Morality (Negative)

The analysis focuses on major parties currently in the Bundestag or represented in the previous term: Greens, AfD, CDU/CSU, FDP, SPD, and The Left.

3 Methods

3.1 Descriptive Analysis

First, to get a broader overview of the data, the manifesto segments were counted by year, party and code. Results are visualized both as raw counts and as within-party-year shares over the four selected codes, to reduce the influence of different manifesto lengths.

3.2 TF-IDF Keyword Profiling

For TF-IDF the text is preprocessed by lowercasing, removing links and basic punctuation, then removing German stopwords. A TF-IDF model is fit with unigrams and bigrams (`ngram_range=(1,2)`), using `min_df=5` and `max_df=0.8`. For each party, the mean TF-IDF weights are computed and the top terms are interpreted.

3.3 Term Co-occurrence

Next, word co-occurrences were counted and extracts frequent term pairs. In this corpus, the method yields limited interpretive value.

3.4 Topic Modelling

Topic modelling is implemented via Non-negative Matrix Factorization (NMF) with eight topics (`n_topics=8`, `random_state=42`). NMF is run (1) separately per party to explore party-specific topic structures and (2) globally to compare average topic weights by party in a shared topic space.

4 Results

4.1 Descriptive Patterns Over Time

Equality (Positive), **per503**, appears far more often than the other categories in this selection. In raw counts, it is most strongly associated with the Greens and The Left. Traditional Morality (Positive), **per603**, shows a shift from being very present for CDU/CSU in

the 2000s to AfD overtaking later. Traditional Morality (Negative), **per604**, peaks across parties in the early 2010s and declines thereafter. Underprivileged Minority Groups, **per705**, peaks around 2009 and then declines after the early 2010s.

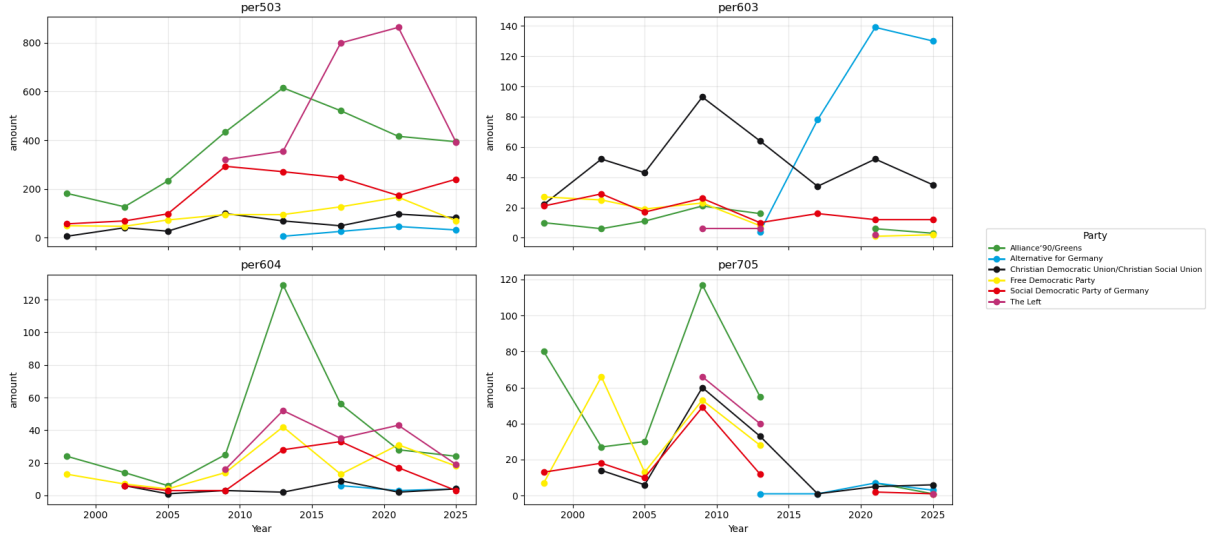


Figure 1: Raw counts of manifesto segments by party and year for each Manifesto Project code (**per503**, **per603**, **per604**, **per705**). Replace the file name with your exported plot.

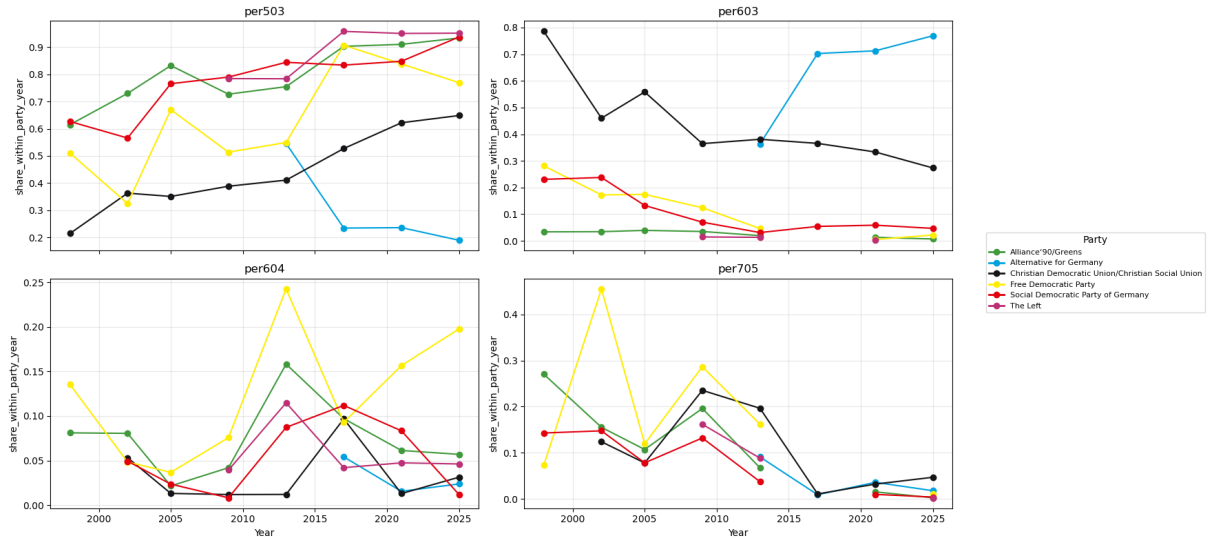


Figure 2: Within-party-year share of manifesto segments by party and year for each Manifesto Project code (**per503**, **per603**, **per604**, **per705**).

The within-party-year share view reinforces these differences. Several parties converge toward very high shares of **per503** over time, while AfD shows a decline in **per503** after its emergence around 2013. For **per603**, AfD becomes a clear outlier with high share, while CDU/CSU declines. FDP stands out in the normalized view for **per604** and **per705** with notable peaks.

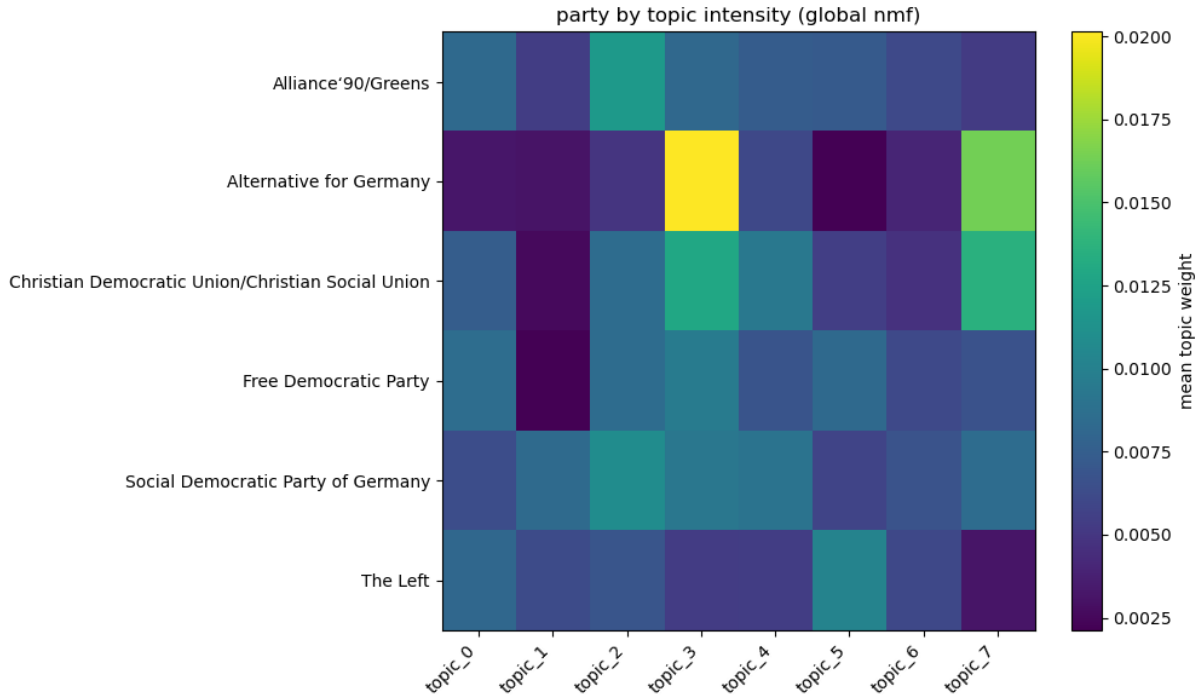


Figure 3: Within-party-year shares of the four codes. Higher values indicate stronger concentration of a party’s content in that year within a given code relative to the other codes in the set. Replace the file name with your exported plot.

4.2 TF-IDF Keyword Profiles by Party

TF-IDF yields distinctive party-level vocabularies. Greens emphasize participation, justice, discrimination and women. AfD is dominated by family-centered vocabulary such as families, children, parents and family policy, alongside terms like gender and abortion, suggesting a focus on family norms and cultural conflict topics. CDU/CSU similarly foregrounds families and children but combines this with language of responsibility, support and strength. FDP terms cluster around freedom and opportunity and include modal verbs such as should and must. SPD combines equality language with social themes such as income and justice, while The Left stands out through strongly ‘material’ language including euro and percent alongside income and work.

4.3 Co-occurrence Analysis

Co-occurrence networks were explored but provided limited additional insight for this corpus, likely because quasi-sentence segments are short and many co-occurring terms are generic.

4.4 Topic Modelling with NMF

Party-specific NMF topics show differences in thematic clustering. The Greens produced topics around disability inclusion, gender equality at work, anti-discrimination rights, participation through education and poverty and diversity. AfD clusters strongly around traditional family models, tax splitting, demography and birth rates, abortion opposition and anti-gender framing. CDU/CSU topics emphasize christian values, marriage and family, parental responsibility, churches, inter-generational issues and disability participation. FDP topics connect modernization and accessibility with self-determined living, equal opportunity regardless of origin and work-family balance. SPD topics foreground social security, cohesion, gender equality, work-family time policy, and support for families including single parents. The Left topics emphasize poverty and accessibility, fighting inequality, redistribution and taxation, east-west equality, and a socially fair transition.

The global topic heatmap highlights the differences between the parties. AfD stands out most strongly on a topic interpreted as party self-positioning and messaging. Greens and SPD show stronger signals on a wealth and taxation-related topic. AfD and CDU/CSU are high on a topic characterized by fiscal totals and revenue figures, while The Left is very low there. The Left is relatively high on the socially fair transition topic and FDP appears comparatively even across topics without a single dominant spike.

5 Discussion

The analysis suggests that parties differ not only in how often they address minority-related issues and how central they are for their program, but also in how they frame them. Greens and The Left emphasize inclusion, rights, participation and social justice. SPD combines equality language with social security and redistribution. FDP frames minority-related themes through individual opportunity, while CDU/CSU connects the discourse to family, responsibility and values. AfD concentrates on traditional family norms and potentially also uses conflict-centered cultural themes, including anti-gender and abortion-related language.

Several limitations should be acknowledged. First, the corpus consists of election manifestos rather than parliamentary speeches or passed policies, so conclusions concern programmatic rhetoric rather than parliamentary debate or political action. Second, the study focuses on four codes, which structures what counts as minority-related discourse and may exclude other relevant categories. Third, TF-IDF and NMF are bag-of-words methods that miss linguistic context such as negation and more complex framing strategies.

Future work could extend the analysis by using more advanced topic modeling methods like LDA or BERTopic. In addition, a supervised classification task predicting party from

segments could also quantify how distinct party rhetoric is and provide complementary validation.