An analysis about the ethnic stereotypes within the plenary records of the German Bundestag from 1983-2021.

Tim Schenkel, Jo-Ann Schüller, Firat Seker and Lotte Haupt

Abstract

The analysis presented here provides insights regarding the development of ethnic stereotypes within the plenary protocols of the German Bundestag, based on all protocols between 1983 to 2021. The results of the analysis will show how the dimensions of stereotypes are developed within the German language (including attributes that refer to low status or also to high threat) and how language has transformed within this time period. For this purpose, the XML files were divided into three time spans, which will be addressed later in the analysis.

1. Introduction (Lotte Haupt)

First of all, the general concept of a prejudice must be clarified. A prejudice is a judgement given to a person, a group, a fact or a situation without a thorough and without a comprehensive investigation, clarification and consideration. There are negative and positive prejudices. Mostly, "prejudice" is meant negatively and is also understood in this way if "positive" is not explicitly prefixed as a characteristic. Prejudices exist in all societies and all social groups, classes and strata to a greater or lesser extent. (Bergmann 2008) Scientific prejudice research is concerned with the critical study of prejudices. Ethnic stereotypes are prejudices against people who belong to an ethnic or religious group. They include harmful stereotypes and less obvious but still dangerous aspects, such as bias. Recognising ethnic prejudice is important not only because it leads to more serious cases of racism, but also because it violates the constitutional right to equal treatment. This paper deals with the topic of ethnic stereotypes using plenary protocols of the German Bundestag between 1983 and 2021. A plenary protocol is the verbatim transcript of a meeting (plenary). After the end of the session, these are submitted to the President of Parliament, who checks them if necessary or receives objections to the record. In the German Bundestag, speakers also have two hours to correct the minutes, whereby the meaning of the speech or parts of it may not be changed by the corrections. As soon as the President of Parliament has signed the minutes, they are printed and published.

2. Motivation (Firat Seker)

Migration is a mass phenomenon in Germany: more than one in four people had a migration background in 2020, which is a total of 22 million people. Migration is an essential part of German history. This contribution traces the development of migration movements from the founding of the German Empire in the late 19th century to the reformation of nationality law in 2000. From the beginning of its settlement, migration as a basic element of human history has shaped the territory of today's Germany. (Jochen Oltmer 2021) Prejudices persist, poison the political climate and make a factual discussion about the topics of flight and displacement impossible. Refugees thus find it difficult to arrive, and the people and organisations that want to help them find it difficult to work. On closer examination, however, it turns out that they in no way do justice to the complex social reality or even lack any basis. Dissemination towards such stereotypical associations can contribute to the development of stereotypical beliefs by reinforcing mental links between social groups and biased attributes. This year, UN High Commissioner Filippo Grandi said the following about the current issue: "I have been working with refugees for over three decades. But in those three decades, I have never seen such cynicism, such venom in the language of politics, in the media, in social media, even in everyday conversations, as today." (Grandi 2021) Building on this thesis, we became interested in this topic and made the following three hypotheses at the beginning of our work in order to conduct a detailed analysis of the plenary transcripts.

H1: Do politicians associate ethnic foreign groups more strongly with low status characteristics than ethnic own groups?

H2: Do politicians associate foreign ethnic groups more strongly with characteristics of high threat than own ethnic groups?

H3: Over time, the strength of implicit stereotypical associations increased for ethnic foreign groups, while this is not the case for ethnic native groups.

3. Data-Set (Jo-Ann Schüller)

The data set we prepared contains a total of 1559 XML files. We downloaded them from the website of the German Bundestag (*Plenarprotokolle* 2022). The XML files include the plenary protocols of the German Bundestag of the period 1983 to 2021 and were available in the form of zipped files. An XML file is structured as follows: First the version and then the encoding (UTF-8) of the document was defined. The structure consisted of election period, document type, number, date, title and the text. The text section also includes the names of the deputies and speakers present, as well as the present parties and additional comments. The language is German. The downloads include all files up to and including the 20th parliamentary period. After the files were downloaded, we saved them to a Google Drive folder. To be able to work with them properly, all files were summarized into a pandas data frame. After this was done, the data cleaning followed.

The first cleaning iteration seems simple, but it had a fundamental effect:

1. Removal of HTML tags and </br> characters.
2. Only ASCII, and European characters and spaces, digits were removed.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE DOKUMENT SYSTEM "BUNDESTAGSDOKUMENTE.dtd">
<DOKUMENT>
  <WAHLPERIODE>10</WAHLPERIODE>
  <DOKUMENTART>PLENARPROTOKOLL</DOKUMENTART>
  <NR>10/1</NR>
  <DATUM>29.03.1983</DATUM>
  <TITEL>Plenarprotokoll vom 29.03.1983</TITEL>
  <TEXT>Plenarprotokoll 10/1
Deutscher Bundestag
Stenographischer Bericht
1. Sitzung
Bonn, Dienstag, den 29. März 1983
Inhalt:
Eröffnung der Sitzung durch den Alterspräsidenten Brandt          1 A
Bestellung vorläufiger Schriftführer    . . .     1 B
Ansprache des Alterspräsidenten          1 A
Zur Geschäftsordnung
Stratmann GRÜNE          4 A
Alterspräsident Brandt   4 C
Dr. Schäuble CDU/CSU     4 D
Wahl des Präsidenten verbunden mit Namensaufruf und Feststellung der Beschlußfähigkeit
Brandt, Alterspräsident         5 A, B, 6 A, B
```

Figure 1: Example of the structure of the XML files of the plenary protocols of the German Bundestag (Plenarprotkoll 10/1 from 29.03.1983).

3. Removal of single letters.
4. Convert all spaces (tabs etc) to single spaces.
5. Adjust punctuation.
6. Remove stop words.
7. Stemming the words.

Executing these steps should first create a solid foundation so that we could work properly with the data-set and receive viable results. Since we only refer to the language of the plenary protocols, numbers, characters and other symbols within the texts are irrelevant to us and could possibly distort the analysis. So are stop words, such as "and" and "or." Also, this cleanup allows for better readability for our program itself.

4. Method (Tim Schenkel)

In our approach, we were inspired by the study "Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage" (Anne C. Kroon 2020), which investigated ethnic stereotypes in news articles within the Netherlands. This study provides new empirical evidence on the nature of ethnic stereotypes in news content by drawing on a sample of more than 3 million Dutch news articles. First, we focused to data preparation, as already mentioned. Then, we performed model training and generated word lists containing words that convey high threat (e.g., art thief, mafia, etc.) or low risk (e.g., drug dealer, alcoholic, etc.). We used the Word2vec implementation. We used Python to test the respective hypotheses and to create our final visualisations. Regarding ethnic stereotypes, word embeddings are particularly suitable, since our question is about recurring word analogies. The collection of 1559 plenary

protocols of the German Bundestag is chosen as input for the analysis, which reflects relevance through weighting. Based on the context, the model learns the meaning of words through their occurrence, which is then used for further work. Words and synonyms are represented as neighbouring points in the vector space. Such neighbouring words in a vector space can be synonyms or words that have a similar topic or context. As a result, fruit and apple, for example, have a common semantic meaning, which is why these two points are embedded or represented as close neighbours in the vector space. Using these word embeddings and this method gives us deeper insights into the ethnic stereotypes in the plenary transcripts. The model training is done in two steps. First comes the whole corpus for the time-invariant hypotheses to simply see the natural form of the ethnic stereotypes and then in the second step the word embeddings from the different years are done to test the time-variant hypotheses. Here, the computer learns each word separately to detect variations. Several steps were taken to ensure the quality of the baseline embedding model. First, we checked that references to target categories in our dataset were frequent > 5,000. This is important in order to define a framework for the extent to which an evaluation is performed. The word analogy task is one of the most popular methods for evaluating the quality of word embeddings. This method is based on the idea that people should be able to predict mathematical operations in a vector space. For the identification of low-status, high-threat associations, which are we want to measure, we need to identify words that relate to this schema. This means we identify words that capture both concepts from the data under study because different language and or jargon is used in politics.

To measure the association of "low status" and "high threat", we need to identify words that reflect these concepts. We extracted these words from an existing study collected in 2020 titled: Associative Guilty: Using Word Embeddings to Measure Racial Stereotypes in News Reports. These words use a bottom-up approach to build word lists by using the most similar results from embeddings trained on a corpus of all news content for each news source. The authors then manually modify the generated word list and classify as follows. Only consider words that have a negative connotation and therefore reflect a certain stereotype of a certain group. Anne C. Kroon 2020 High threat indicators are defined as words associated with hostility, deviance, threatening behavior or objects, criminal and/or illegal activity (eg, theft, robbery, and murder). We also included words related to law enforcement (such as police or police) because we think these words conjure up hostility and crime. Low-status indicators were defined as words associated with low social class, (not) intelligence, low educational attainment, unemployment, addiction, and/or homelessness. Words with a negative connotation that did not fit into one of these categories were not included in the analysis. We excluded issue-specific words related to foreign affairs, specific events or issues, and/or a select group of ethnic categories (such as dictators, child soldiers and radicalism), as we are explicitly interested in the general dimensions of stereotype content. In several rounds, the selection of words was critically discussed and carefully revised by the authors. A target word list was created that included various ethnic categories. We included the eight largest non-Western ethnic groups living in Germany (i.e. Surinamese, Turks, Syrians, Antilleans, Moroccans, Somalis, Afghans and Iraqis). In addition, we used only the Germans themselves as a group within Germany. For the sake of completeness, we also included designations that are frequently used in the plenary minutes of the German Bundestag to designate ethnic categories (e.g. immigrants or foreigners). For each ethnicity and designation, both the singular and plural forms were included in the target word list (e.g.

Moroccans, Moroccans, immigrants, immigrants).

5. Analysis Results (Tim und Jo-Ann)

(Jo-Ann Schüller) For the analysis, we first trained the Word2Vec models to detect the correlations between ethnic groups, which we summarized into a word list, and a second list which includes the labels for the predefined attributes "high threat" and "low status". We retrieved the word lists from the study "Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage," as we believed they were suitable. We had to be translate them into German before we started so that we could use them. For Run 1, we decided to use an existing word list for ethnic groups. However, this also had to be translated into German beforehand, which caused that translation errors occurred and distorted the results. The frequencies of the words were therefore not valid and quite low in the first run. Since words like 'pool' were also included, some population groups were unrepresentative.

```
[ ]  frequencies = compute_frequency(all)
     frequencies

     {'deutsch': 1014288,
      'griechisch': 75521,
      'migrant': 95929,
      'pool': 5111,
      'somali': 8466,
      'syrisch': 56380,
      'turk': 55202}
```
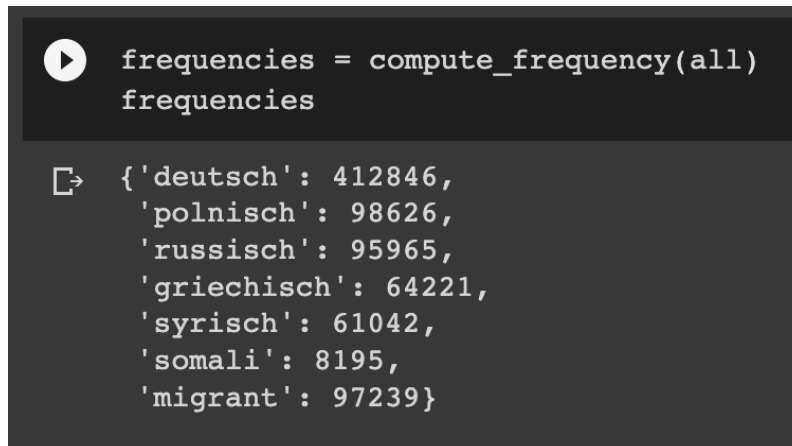
Figure 2: Term frequencies of run 1

	frequencies	scores_lowstatus	scores_highthreat	names:group
0	75521	0.069023	0.048130	griechisch/outgroup
1	95929	0.167606	0.058967	migrant/outgroup
2	56380	0.105655	0.087850	syrisch/outgroup
3	8466	0.161149	0.070071	somali/outgroup
4	55202	0.065094	0.035421	turk/outgroup
5	5111	0.208172	0.100422	pool/outgroup
6	1014288	0.000397	0.008355	deutsch/ingroup

Figure 3: Term Scores of run 1

To correct this error, we created our own word list based on the first list. We also used

current historical events as reference for our choices, such as the refugee crisis or the Greek government-debt crisis. Run 2 with the second word list then produced better results. As you can see in the figure 4, the frequencies have become much higher and thus obviously better.

```
frequencies = compute_frequency(all)
frequencies
```

```
{'deutsch': 412846,
 'polnisch': 98626,
 'russisch': 95965,
 'griechisch': 64221,
 'syrisch': 61042,
 'somali': 8195,
 'migrant': 97239}
```

Figure 4: Term frequencies of run 2

5.1. Hypothesis 1 and 2 (H1 and H2) (Jo-Ann Schüller)

H1: Do politicians implicitly associate ethnic outgroups more strongly with low-status traits than ethnic ingroups? H2: Do politicians implicitly associate ethnic outgroups more strongly with high-threat traits than to ethnic ingroups.

To be able to prove or disprove the above hypotheses, let's look at the first analysis. In figure 5 you can see a scatterplot, which is used to visualize the results. As expected, the scores for the ingroup 'German' (low-status: 0.0073, high-threat: 0.0077) are relatively low, even though it has the highest term-frequency (412,846). The outgroups 'polish' (frequency: 98.626), 'russian' (frequency: 95.965) and 'migrant' (frequency: 97.239) all have a similar high term frequency, but they have big differences when it comes to scoring on attributes. It so happens that 'Polish' has a low-status score of 0.0295 and a high-threat score of 0.0322, thus relatively balanced. With the ethnic group 'Russian' a first big difference is already visible: the low-status-score is 0.0601 and the high-threat-score is 0.0564. Although the terms have an equal frequency, there seem to be more correlations of the negatively tainted attributes to the Russian population group within the protocols. Looking now at the term 'migrant', the difference becomes even more clear: the low-status-score is 0.1685 and the high-threat-score is 0.0571. These terms seem to be correlated very often with negative attributes and suggest that ethnic stereotypes can be seen here.

Also interesting to consider are the remaining outgroups 'Greek', 'Syrian' and 'Somali'. The terms 'Greek' have a frequency of 64.221 and 'Syrian' 61.042. 'Somali' has the lowest frequency of all, namely 8.195. Despite the low occurrence of the term 'Somali', it has the highest score both as low-status (0.1756) and as high-threat (0.0814). Thus, the score for low-status is even more than twice as high as that for the term 'Greek' (low-status: 0.0831). 'Greek' is ranked
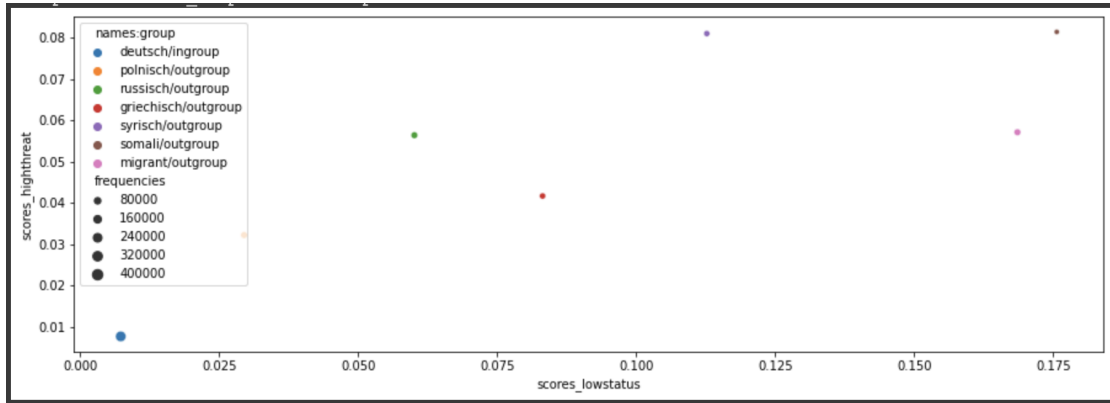
Figure 5: Visualization of term frequencies and scores of run 2

fifth out of seven with a high-threat score of 0.0417. 'Syriac' has a low-status score 0.1126 and a high-threat score of 0.0810. What is very interesting to observe is that 'Syriac' and 'Somali' have by far the highest high-threat score and low-status score.

Low status characteristics are clearly associated with outgroups, which is shown by the high scores. The closer the considered outgroups are to the ingroup 'German', their status improves, which can be observed by the representation of low high-threat and low-status scores. On the other hand, we can observe that the test group 'Migrant' also show scores higher than 0.1 and 0.05 and therefore represent low status. These results suggest that ethnicity is related to attributes that convey low status or threat.

Reasons, why these results occur, could be past historical events. As suggested earlier, past crises could be the cause of the classification. These crises could also be related to the federal elections. Right-wing extremist parties such as the NPD and AfD, have gained votes over the course of the last few years. The AfD even made it into the Bundestag in 2017 and 2021. The NPD has not been able to pass the 5% hurdle, but its share of the vote has increased from 0.4% (2002) to 1.6% (2005), i.e. 4 times as many people voted for the NPD in 2005 (*Stimmenanteile der AfD bei den Bundestagswahlen von 2013 bis 2021* 2021; *Stimmenanteile der NPD bei den Bundestagswahlen von 1965 bis 2021* 2021). Therefore, we cannot disprove H1 and H2, but we have clear evidence here that ethnic outgroups are associated with high-threat and low-status traits than ethnic ingroups.

5.2. Hypothesis 3 (Tim Schenkel)

The third hypothesis was primarily about the time variable predicting implicit stereotype association strength. Here we use the word embeddings obtained for consecutive years from 1994-1998, 1998-2002 and 2002 to 2021. The strength of implicit stereotype association was expected to increase over time for the ethnic outgroup, but not for the ethnic ingroup (H3). We analysed this proposition using the different temporal periods. In the first analysis, we focused on the period from 1994-1998. For this purpose, all ethnic groups were included that had a frequency of >5000 within this period.

Figure 6: Term frequencis from Period 1994-1998

These values were then used to create an ingroup with the German population and 4 outgroups: Polish, Russian, Greek and migrants. Our German ingroup had the lowest values for both word lists in the analysis, i.e. this ethnic group was least associated with the hightreats and lowscores lists. The value for the lowstatus words was -0.015194 for the term German and 0.04638 for the highthreat words. All other ehtnic groups had higher values for both word lists. The highest value for the highthreat word list comes from the ehtnic group Greek with a value of 0.115041. This means that this ethnic group is significantly more negatively loaded with words such as murder in years between 1994-1998 compared to all other groups. The highest value for the low-score words has the ethnic group of migrants with a value of 0.052130. Here, too, the value is significantly higher compared to the value of the ingroup. For this year, we can definitely confirm that the outgroups are associated with significantly more negative words. Now the analysis was mainly about the change over the years. Therefore, we now compare the different periods to see how the individual groups have developed. In the period from 1998 to 2002, a similar development can be observed.

In the second period, we looked at the time frame from 1998-2002. This is shown in Figure 8. Here we have used the same in- and outgroups as in the 1994-1998 time frame. Here the results for individual groups have changed.

Our ingroup German continues to be the lowest with the values for the lowscores at -0.030383 and the highthreat words at 0.033649. For the ethnic group of migrants, however, the value has changed. In the previous period, the value for the highthreat words was 0.065034. This has increased to 0.135040 in the current period. The value has therefore doubled in this period. This means that the negative association of the migrant group has increased significantly within 4 years. The value for the low-score words has also increased from 0.052130 to 0.075316. Thus, for this outgroup, the scores for both word lists have increased. The values for the other outgroups have also changed in the period 1998-2002. The ethnic group of Greeks is also an example of this. Above all, the value of the low-score terms has increased from 0.016017 to 0.051222 compared to the years 1994-1988. While the value for our German ingroup has decreased in
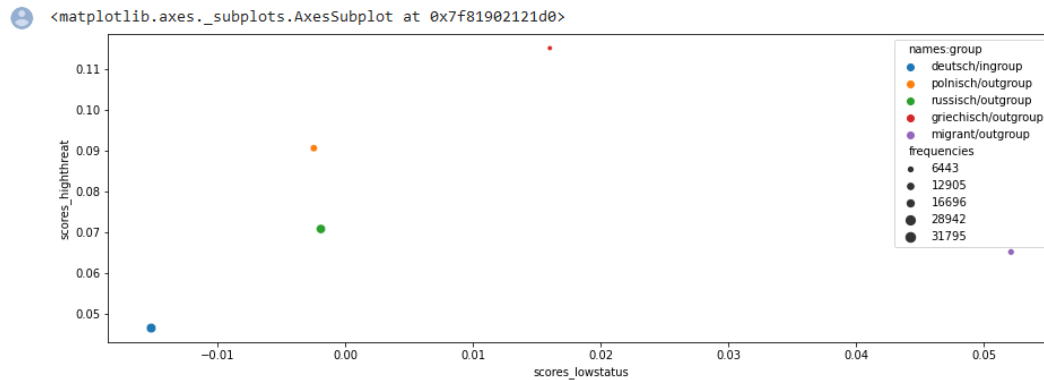
Figure 7: Results of the Time Period between 1994-1998

	frequencies	scores_lowstatus	scores_highthreat	names:group
0	36465	-0.030383	0.033649	deutsch/ingroup
1	7093	0.043221	0.120962	polnisch/outgroup
2	14285	0.023702	0.084765	russisch/outgroup
3	6936	0.051222	0.132389	griechisch/outgroup
4	5377	0.072655	0.135040	migrant/outgroup

Figure 8: Term frequencis from Period 1998-2002

both categories, the value for all outgroups has increased. So far, this confirms our thesis that an increase in ethnic stereotypes in the Bundestag can be seen over time. Finally, the share of ethnic stereotypes in the years 2002-2021 is presented.

In the third period between 2002-2021, the development is somewhat different compared to the previous years. Here, the values for all outgroups have fallen significantly in relation to the highthreat words.

In the previous year, the value for the lowercase words for the ethnic group of migrants was 0.072655. In the new period, this value is 0.135974, which is approximately double the value. This value has also increased for all other outgroups. Over time, the strength of implicit stereotypical associations in plenary documents content will increase for ethnic outgroups, while this will not be the case for ethnic ingroups. Over Time, we can observe from the 3 plots of 3 diffrents periods, that the points represnting outgroups are moving to to right and to the top, period by period which mean that their low status/ high treat scores are increasing over time for outgroups. In the other hand, In groups are reciving almost the same low scores over
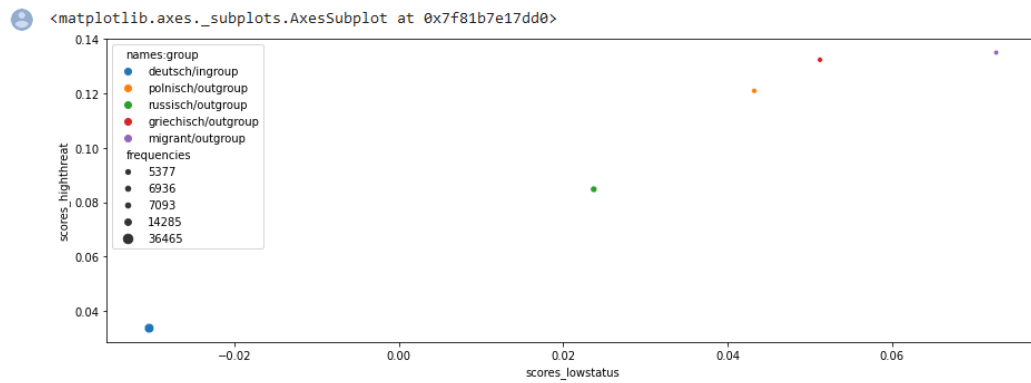
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f81b7e17dd0>
```


Figure 9: Results of the Time Period between 1998-2002

```
[ ]  period3_results
```

	frequencies	scores_lowstatus	scores_highthreat	names:group
0	268595	0.024043	0.012452	deutsch/ingroup
1	48509	0.046561	0.040179	polnisch/outgroup
2	64352	0.038357	0.066498	russisch/outgroup
3	60756	0.056899	0.047263	griechisch/outgroup
4	64790	0.084018	0.086614	syrisch/outgroup
5	236306	0.105968	0.085784	somali/outgroup
6	75697	0.135974	0.072779	migrant/outgroup

Figure 10: Term frequencis from Period 2002-2021

time.

An additional interesting point about this period is the fact that additional outgroups have been added that did not appear in the two previous periods. These include the ethnic minorities from Somalia and Syria. They also exchange with a very high frequency compared to the other outgroups. In addition, the two groups have the highest values for highthreat words within the analysis. Syrians have a value of 0.084018, which is the third highest value in this time interval. The second highest value belongs to the ethnic minority of people from Somalia with a value of 0.105968. The highest value of highthreat words for this time is 0.135974 with the value for migrants in general. As our analysis has shown, we expect stereotyping of plenary protocols about ethnic minorities to continue to increase over time. Marginalised ethnic groups from non-European countries were uniformly rated negatively on both dimensions. In particular,
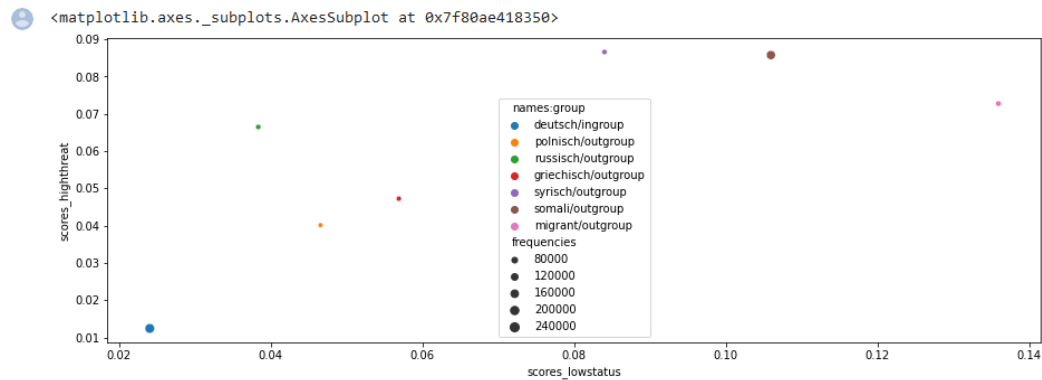
Figure 11: Results of the Time Period between 2002-2021

ethnic subgroups such as Somalis or Syrians could be uniformly placed in the high threat and low status quadrants of stereotype content. This suggests that politicians in the German Bundestag portray immigrant subgroups largely in terms of the prevailing societal perception about them.

6. Next Steps (Jo-Ann Schüller)

Our results could be further elaborated by adding a table which could include the results for the individual years. This procedure would simplify the comparison of these years with historical events. An example of the accumulation of (negatively) charged terms for ethnic minorities in Germany could be the "refugee crisis" in 2015. This subdivision also simplifies the values of terms that appear in connection with nationalities. Creating a list of positive words could also ensure that more accurate scoring is possible. To obtain even more accurate results with our previous approach, all documents could be translated into English beforehand to eliminate word variants. This would lead to better results, since the English language is simpler and generally has fewer words than the German language, because especially word neologisms are not possible in English. However, at the same time this brings another problem: translation errors. These could then distort the results again. Performing a cross-check to compare these results with each other could solve this.

7. Conclusion / Discussion (Jo-Ann Schüller)

In summary, our analysis clearly showed that ethnic minorities are more likely to be associated with negative attributes than ethnic ingroups. However, as noted earlier in the analysis, several factors may have contributed to obtaining these results. On the one hand, historical events are to blame for the increased appearance of a term; on the other hand, perhaps the rise of far-right parties in the Bundestag (i.e. AFD) could also be the reason for the more frequent appearance of ethnic stereotypes. It was easy to see that over time the occurrence of different terms contributed to different scores, making it likely that the same conditions do not apply

to every term and that there must be further external correlations for the occurrence of the attributes. Accordingly, our analysis could not consider whether the attributes always referred to the ethnic outgroup, but the scope of this analysis and the technical possibilities do not quite allow for this.

8. Citations and Bibliographies

References

Anne C. Kroon Damian Trilling, Tamara Raats (2020). *Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage.* url: https://journals.sagepub.com/doi/10.1177/1077699020932304.

Bergmann, Werner (2008). *Was sind Vorurteile?* url: https://www.bpb.de/shop/zeitschriften/izpb/9680/was-sind-vorurteile/.

Grandi, Filippo (2021). *Vorurteile gegen Flüchtlinge auf dem Prüfstand.* url: https://www.uno-fluechtlingshilfe.de/informieren/faktencheck.

Jochen Oltmer, Vera Hanewinkel (2021). *Geschichte der Migration nach und aus Deutschland.* url: https://www.bpb.de/themen/migration-integration/laenderprofile/deutschland/341068/geschichte-der-migration-nach-und-aus-deutschland/#node-content-title-3.

Plenarprotokolle (2022). url: https://www.bundestag.de/dokumente/protokolle/plenarprotokolle.

Stimmenanteile der AfD bei den Bundestagswahlen von 2013 bis 2021 (2021). url: https://de.statista.com/statistik/daten/studie/753961/umfrage/stimmenanteile-der-afd-bei-den-bundestagswahlen/.

Stimmenanteile der NPD bei den Bundestagswahlen von 1965 bis 2021 (2021). url: https://de.statista.com/statistik/daten/studie/368838/umfrage/stimmenanteile-der-npd-bei-den-bundestagswahlen/.