CESifo WORKING PAPERS

10026
2022

October 2022

Immigrant Narratives

Kai Gehring, Joop Adema, Panu Poutvaara

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10026*

Immigrant Narratives

Abstract

Immigration is one of the most divisive political issues in many countries today. Competing narratives, circulated via the media, are crucial in shaping how immigrants' role in society is perceived. We propose a new method combining advanced natural language processing tools with dictionaries to identify sentences containing one or more of seven immigrant narrative themes and assign a sentiment to each of these. Our narrative dataset covers 107,428 newspaper articles from 70 German newspapers over the 2000 to 2019 period. Using 16 human coders to evaluate our method, we find that it clearly outperforms simple word-matching methods and sentiment dictionaries. Empirically, culture narratives are more common than economy-related narratives. Narratives related to work and entrepreneurship are particularly positive, while foreign religion and welfare narratives tend to be negative. We use three distinct events to show how different types of shocks influence narratives, decomposing sentiment shifts into theme-composition and within-theme changes.

JEL-Codes: F220, J150, C810, Z130, D720.

Keywords: narrative economics, immigration, media, newspapers, voting.

Kai Gehring
Wyss Academy
University of Bern / Switzerland
kai.gehring@unibe.ch

Joop Adema
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Munich/Germany
adema@ifo.de

Panu Poutvaara
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Munich/Germany
poutvaara@ifo.de

October 10, 2022
We would like to thank Manfredi Aliberti, Matteo Grigoletto, Jan Gromadzki, Melina Liethmann, Jonathan Öztunc, Charlotte Robert, and Bahar Zafer for valuable research assistance. Furthermore, we would like to thank Thomas Stratmann and the participants of the CEMIR Junior Economist Workshop, the Silvaplana Political Economy Workshop, and the EEA and VfS annual congresses for valuable comments. We acknowledge funding from the SNSF Ambizione grant and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101004703.

1 Introduction

Narratives are simple stories or fragments relating to such stories that do not just transport neutral, objective information, but reflect an emotion, opinion, or sentiment about an actor or issue. Shiller (2017) highlights how individual narratives and their spread can crucially affect personal, economic, and political decisions. Recent studies show that narratives influence a wide range of outcomes from inflation expectations (Andre et al., 2021) to racism (Esposito et al., 2021) and behavior during the COVID-19 pandemic (Bursztyn et al., Forthcoming). In contemporary societies, narratives are reflected, shaped and spread in particular through mass media like TV or newspapers (Ash et al., 2022; Durante, Pinotti and Tesei, 2019; Durante and Zhuravskaya, 2018; Zhuravskaya, Petrova and Enikolopov, 2020).

Competing narratives are particularly important when it comes to immigration. Immigration is one of the most controversial topics that constitutes a key political cleavage in all Western democracies (Bonomi, Gennaioli and Tabellini, 2021; Gethin, Martínez-Toledano and Piketty, 2022). Controversies about immigration have influenced the rise of far-right parties (Otto and Steinhardt, 2014; Halla, Wagner and Zweimüller, 2017; Edo et al., 2019; Hangartner et al., 2019) as well as key events like Brexit or Trump's election as US president (Norris and Inglehart, 2019). Competing narratives claim, for instance, that immigrants are either a burden to the welfare state or needed to save that same welfare state. While existing papers on narratives often focus on single events or specific narratives, our paper provides comprehensive coverage of seven immigrants narrative themes. This enables us to better understand which narratives are most salient, what sentiment they tend to transport, and how they are influenced by societal conditions and key events.

We propose a novel way to study narratives using text-as-data, and apply it to immigrant narratives in newspapers. We focus on Germany as the largest member state of the European Union that has a large and diverse immigrant population and is the main destination country of asylum seekers. Germany also features a rich and diverse landscape of regional newspapers, opening up the possibility to link immigrant narratives to specific local conditions. Our definition of narratives is rather broad, including not only causal statements about the role of immigrants, but also statements characterizing immigrants as an actor or group. The unit of analysis are individual sentences as the fundamental building blocks of longer texts like newspaper articles.

To provide a comprehensive dataset capturing immigrant narratives, we combine more traditional dictionary-based approaches with the possibilities of modern Natural Language Processing (NLP) packages that allow detecting linguistic features like grammar, word types, and dependencies. For each sentence, our method aims to detect (i) whether the sentence is about immigrants; (ii) if it fits into one or more of seven narrative themes that we identify; and (iii) if it has a (theme-specific) negative, neutral, or positive sentiment. Instead of following an unstructured topic-modelling approach, we classify narratives into seven pre-defined themes. Those themes are based on key topics in the economics of immigration literature and our own reading of 500 randomly chosen German newspaper articles about immigrants. They contain the economy-related narratives *Work*, *Welfare*, and *Entrepreneurship*, the society-related narratives *Foreign Religion* and *Cultural Integration*, as well as *Immigrant Criminality*, and *Immigrants-as-Victims*. The last theme differs from the others by focusing on all narratives that depict immigrants as victims, e.g. crimes but also and discrimination against immigrants. [1] Our final sample contains 107,428 articles from five national and 65 regional

[1] For that reason, we exclude this theme when calculating average sentiment.

newspapers about immigrants over the period from 2000 to 2019.

Our main results are the following. First, we analyze the composition of themes in newspaper articles in a balanced sample of newspapers from 2005 to 2019 (newspaper coverage improves substantially starting from 2005). We find that the three economy-related themes *Work*, *Welfare*, and *Entrepreneurship* were the subject of only 12 percent of narrative sentences, *Immigrant Criminality* of 12 percent and *Immigrants-as-Victims* of 7 percent. A vast majority of immigrant narrative sentences relates to the society themes *Foreign Religion* (45 percent) and *Cultural Integration* (23 percent). Second, we find that sentiment differs a lot across themes, with more limited shifts within themes over time. For instance, *Foreign Religion* narratives are mostly negative, whereas *Entrepreneurship* tends to be a very positive theme. Third, our approach allows us to propose a method to decompose sentiment shifts in compositional effects and within-theme shifts. We use this to analyze the effect of three important immigration-related events in recent German history and their influence on immigrant narratives.

Most studies using text-based measures are not able or do not try to provide a systematic assessment of the quality of the measures. Developing specific functions and combining NLP tools with dictionaries for our approach requires effort, hence it is important to evaluate if and which parts of the effort are justified by improved performance. One of the contributions of our paper is to use human-coded articles to investigate both heterogeneity in the human interpretation of narratives and to compare the quality of our method with simpler alternatives. While our dataset can be applied to many research questions, this investigation helps to better understand the general potential of the approach we propose.

For the purpose of validating our approach, we recruited 16 human coders among native German-speaking university students from different parts of Germany and carefully trained them intensively for the task. Each student coded a batch of 437 articles, which equals around 18,000 to 20,000 sentences. We use the sample of articles coded by students to study heterogeneity among human coders and to assess the performance of our algorithm compared to more standard approaches. Our algorithm has an accuracy rate of 96.6%, and clearly outperforms alternatives based on simple keyword matching, providing the best balance between true positives and false negatives. While our new theme-specific dictionaries contribute a lot to the initial classification performance, correct sentiment assignment is particularly improved by the use of specific NLP functionality and our new sentiment-adjustment functions.

We discover several interesting insights regarding heterogeneity in the human understanding of narratives. There is a large heterogeneity across human coders for classification and sentiment assignment, which differs between themes. Complete alignment among humans is an exception. Given the controversial and partisan nature of the immigration discourse this might not sound surprising, but the extent of the heterogeneity is, nonetheless, noteworthy. It highlights a potential problem for supervised machine learning approaches that require a ground truth to learn from. Our algorithm tends to classify sentences as immigrant narratives slightly more conservatively than the human average, but well within the range of the human assessments. If transparency is desirable or other reasons make a supervised machine learning approach infeasible, combining a dictionary-based approach with modern NLP features to exploit linguistic features like grammar thus seems a good option.

Our main contributions are the following. First, we develop new immigration-specific dictionaries and combine them with NLP tools in a novel way that delivers promising results. Second, we provide a large dataset of immigrant narratives in seven themes and their sentiment for German national and regional newspapers. Third, to the best of our knowledge, we are the first to collect a large set of human coders to evaluate the human understanding of narratives and use it as a basis of evaluating the methodology. Fourth, we demonstrate how that dataset can be used to understand both larger trends over time and across space, as well as analyze individual events in detail.

Our paper relates to four different strands of literature. First, it contributes to the strongly growing literature on the economics of narratives, sparked among others by Shiller (2017). Narratives can have a crucial role in shaping human incentives (Bénabou, Falk and Tirole, 2018). Bursztyn et al. (Forthcoming), for instance, find that exposure to one of two distinct narratives about the COVID-19 pandemic had a strong impact on behavior. Esposito et al. (2021) show how a racist narrative spread through a popular movie. In contrast to papers focusing on a single narrative, we can measure narratives comprehensively across several themes and decompose the effect of important events in shifts across themes and within-theme sentiment changes.[2]

Our methodological contribution to the economics of narratives literature is the combination of dictionaries with modern NLP tools that capture grammatical features of words and sentences. By doing so, we create the most comprehensive dataset of immigrant narratives and their sentiment so far. Like Ash, Gauthier and Widmer (2021), we focus on shorter narrative fragments (contained in an individual sentence) that might be part of an over-arching grander narrative. In contrast to their paper, we do not rely on unstructured topic models for theme selection. We also define narratives in a broader way beyond simple cause and effect statements, in line with Shiller (2017).[3] One of our innovations is evaluating the quality of our approach using human coders and quantifying the advantages compared to simpler alternatives.

Second, while behavioral economics and lab experiments highlight the importance of individual-level psychological biases, many important psychological mechanisms operate at the group-level. Narratives are collective, shared stories that can be regarded as a technology both influencing and being influenced by other group-level phenomena. Those include culture (e.g., Guiso, Sapienza and Zingales, 2016), moral values (e.g., Enke, 2020), and group identity (pioneered by Akerlof and Kranton, 2000).[4] Our comprehensive immigrant narrative dataset enables researchers to gain a better understanding of how narratives as one such mechanism correlate with societal conditions or important events.

Third, we relate to the literature on media economics. Access to different types of media can have important effects on a variety of political, economic and personal outcomes (Ash et al., 2020; Bursztyn et al., Forthcoming; Campante, Durante and Sobbrio, 2018; Galletta and Ash, 2019; Kearney and Levine, 2015). Some important papers examine social media specifically (Enikolopov, Petrova and

[2] Narratives concern real people, or at least people claimed to be real in case of fake news. In a related strand of literature, Michalopoulos and Xue (2021) explore a large collection of folklore stories and how they are shaped and linked to societal and environmental conditions. Folklore is defined as traditional stories linked to a specific community and passed through generations by word of mouth. Michalopoulos and Xue (2021) analyze a catalog of folklore stories compiled by anthropologist and folklorist Yuri Berezkin, and show that motifs in a group's oral tradition are strongly correlated with its physical environment. Furthermore, they show that folklore-based measures of historical attitudes help to predict contemporary stereotypes, values, and economic choices.

[3] Given the rapidly evolving nature of the field, it is impossible to give a fair overview of all new developments. There are many exciting developments that integrate machine-learning, for instance (e.g. Card et al., 2022). It is very hard to fully automate the process. For instance, topic models all involve a manual choice on how many topics to select and what the topics proposed by the algorithm represent. With our method, themes are transparently selected before designing the dictionaries.

[4] Economists have started to study the origins of group identity (Dehdari and Gehring, 2022), its influence on political preferences (Fouka, 2019; Gehring, 2021) and how current events can shape it (Depetris-Chauvin, Durante and Campante, 2020; Gehring, 2022).

Zhuravskaya, 2011; Cagé, Hervé and Mazoyer, 2022; Zhuravskaya, Petrova and Enikolopov, 2020), others focus on newspapers (Besley and Burgess, 2002; Gentzkow, Shapiro and Sinkinson, 2011; Snyder and Strömberg, 2010) or the relationship between different types of media (Cagé, Hervé and Viaud, 2020). Newspapers are still a major source of information – often the basis for social media discussions – and regional newspapers allow linking narratives to local characteristics. We augment studies like Couttenier et al. (Forthcoming) and Keita, Renault and Valette (2021), which use a dictionary approach to specifically study coverage of immigrant criminality, by capturing immigrant narratives comprehensively across seven themes.[5]

Fourth, our findings have important implications for research on the effects of immigration on host societies. Economic research on immigration has largely focused on labor market and public finance effects (Scheve and Slaughter, 2001; Borjas, 2003; Facchini and Mayda, 2009; Ottaviano and Peri, 2012; Hatton, 2017; Battisti et al., 2018). A notable exception is Card, Dustmann and Preston (2012), who emphasize the perceived threat from immigrants to compositional amenities like schools. Ethnic diversity has been also found to undermine general support for redistribution (Dahlberg, Edmark and Lundqvist, 2012; Alesina, Miano and Stantcheva, Forthcoming). Alesina and Tabellini (Forthcoming) review the existing literature and conclude that culture and ethnic background play an important role in determining the political reaction to immigrants. Our results highlight the potential role of media in this process: just as voters, the newspapers in our sample emphasize other societal aspects more than economy-related narratives when discussing immigration.

2 Methodology

2.1 Measuring Immigrant Narratives in Newspaper Texts

> [...] narrative to mean a simple story or easily expressed explanation of events that many people want to bring up in conversation or on news or social media [...] (Shiller, 2017)

Based on Shiller's article with this initial definition and related to popular science books like Harari's "Sapiens" (2015), the study of narratives has been gaining attention in the economics literature as well as in the popular press. Since the work by Shiller on narrative economics, both theoretical (Bénabou, Falk and Tirole, 2018) and empirical (Esposito et al., 2021; Bursztyn et al., Forthcoming) papers increasingly use the concept and try to measure narratives. However, there is no universally agreed upon definition of a narrative in economics yet. Some papers define narratives in a narrow sense as causal statements linking a subject to an object (e.g. Ash, Gauthier and Widmer 2021), while others like Shiller suggest a broader understanding. In this paper, we adopt a broader definition of narratives that includes causal statements as one type of narratives but is not restricted to those.

The reason for choosing a broader definition of narratives is based on our aim to study immigrant narratives in (German) newspapers. A qualitative review of actual newspaper articles reveals that narratives about immigrants are of very different types, that could all contribute to shaping readers' perceptions. Newspapers differ from other sources like social media because content is written by professional journalists, who usually avoid using overly emotional language or expressing strong

[5] Djourelova (2022) exploits an abrupt ban of the politically charged term "illegal immigrant" in the dispatches distributed by the Associated Press to show that media slant has a causal effect on individuals' views about immigration policy. The finding that the use of a single politically charged term affects views on immigration policy suggests that accounting for the overall slant in media narratives about immigration may well have important implications on attitudes towards immigrants.

sentiment. Instead, journalists often shape narratives by selecting which news to report and how intensively. They also often use more complex sentence constructions to convey sentiment, rather than just using a strongly polarized term to achieve that purpose. Hence, simple word matching misses many of these subtle, but very common ways to assign a sentiment to a sentence. Our guiding criterion to consider a sentence in an article as a narrative fragment is if it is (i) about immigrants, (ii) fits into a pre-specified narrative theme and (iii) expresses an opinion, emotion or interpretation about the role of immigrants in the host society.

We use seven pre-defined *narrative themes* that correspond to the most salient areas of political controversy and research about immigrants. We prefer this to an automated data-driven topic generation because the most relevant themes within the migration discourse can be well defined based on qualitative research and manual inspection.[6] By pre-defined *narrative themes* we mean the most relevant areas of public discourse about immigrants, defined in a way that they are sufficiently homogeneous internally and easily distinguishable from each other. For instance, the economic impact of immigrants is clearly a highly relevant theme, especially among economists, and can be distinguished clearly from possible linkages between immigrants and crime as another theme. The advantage of defining a few clear themes is their transparency and clear relationship with existing quantitative and qualitative research strands. Automated topic-models provide useful interpretations of generated topics but suffer from the sensitivity to adaptable parameters. We outline and explain our seven themes in more detail below.

Narrative fragments - building on Ash et al. (2020) - are parts of a larger narrative that can be more complex and long, but are often represented through narrative fragments that trigger an association with a grander narrative known to readers. For instance, consider the grander narrative that immigrants are a large burden for the welfare state, because they lack certain skills and are hence unemployed more often. Readers of newspapers are familiar with those grander narratives. The logic of our approach is that simple narrative fragments like "unemployment among immigrants continues to be very high" or "most immigrants lack a formal education" will thus trigger people to think about the grander narrative. It is very rare to observe complete complex narrative or strong specific statements like "immigrants are causing lower wages for natives" in the newspapers we cover, instead such narrative fragments provide such cues to the over-arching narratives. We refer to narrative fragments related to the immigrant narrative themes simply as immigrant narratives in the remainder of the paper.

Technically, three main approaches to measure narratives and their sentiment in texts can be distinguished: manual dictionary methods, unsupervised machine learning (ML), and supervised machine learning. Traditional text-as-data analysis uses dictionaries that can either contain theme-specific words or words that transport a sentiment (Gentzkow, Kelly and Taddy, 2019). The words in those dictionaries are then matched to the text, and based on match frequency a theme and a sentiment can be assigned. Unsupervised learning, like topic models, tries to automatically detect patterns in data, much like cluster analysis. Regarding ML, supervised learning requires human training data, which is then usually processed with neural networks to conduct out-of-sample text classification. ML also empowers new NLP packages (like the Python package *spaCy*) that can auto-

[6] Different approaches are useful for different purposes, conditional on the underlying data and concept. Unsupervised learning in the form of topic models allows to descriptively understand clusters in the data, which is the more useful the less is known about a topic or the less specific the existing qualitative literature or political discourse. Pure word-matching approaches based on existing dictionaries are efficient if the goal is a simple general sentiment assessment in arbitrary sentences that correlates strongly enough with a latent concept so that in can be used in a regression framework.

matically process sentences and provide linguistic features such as dependencies between the words composing that sentence.

All methods exhibit advantages and disadvantages, summarized for instance in Osnabrügge, Ash and Morelli (2021). Simple word-matching requires effort to define the dictionaries, but those dictionaries are then transparent and allow easy replication by other researchers on other text sources. However, as we will demonstrate in detail, word-matching alone fails in many instances and causes both biased measurement and measurement error. Supervised ML requires a definition of target criteria (e.g. themes) and a sufficient amount of manually coded training material. It is able to capture more nuances of texts that are lost or misinterpreted by simply counting word-matches. The resulting predictions, however, are based on a black-box-process. This is more problematic when there is large and incompletely understood heterogeneity among humans as it requires a plausible definition of a ground truth.

When approaching the task of measuring immigrant narratives, our own initial manual coding of a substantial amount of articles and sentences suggests large heterogeneity among humans in classifying a sentence and assigning a sentiment. The lack of agreement on the "correct" assessment among humans is one argument against using supervised ML in this case. The second reason is that while supervised ML clearly excels at many tasks, we want to ensure that there is transparency about the criteria behind a prediction. One reason is that pre-testing revealed that some sentences were assessed quite differently. That is why we opt in favor of a more simple approach that combines the creation of theme-specific dictionaries with NLP tools and self-built functions using one leading ML-based NLP tools. Specifically, we use the Python package *spaCy* (Honnibal et al., 2020), which allows us to extract linguistic features such as dependencies and word-types.[7]

This allows us to (i) classify a much larger share of sentences correctly while maintaining a low false selection rate, (ii) cope with linguistic complexities like negation, qualifying (e.g. a positive or negative aspects improves or worsens), and (iii) assign theme-specific sentiment much more accurately than simple dictionary methods. Figure 1 summarizes the custom dictionaries, NLP tools, and self-coded functions that we use. In strongly simplified terms, we identify narratives about immigrants in German newspapers in following steps:

1. Identify articles about immigrants in Germany by using a comprehensive filter on the *Factiva* database and further filtering using a German location dictionary (see section 3.1). In the following, we refer to these as articles about immigrants in Germany.

2. Classify sentences as containing immigrant narratives based on keywords from custom dictionaries and NLP tools (see section 2.2.2). In the following, we refer to these as sentences containing immigrant narratives or simply classified sentences.

3. Categorize immigrant narrative sentences into seven narrative themes using dictionaries and NLP tools. In the following, we refer to these as classified into a theme.

4. Assign a theme-specific negative, neutral, or positive sentiment for each sentence when the sentence is classified into the theme using dictionaries, NLP tools, and self-constructed evaluator/negation/qualification functions (see section 2.2.3).

[7] These include features such as dependency parsing, Part-Of-Speech (POS) tagging, morphology, lemmatization, parse trees, and Named Entity Recognition (NER). For an introduction into these and other features developed in the computational linguistics literature, we refer the reader to `https://web.stanford.edu/~jurafsky/slp3/`

Figure 1: Dictionaries, NLP tools, and Custom Functions Used for the Algorithm

- **A - Custom dictionaries** are manually compiled word-lists. These include bi-grams and tri-grams, gendered singular and plural versions of each word, as well as German compound words. Bi-grams and tri-grams refer to combinations of two or three words that appear in a fixed order, e.g. "gut integriert" ("well integrated"). Compound words are extremely common in German, for instance "Ausländerkriminalität" ("foreigner criminality"). Those dictionaries are based on human reading of a random sample of 500 articles, plus additional selective articles for specific dictionaries.

  Examples of dictionaries are common immigrant terms word-lists, foreign names word-list, and foreign nationalities word-list used in section 2.2.2, as well as theme-specific word-lists, evaluators, negation, and qualifiers used in section 2.2.3.

- **B - Python NLP tools**
  - **B1 - spaCy NLP tools** are linguistic features from the spaCy Python package (Honnibal et al., 2020). These include lemmatization, Part-Of-Speech (POS) tagging, dependency parsing, and named entity recognition (NER).
  - **B2 - Coreferee Pronoun function** builds upon the Coreferee Python package (Hudson, 2022). It relates pronouns to the most likely nouns they refer to within and across sentences.

- **C - Custom sentiment-adjustment functions**
  - **C1 - Evaluator function**: adjusts the sentiment if its meaning is modified by contextual words from theme-specific evaluation dictionaries, e.g., *find/lose jobs*, *contribute/burden to the social security system*.
  - **C2 - Negation function**: combines negation dictionary with NLP tools to reverse the sentiment of a negated sentence or adjust it to neutral. Based on more than 100 relevant sample sentences, we validated with three human coders that reversing or adjusting brings the sentiment closer to human judgement.
  - **C3 - Qualifier function**: combines the qualifier dictionary with NLP tools to adjust the sentiment linked to a theme dictionary word by combining qualifier words (e.g., *increase*, *decrease*, *big*, *small*) with NLP tools. Based on 100 relevant sample sentences, we validated with three human coders that reversing or adjusting brings the sentiment closer to human judgement.

2.2 Examples and Detailed Process

2.2.1 Challenges to Classify Sentences Correctly

A few examples help to illustrate how we approach the challenge of capturing narratives in newspapers using the dictionaries, NLP tools, and functions outlined in Figure 1, and to understand some of the benefits in comparison to alternatives. First, there are sentences that could be recognized as describing immigrants by simply matching keywords, and as positive or negative by featuring a negative or positive word. However, it turns out that such sentences are quite rare and sentiment is often contained not in emotional words but in telling a certain narrative. Moreover, negation can reverse the meaning of a sentence if not recognized. The following two examples show such cases:

"Viele **Ausländer** würden sich gar **nicht integrieren** wollen."(*Die Welt*, November 26th 2004)
TRANSLATION: Many foreigners would not even want to integrate (into society).
"Es sind **nicht** alle **Muslime bildungsfern**."(*Berliner Morgenpost*, September 2nd 2010)
TRANSLATION: Not all Muslims are uneducated.

Such sentences would be recognized to be about immigrants by simply matching with a dictionary of immigrant terms ("Ausländer/foreigners, Muslime/Muslims"). These are assigned to *Cultural Integration* as they mention integration and a lack of education. SentiWS, a commonly used German sentiment dictionary, assigns a positive sentiment to the first sentence (as the word "integrieren" has a positive sentiment) and fails to assign a sentiment to the second sentence (the composite word "bildungsfern" has a negative meaning, but is not included in SentiWS). The manual dictionary-based approach ensures that "bildungsfern" is recognized as a negative term. The negation tool ensures that the first sentence is classified as negative instead of positive, and the second sentence as neutral instead of negative.

The majority of sentences in (German) newspapers turn out to be more complex and require more sophisticated methods. One such method we use is the pronoun function. It allows us to trace back the origin of personal pronouns to prior sentences, to classify correctly whether they are about immigrants. One example how this improves sentence recognition is:

"Die meisten Migranten kommen aus Rumänien und Bulgarien. Oft kommen sie **bettelarm** und **ohne** **Bildungsabschluss** an." (*Metzinger Uracher Volksblatt*, July 11th, 2016)
TRANSLATION: Most migrants come from Romania and Bulgaria. They often arrive in bitter poverty and without an educational degree.

In such common sentence constructions, the actual narrative appears in the second (or subsequent) sentence, but the first sentence is necessary to classify it as being about immigrants. The *pronoun function* captures most of those cases. The theme-classification of the sentence rests on the words describing extreme poverty ("bettelarm") and (lack of a) formal education. Again, the negation function ensures that the negative connotation is captured for the second term. There are of course also many more complex cases that require combining dictionaries with NLP functions like:

"„Jugendliche **Straftäter** mit **Migrationshintergrund** kommen meistens aus **Familien**, die nicht **integrationswillig** sind“, so die Erfahrungen des **Polizisten**." (*Süddeutsche Zeitung*, March 8th, 2013)
TRANSLATION: "Juvenile offenders with a migration background usually come from families that are not willing to integrate," according to the police officer's experience.

NLP tools, specifically the dependency parsing, allow us to detect that the offender is linked to an immigration background. Moreover, it allows us to link those to a lacking willingness to integrate. This sentence is assigned to two themes, *Immigrant Criminality* due to the mentioning of criminal offenders with migration background, and to *Cultural Integration* due to mentioning families with a lacking willingness to integrate.

In some cases, we hard-code certain terms that are used very frequently and allow distinguishing sentences that would otherwise be regarded as identical by simple word matching. By hard-coding we mean linking theme-specific terms with specific terms that signal a crucial difference and affect whether a sentence is assigned to one theme or another. Take the final example:

"Seit der Zuzug von **Flüchtlingen** nach Deutschland massiv zugenommen hat, steigt auch die Zahl der **Straftaten gegen Asylunterkünfte und Ausländer** dramatisch an." (*Allgemeine Zeitung Mainz*, February 23rd, 2016).
TRANSLATION: Since the inflow of refugees to Germany has massively increased, the numbers of

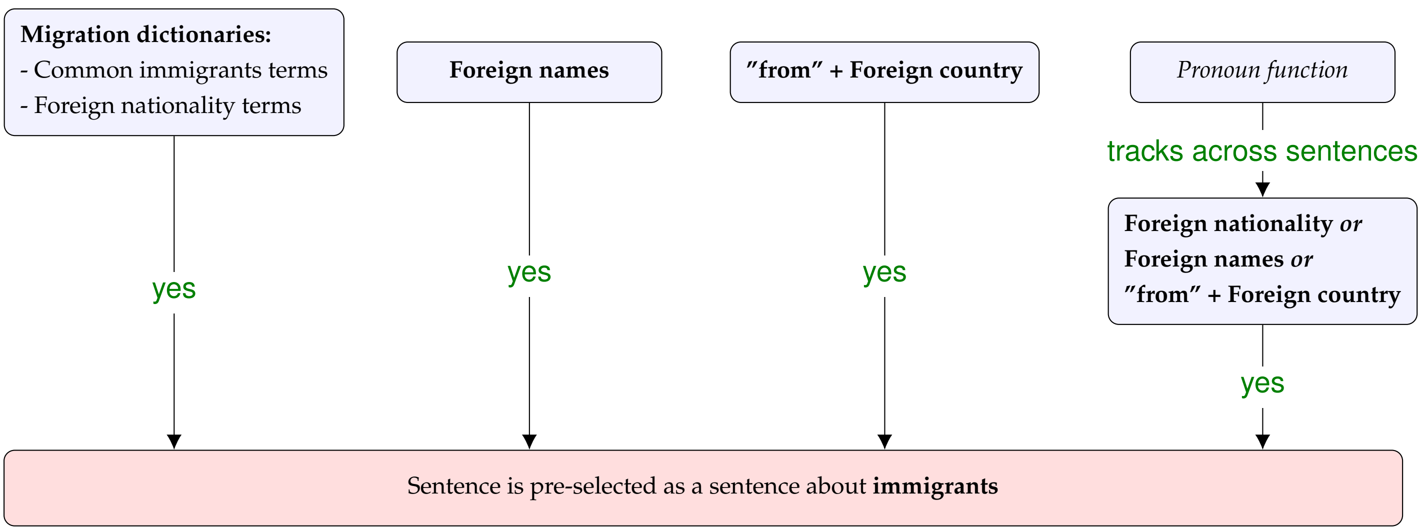
crimes against asylum centers and foreigners also have increased dramatically.

Recognizing that this sentence is about refugees is easy, and could be achieved by having a simple immigrant term dictionary that contains the word. Assigning it is about crime is also relatively straightforward, as it contains the word for ”criminal offenses” (”Straftaten”). However, it turns out that there is a distinct type of narratives that are indicating actions or attitudes against immigrants, instead of portraying the actions of immigrants. Our algorithm detects not only whether there is a crime and an immigrant term in a sentence, but also if the signal term ”gegen” (”against”) appears and whether dependency parsing links it directly to the immigrant term. If it does, as is the case here, the sentence is assigned to *Immigrants-as-Victims* instead of *Immigrant Criminality*. While those examples provide an idea of the challenges faced to measure immigrant narrative in specific cases, we now describe more generally how our methodology approaches theme classification and sentiment assignment.

2.2.2 Identifying Immigrant Sentences

To capture narrative fragments at the sentence-level, the first general step in methodology is to select sentences relating to immigrants in Germany. As outlined above, the first step is to use locations to filter at the article level which articles are about immigrants and about Germany. Conditional on that simple initial filtering, we then use dictionary-based selection procedures to identify immigrant sentences (see Figure 2). The dictionaries include (i) common terms that refer to immigrants; (ii) foreign nationalities; (iii) names with foreign origins signaling a migrant background.[8] Several NLP tools are used to help with the foreign origin identification, with distinguishing whether someone is ”in” or ”from” a country, or to trace personal pronouns across sentences.[9]

Figure 2: Immigrant Sentence Selector



[8] We also include all dual nationalities where on of the nationalities is German such as *Turk-German*, *German-French*, etc.

[9] Foreign names are identified using *spaCy*'s *entity recognition* tool to identify *person entities* in the sentence. If these entities do not intersect with dictionaries of German last and first names from *www.behindthename.com*, they are considered as foreign names. To avoid a biased assessment, we manually inspect the 1000 most frequently matched names and exclude publicly known foreign individuals like politicians, music or sport stars, as well as German politicians with a foreign name.

2.2.3 Classifying into Themes and Assigning Sentiment

We classify sentences in seven themes: *Work*, *Welfare*, and *Entrepreneurship* all relate to the role of immigrants in the the *Economy*, *Immigrant Criminality* to immigrants as perpetrators of crimes, *Cultural Integration* and *Foreign Religion* to immigrants' role in *Society*. *Immigrants-as-Victims* is a separate theme that captures narratives about Immigrants as Victims of crimes, discrimination, or other actions.[10] Figure 3 and 4 illustrate the process of assigning themes and theme-specific sentiments to sentences. A sentence can contain more than one immigrant narrative theme, and we can compute a theme-specific and overall sentiment. If not stated otherwise, our results will use the overall sentiment, but can be decomposed by themes.

The figures simplify the actual algorithm, but illustrate how the different dictionaries and tools from Figure 1 are combined for classification and sentiment assignment. A sentence is identified to be about immigrants either by a combination of the immigrant sentence selector or a foreign religion term with a theme-specific term or, in fewer cases, by a single term that encapsulate both the immigrant and theme aspects. An example for such a single term would be German compound words: "Ausländerkriminalität" ("crimes by foreigners") reflects both crime and immigrants in one word.

Based on the extensive manual reading and coding of sentences, we construct one theme that can be understood as orthogonal to the others. While sentences for immigrant schemes normally describe actions or behavior of immigrants as part of the host society, there is a considerable number of sentences that describe how the host society is treating immigrants. These can include crimes, with the distinction that immigrants are not the perpetrator, but rather the victim of a crime, unlike in the *Immigrant Criminality* theme. To capture who is the victim and who the perpetrator we use various NLP features, most importantly the Part-of-Speech (POS)-tags that allow recognizing subject and object in a sentence. In addition, we classify sentences that include explicit anti-immigrant terms or discrimination towards immigrants into the theme *Immigrants-as-Victims*. Note that our approach here is rather conservative, as our aim is to identify clear cases and avoid judging ambiguous sentences.

[10] The discrimination dictionary in Figure 4 contains terms like "discrimination" that need an additional immigrant term to clarify that the discrimination is directed against immigrants, and not other groups. In contrast, the anti-immigrant dictionary contains terms that directly signal an action, policy or regulation against immigrants (e.g. German compound words like "Fremdenfeindlichkeit").

Figure 3: Themes: Work, Welfare, Entrepreneurship, Foreign Religion, Cultural Integration

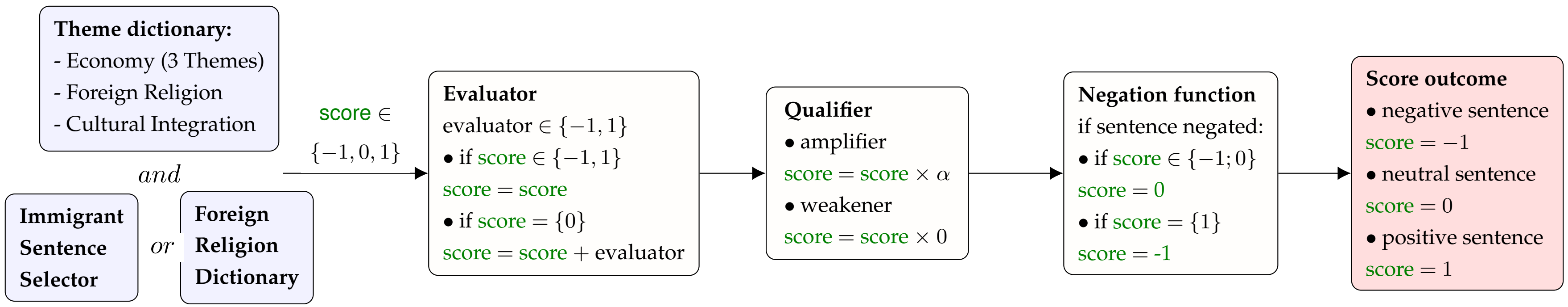
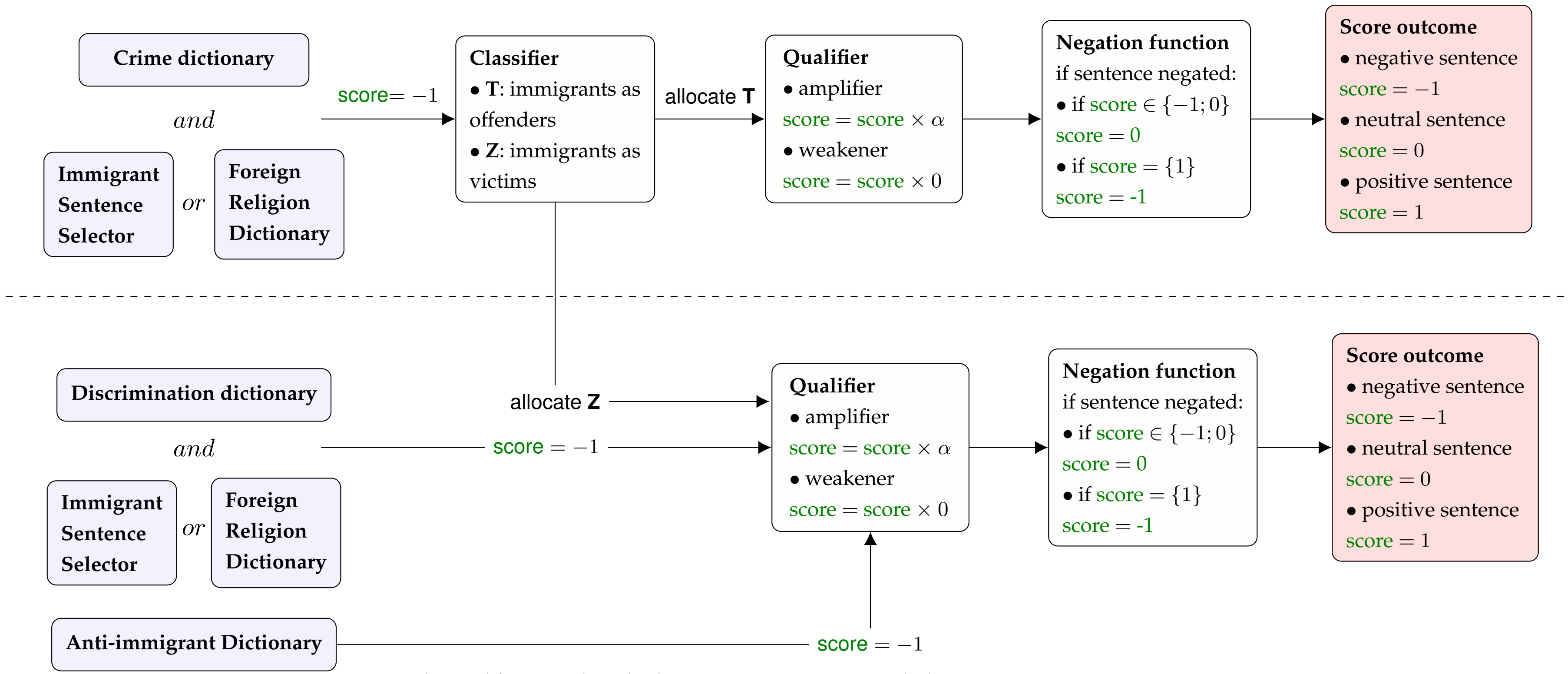


Figure 4: Immigrant Criminality (upper part) and Immigrants-as-Victims (lower part)



The amplifier can only make the sentiment more extreme, which remains an integer: $\alpha > 1$.

3 Data

3.1 Newspaper Articles: Factiva

In Germany, local and regional newspapers have an important place in the media landscape. Local and regional newspapers are being at least occasionally read by around two thirds of the population (GLES, 2019). We obtain individual articles about immigrants published by German national and regional newspapers from Factiva, an international newspaper database. Figure 5 outlines the steps taken to obtain relevant articles. We begin by querying articles from Factiva using a Boolean search filter (logical AND, OR, and NOT operators) that combines immigrant-specific search terms with a geographic location within Germany.[11]

To have a sufficient number of articles per newspaper, we obtain up to 3,000 (6,000) newspaper articles for regional (national) newspapers from Factiva.[12]

We develop an approach based on Entity Recognition and lists of foreign and German locations to further ensure that an article is concerned with immigrants in Germany. For details on this approach, see Appendix A. Overall, we obtain a dataset comprising 107,428 newspaper articles from 70 newspapers.[13]

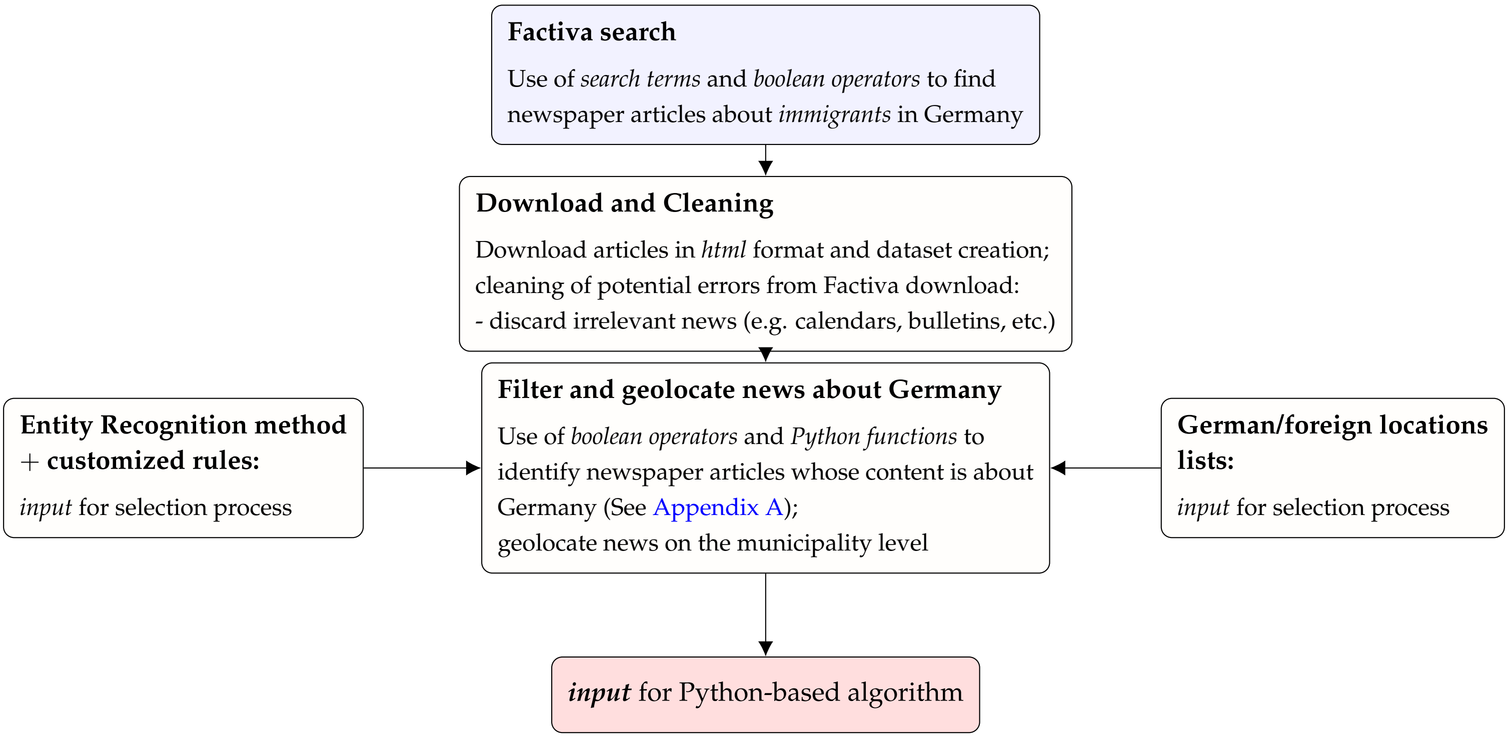
Table B.1 provides an overview of the scope, total number of articles, share of articles with at least one immigrant narrative sentence, as well as with the share of articles obtained from a press agency such as Reuters or the German Press Agency (DPA). On average, more than 80 percent of all articles contain at least one sentence containing one of the themes and 14 percent of all articles are written by press agencies. Hence our simple filters seem to work quite well, and the share of original (non-press agency) articles seems high enough to provide also spatial variation across newspapers.

[11] The filter includes all words starting with: Einwanderer (immigrant), eingewandert (immigrated), Migra (as in migrant or migration), or Ausländer (foreigner). Location terms are Germany, all 16 German federal states (German: *Bundesländer*), the 50 biggest cities in Germany and "Ruhrgebiet", the term for a large conglomerate of cities in the West of Germany, home of many immigrants. The filter excludes certain terms that seem to create many false positives, like event-calendars, bulletins, or sport results. We also omit articles with less than 100 words and those that are most likely about immigrants outside Germany, using the count of non-German places that are found in an article. For the full filter, see Appendix A.

[12] The Factiva subscription limits article search and download and does not allow systematic webscraping. Hence, we manually download up to 1,500 articles per newspaper, which is the threshold to which the subscription limits availability for many newspapers. Factiva sorts articles according to relevance, and the same algorithm is applied consistently to each newspaper. To increase the number of articles, we divide our sample period into sub-periods for these manual downloads. For each national newspapers, we subdivided the period 2000-2020 in 4 periods, and for each regional newspaper in 2 periods. This leads to a maximum of 6000 articles per national newspaper and 3000 per regional newspaper. The only exception to this is the *Weser Kurier* from Bremen, as its weekend edition was included in Factiva under a separate name.

[13] We include 5 national (Süddeutsche Zeitung, Die Welt, Der Spiegel (weekly), Die Zeit (weekly), and BILD) and 65 regional German newspapers, listed in the appendix. Although Factiva contains identifiers for 89 newspapers, those do not all represent individual newspapers with independent editorial offices. We collected information about editorial offices from the official websites to identify the editorial offices.

Figure 5: Steps to Construct Dataset



3.2 Newspaper Sales and Background Information: IVW

We use municipality-level data by the German Audit Bureau of Circulation (IVW), an independent auditing organization that records and certifies German newspaper sales and circulation data.[14] Beyond the data on the distribution zones (areas served by a regional newspaper), we reached a bilateral agreement to get access to municipality-level sales data (*Gemeinde*) for the year 2019, covering 98 percent of German municipalities.[15] We were able to match most of the Factiva newspapers to the IVW data, using a combination of automated matching with manual inspection or research in case of different spelling or complicated ownership structures.[16] We thus need to define what constitutes an independent newspaper and estimate market shares for all years based on 2019 data. While there were a few changes like mergers and acquisitions, the geographical reach of specific regional newspapers was not affected.

3.3 Newspaper Coverage across Space and over Time

Figure 6 gives an overview of the geographical coverage of regional newspapers in our Factiva sample over time. Coverage is rather limited in the first five years from 2000-2004, which is why we show those years only in few selected applications over time. Coverage in the west is usually quite high already in 2005, but initial coverage is limited in the Eastern states of Saxony and Thuringia. Geographical coverage of regional newspapers substantially improves over the years: while 47 percent of municipalities were covered in 2005, over 66 percent of municipalities are covered from 2012 on-

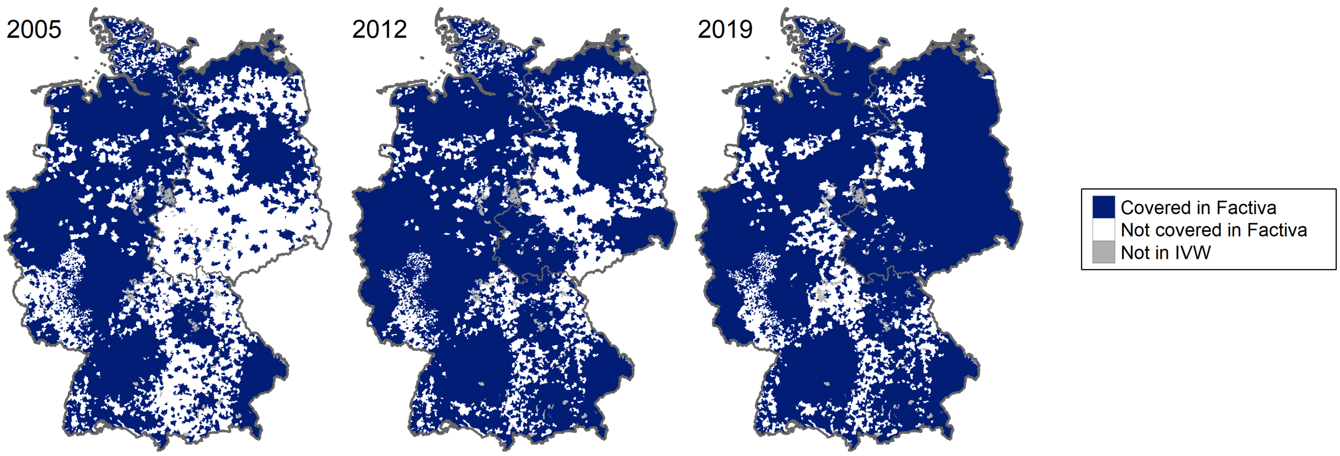
[14] Participating publishing houses report quarterly sales and distribution data to the IVW following common guidelines; the data is then audited by IVW employees before being published.

[15] IVW measured sales in the reference week of 4th to 10th of November 2019. This includes subscriptions and single copy sales. It includes electronic sales only for few selected newspapers.

[16] Der Spiegel, Die Zeit, Bayerische Gemeinde Zeitung, and Bayerische Staatszeitung do not submit (municipal-level) sales data to IVW and we thus do not know their geographical coverage.

ward. If we include national newspapers, over 90% of municipalities are covered from 2012 onward. Appendix Figure B.2 shows that the overall number of articles about immigrants in our sample declines somewhat between 2006 and 2012, and increases strongly again afterwards. Appendix Figure B.3 displays the coverage for all individual newspapers over time.

Figure 6: Geographical Coverage of Regional Newspapers in Factiva in 2005, 2012 and 2019



The maps above show the municipalities for which our Factiva dataset contains articles from regional newspapers in the years 2005, 2012, and 2019. Our dataset covers 47 percent of German municipalities in 2005, 66 percent in 2012, and 73 percent in 2019. Those municipalities house 86, 92, and 93 percent of population, respectively in 2005, 2012, and 2019. IVW does not have newspaper circulation data for the municipalities colored in grey.

3.4 Other Data

We obtain data about local characteristics as well as developments over time from a variety of sources, explained in detail in the appendix. To link newspapers to local characteristics, we first retrieve municipal-level German AGS-identifier codes from merging our IVW dataset with an official administrative dataset. This also allows us to depict the newspapers and our narratives spatially on a map. It also allows us to use any other official administrative or other data source that includes a local identifier. Those include, among others, data from the German statistical office about the migrant share, unemployment, or the share of Muslims in a municipality.[17]

4 Theme Composition and Sentiment of Immigrant Narratives

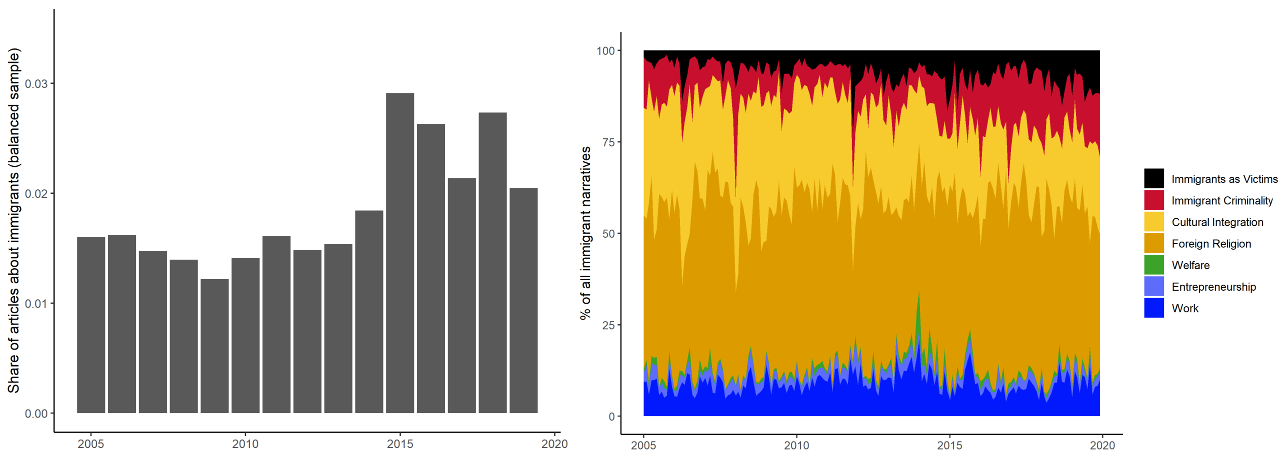
Our new dataset provides a detailed understanding of the prevalence, themes, and sentiment of immigrant narratives in Germany. Germany is a particularly interesting case, offering large heterogeneity across space due to its federal system and Cold War history as well as variation over time due to a large number of relevant shocks in its recent history. As in other Western countries, the population share of immigrants has increased substantially in recent decades and immigration is a controversial political issue.

[17] Appendix Figure B.1 shows the share of foreigners by nationality in 2019, indicating the 16 federal states and the largest cities. Across Germany, 13 percent of population has a foreign nationality. As foreigners may have acquired German nationality, this is slightly lower than the share of foreign-born population. Furthermore, it does not include second (and higher) generation immigrants. The share of foreigners is higher in the former Western Germany than in the East, and higher in urban than in rural areas.

4.1 Salience of Immigration and Theme Composition over Time

The first major descriptive question we answer is about the prevalence of immigrant narratives in German newspapers, measured by the share of articles that are about immigrants. We focus on a balanced sample of newspapers from 2005 to 2019 to avoid changes in newspaper coverage driving changes in the estimated salience of immigration.[18] Figure 7a shows the share of articles about immigrants by year. The salience of immigration remains relatively flat from 2005 to 2014, averaging at 1.5 percent, and increases considerably in 2015 when almost 3 percent of articles printed in German newspapers were at least partly concerned with immigrants. The share of articles related to immigrants somewhat declined in subsequent years, averaging 2.4 percent from 2016 to 2019.

Figure 7: Salience of Immigration and Narratives by Theme between 2005 and 2019



(a) Share of articles about immigrants

(b) Share of narratives by theme

Panel (a) shows the share of all articles in Factiva that are about immigrants in Germany, by year. To ensure comparability over time, this figure is constructed using only 24 newspapers that are available in Factiva since 2005.
Panel (b) shows the share of immigrant narratives about any of the seven themes over time, for a balanced panel of 24 newspapers. The total number of immigrant narratives is the sum of all narratives identified. As sentences can contain more than one narrative, this may exceed the number of sentences carrying narratives.

Our second major descriptive question is how the shares of immigrant narratives by theme develop over time, again focusing on a balanced panel of newspapers. To study the relative prevalence of the different narrative themes over time, we sum up the number of narratives within a theme and divide by the sum of all narratives identified. When doing this, a given sentence can be classified into more than one theme. For example, a sentence about unemployment among young Muslims would be classified to be about both *Work* and *Foreign Religion*, while a sentence about unemployment among Turkish immigrants would be counted to be only about *Work*. Although newspaper narratives are not equivalent to the spread of narratives among the population, they provide a useful way to assess this. The composition and sentiment of a sentence is influenced by many factors that include the journalist's private preferences as well as – given the profit-orientation of newspapers – the preferences of their readers.

[18] There are six newspapers already available in 2000, including two national newspapers (Der Spiegel and the Süddeutsche Zeitung), but the majority is added later to Factiva. We restrict our analysis to the period starting in 2005 in order to keep a balanced sample with 24 newspapers.

Figure 7b shows the composition of immigrant narratives by month.[19] The results are quite striking. While economists usually highlight the economic implications of immigrants for the labor market or welfare state, media coverage focuses much more on *Foreign Religion* and *Cultural Integration*. *Economy* narratives are relevant, but at a clearly smaller scale than the societal themes and with considerable fluctuations. Over the entire 15 years, out of all sentence-level narratives identified in our sample, 12 percent concerned the *Economy*, 45 percent *Foreign Religion*, 23 percent *Cultural Integration*, 12 percent *Immigrant Criminality* and 7 percent *Immigrants-as-Victims*.

4.2 Narrative Sentiment by Theme over Time

Our third major descriptive question is how average sentiment and theme-specific sentiment change over time. Every sentence that is classified into a particular theme is also assigned an integer sentiment (see section 2). To capture whether sentence-level sentiments related to the themes change over time, we first truncate the sentence-level sentiment to {-1,0,1}, to only capture the extensive margin of sentiment. Thereafter, we compute the average sentiment for all sentences assigned to a specific theme (or several themes). For some themes, like *Cultural Integration*, it is a priori unclear whether overall sentiment can be expected to be positive or negative. For other themes, like crime, it seems obvious that the sentiment should be negative. For all themes, how sentiment has changed over time is an open question.

Figure 8 shows the monthly aggregated average sentiments as well as the average theme-specific sentiments. The average aggregated sentiment takes into account *Work*, *Welfare*, *Entrepreneurship*, *Foreign Religion*, *Cultural Integration*, and *Immigrant Criminality*. We do not include narratives about *Immigrants-as-Victims* in this aggregation as this is about how recipient society treats immigrants, rather than narrative about immigrants. From the year 2005 to 2014, the average sentiment is close to zero, ranging from -0.08 in year 2005 to 0.02 in year 2011. From 2014 onward, the average annual sentiment ranges between -0.20 and -0.12, reaching as its lowest monthly value -0.49 in December 2016 (during a series of Islamic State terrorist attacks).

Next, we consider theme-specific sentiment. We begin with the *Economy* sub-themes *Work*, *Entrepreneurship*, and *Welfare*. The average sentiment related to *Work* and *Entrepreneurship* is always positive, and that related to welfare always negative. Average weighted sentiment related to *Economy* is predominantly positive.[20]

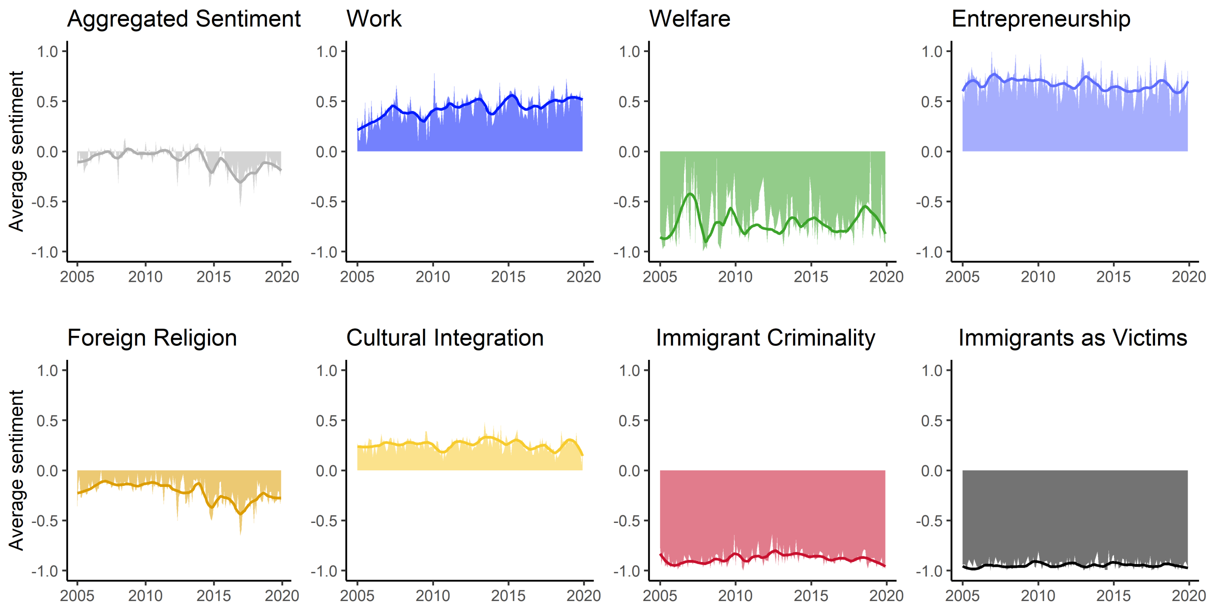
There is an intriguing difference in sentiments related to *Foreign Religion* and aspects of *Cultural Integration*. Sentiments related to *Foreign Religion* are mostly negative. At the annual level, average sentiment related to *Foreign Religion* between 2005 and 2019 was -0.20. The average annual sentiment related to *Foreign Religion* between 2005 and 2014 ranged between -0.22 and -0.11. From 2015 to 2019, the average sentiment related to foreign religion ranged between -0.47 and -0.32. The average sentiment related to *Cultural Integration*, instead, is positive in every month, with average annual sentiment between 2005 and 2019 being 0.24. Sentiment in sentences related to *Immigrant Criminality*

[19] We also analyzed themes over time separately for three national and 21 regional newspapers. As shown in Appendix Figure C.2, themes develop quite similarly in national and regional newspapers over time. For completeness, we show the theme shares between 2000 and 2019 for all 70 in Appendix Figure C.3. Theme shares are also comparable to Figure 7b Moreover, when considering the sub-sample of six newspapers that are present in our sample since 2000 in Appendix Figure C.4, the development of theme shares looks qualitatively similar. Furthermore, we analyzed themes by weighting each sentence by the sales of the newspaper it appeared in. Appendix Figure C.5 shows that the resulting pattern of themes over times is near indistinguishable from that shown in Figure 7b.

[20] The only months in which the average *Economy* sentiment is negative are December 2013 and January 2014, which is attributable to an increased salience of free mobility of labor for Bulgarians and Romanians since 2014.

and *Immigrants-as-Victims* narratives are almost always negative, with average annual sentiment for *Immigrant Criminality* ranging between -0.92 and -0.77, and for *Immigrants-as-Victims* between -0.95 and -0.86.

Figure 8: Theme-specific Sentiments between 2005 and 2019



This Figure shows the average aggregated sentiment per sentence about immigrants (upper left panel) and the average sentiment per sentence about a particular theme (the other panels) between 2005 and 2019, for a balanced sample of 24 newspapers, by month. After sentiments are assigned by the algorithm, we truncate the sentiments at -1 and 1, such that every sentence about a particular theme has a theme-specific sentiment of -1, 0, or 1. To calculate the aggregated sentiment, we calculate the simple sum of all theme-specific sentiments and truncate it again at -1 and 1, such that each sentence is negative, neutral, or positive. To calculate the average sentiment over a large collection of newspaper articles, we calculate unweighted averages. To guide the eye, each of the plots is supplied with a locally smoothened regression line (Loess) with $\alpha = 0.15$.

5 Human Validation

Our goal is to provide a dataset that accurately reflects immigrant narratives in German newspapers on a granular spatial and temporal level. A more limited goal would be to construct a measure that correlates positively with the underlying latent concepts and can be used in a regression framework, but is not reliable and precise enough to provide interesting descriptive insights. Simple keyword-matching of short lists of topical terms or specific dictionaries can be sufficient for such a more limited application. However, such methods may not be able to capture (the sentiment of) narratives over time and space.[21]

Before studying the findings of our method in depth, it is crucial to validate the performance of our approach and show that it performs better than existing approaches. As we are interested in narratives categorized in themes as humans perceive them, the only feasible way to evaluate our algorithm is using human evaluators as a comparison. We proceed in the following way:

1. We recruited and instructed human coders to classify each sentence of a random sample of

[21] As an example, a narrative may consist of new expressions which are interpreted based on the context. By employing theme-specific dictionaries with human-assigned sentiments we are able to accurately capture a narrative's theme and tone.

articles according to whether it is about immigrant narratives, and if so then assign themes and sentiments. Human coders were provided a clear and transparent codebook and a random sample of articles in our dataset. The human-coded articles serve as a benchmark to estimate the performance of our algorithm.

2. We study the share of sentences classified to contain immigrant narratives, theme assignment, and agreement of our algorithm and human coders.

3. We isolate the main features of our algorithm to understand their marginal contribution to the overall ability of the code to capture narratives.

4. We compare our algorithm with simpler alternatives with regard to classification and sentiment assignment.

5.1 Data Collection Process

To evaluate human heterogeneity in understanding narratives and to serve as a basis for assessing the performance of our algorithm, we recruited 24 students from five German universities to classify and assign sentiment to overall more than 2,000 articles. To understand the extent to which there is heterogeneity in human perception of narratives and sentiments, we divided the articles over six batches of 437 articles each. Hence, each article and sentence was assigned to four coders. Because not all students completed the task, our subsequent analysis is based on four batches coded by four human coders.[22] Coders received a spreadsheet with the articles split up in sentences. For every sentence, coders were asked to (i) identify whether the sentence is about immigrants, (ii) if so to assign it to one or more of the themes, and (iii) assign a sentiment of -1, 0 or 1.

The quality of the human input is crucial to provide a reliable assessment of heterogeneity and to assess the performance of our algorithm. To ensure a high quality, we organized an intensive preparation workshop, distributed a detailed codebook, and conducted automated quality checks. Day 1 consisted of joint instruction, questions on the codebook, and random in-class checks to test if instructions were well understood. Every student then completed practice sentences at home, which were discussed in detail on day 2. Throughout the coding process, we were available for clarification questions but never tell that there is one "correct" answer for a sentence in the test set. The ex-post verification suggests that no coder used simple random patterns and that there was some fatigue, but with negligible impact.[23]

In the following subsections, we study heterogeneity among human coders in how they classified sentences, and assess the performance of our algorithm compared to more standard approaches using human coding as a benchmark.

The first question we are interested in is to which extent there are differences among humans in classifying and interpreting narratives. This is an important step that other studies usually ignore. Given the intensive preparation, codebook, and quality checks, we find it plausible to interpret differences across students as real differences in the understanding and interpretation of narratives.

[22] The students were remunerated at a net hourly wage of €15. 20 out of 24 initial participants completed the task. Four batches were coded by all four human coders, which comprise the dataset for subsequent analysis. 14 out of these 16 participants also provide personal background information. Out of the 14 students, seven are female, seven are male, and their age ranges from 19 to 27, with a mean of 22.4. Three students have a migration background.

[23] The codebook and further instructions can be found in Appendix E. Correlation coefficients of classification rates between any two coders coding the same batch ranged from 0.35 to 0.68, suggesting that no coder randomly coded sentences but also that agreement between humans is limited.

While it is difficult to precisely assess the exact extent of differences in understanding a given sentence, we take several steps to minimize measurement error stemming from faulty or sloppy coding, as outlined above.

The second and main purpose is to use the human coding data as a basis for assessing the performance of our algorithm. Performance consists broadly of two aspects. The first aspect is the accuracy of the immigrant narrative sentence classification. The second aspect is whether the algorithm's sentiment assignment aligns with that of human coders, conditional on classification as an immigrant narrative sentence. Our aims are to (i) assess if performance is good enough to be reliably used in further analysis, (ii) which features of our algorithm contribute how much to classification performance, and (iii) how our algorithm performs relative to existing simpler alternatives.

5.2 Classification of Immigrant Sentences and Theme Assignment

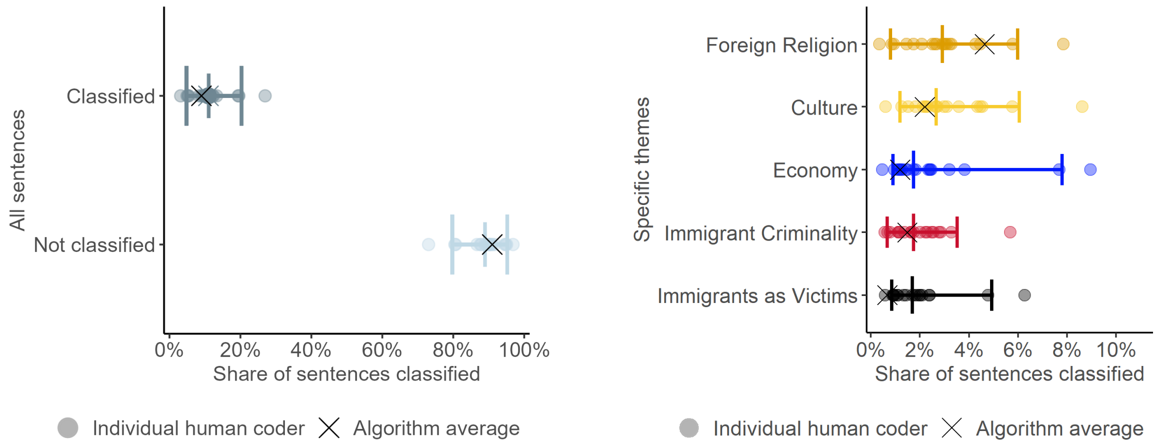
Figure 9a focuses on the classification of sentences by human coders and our algorithm, and Figure 9b on the assignment to a specific theme. The dots indicate the distribution of the share of sentences classified to contain immigrant narratives by human coders, the middle vertical line indicates the coder average, and the cross indicates the share of sentences classified by our algorithm to be about immigrants.

Figure 9a shows that coders classify on average 11.1 percent of all sentences as containing immigrant narratives, compared with 9.0 percent by the algorithm. However, there is considerable variation among coders which we indicate with a 90% confidence interval. Overall quality of our algorithm depends on both recognizing immigrant sentences correctly, as well as on correctly detecting non-immigrant sentences, which are the majority of sentences. Hence, we were rather conservative in the construction of the dictionaries. When it was unclear if a particular term added more true positives than false positives, we did not include it in the dictionaries. The fact that the algorithm turns out to be slightly more conservative, but well within the range of human disagreement is thus reassuring.

Looking at classification by theme in Figure 9b, we see some interesting differences in human heterogeneity and differences compared to the algorithm. Human heterogeneity is the largest for *Economy* and lowest for *Immigrant Criminality*. A small number of coders tend to classify a particularly high share of sentences, shifting the human mean upwards. The algorithm tends to classify below the mean, but much closer to the median of human coders. *Foreign Religion* is the only theme for which the algorithm classifies closer to the upper end of the human distribution, and *Economy* and *Immigrants-as-Victims* the ones where it classifies at the lower end. Reasons for differences across themes could depend on the ease of capturing the theme with a limited dictionary and the complexity of theme-specific narratives. The results are in line with our expectations: religion is rather easy to define unambiguously because what constitutes a religions term is easy to define. In contrast, *Economy* and *Immigrants-as-Victims* are very broad themes where our approach can only cover a certain, but sizeable, share of all sentences.

So how should we interpret the heterogeneity among human coders? It is striking how large differences across humans are. This would make it very difficult to train a machine learning algorithm, because it is not obvious what the objective truth is. At the same time, it is highly plausible. Immigration is a controversial topic, often linked to political preferences and ideology. Differences in sentiment assignment did arise already when creating the dictionaries and algorithm among the authors of this paper and the initial research assistants. Nonetheless, it is surprising that such a large

Figure 9: Human Coders and Algorithm Classification Rates



(a) Classification of migrant sentences

(b) Theme assignment

Panel (a): The graph shows the distribution of the share of sentences containing immigrant narratives, across human coders and our algorithm. A sentence contains an immigrant narrative when it is assigned to at least one immigrant narrative theme. Each dot represents a human coder and the crosses represent the average of the algorithm over all sentences used for human validation. The average classification rate by human coders is reported with the 5th and 95th percentile of the estimated distribution. The distribution is estimated using a widely used estimator for approximating a distribution from observational data using piecewise linear interpolation, described as type 7 in Hyndman and Fan (1996). **Panel (b)**: The graph shows the classification rate into themes: a sentence is classified into a theme if it is about immigrants and fits into one of the pre-defined themes. The average classification rate of human coders is reported with the estimated 5th and 95th percentile, as in **Panel (a)**. A sentence can be classified into more than one theme.

heterogeneity was already visible in the classification step, rather than in the sentiment assignment. Some of it can be measurement error (i.e. human sloppiness), but we have no reason to believe this is the major cause of differences.

Even with intensive training and a detailed codebook, there is large heterogeneity among human coders in classifying sentences as containing immigrant narratives and assigning sentiment.[24] This heterogeneity suggests that to create a comprehensive immigrant narrative dataset, a transparent approach that combines dictionaries with NLP features has some advantages over ML. Our algorithm provides an assessment that is very close to the median human in overall classification. For individual themes, it tends to be conservative compared to human classification for all except one theme. This suggests that while it might miss some narrative fragments, it provides an assessment of narratives that is quite close to how the median human would interpret a sentence.

As the next step, we further use the human-coded data to assess the quality of our algorithm and its individual parts. Figure 10a uses the following simple logic. The more humans agree on a narrative, the less subjective the narrative. The less subjective the narrative, the better our "objective" algorithm should be able to classify a sentence. This is indeed the case. The share of sentences classified by our algorithm as immigrant narratives increases strictly in the share of human coders agreeing on classifying it that way.

In the following, we use sentences with full consensus among humans as a reference to assess the quality of our algorithm. This means we take all sentences that no coder classified as containing an immigrant narrative as a proxy for a false positive, and the sentences all four coders classify as

[24] Heterogeneity is larger for theme assignment than for classifying if a sentence contains any immigrant narrative. It is also larger for sentiment assignment than for classification, potentially because sentiment assignment might be influenced more by personal preferences. One reason for substantial heterogeneity is that sentences that very obviously signal strong sentiment are rare in professional journalism, which avoids strong emotions and value judgements.

immigrant narratives as a proxy for a true positive.

Figure 10b then uses the sentences all four coders and our algorithm classify as immigrant narratives to assess the importance of specific features of the immigrant-sentence-selector and the theme-specific dictionaries and approaches. We find that about half of the classifications by our algorithm are based on general immigrant dictionaries in combination with the NLP tools and our own functions (e.g., qualifier and negator). 28.9% originate from the simple immigrant term dictionary, 13.2% from our nationalities and foreign country dictionary, and 7% from on the foreign names dictionary. Almost half the classifications that are in line with consensus among human coders require the theme-specific dictionaries and algorithm features. The pronoun-function specifically adds another 2.8%. This highlights the contribution of collecting the specific dictionaries, especially the theme-specific ones, even for simply classifying whether a sentence contains an immigrant narrative.

Performance in Comparison with Simpler Alternatives

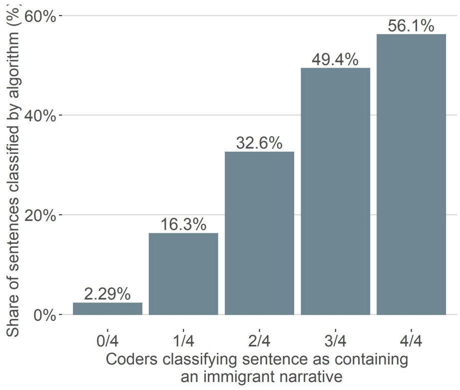
So far, we compared the algorithm to human classification and assessed the contribution of specific features of the algorithm for correct classification. The results suggest a very good classification performance of the code, and an important contribution of, in particular, the newly created theme-specific dictionaries. However, it is important to note that while the features that we assessed are based on dictionaries, the algorithm does much more than simply matching words. It is important to understand if the additional effort invested in combining dictionaries with NLP tools and designing our own functions adds much additional value. Even though the resulting dataset from this specific effort can be widely used, understanding this added value is an important general insight for researchers considering alternative methods to analyze text.

Figure 10c evaluates the performance of our algorithm compared to an approach that would rely solely on simple word matching. We compare our full algorithm to two alternatives: matching using only the immigrant dictionaries from the immigrant sentence selector and matching combining the immigrant dictionaries with at least one term from our theme-specific dictionaries. The overall accuracy rate measures what share of all sentences is correctly classified as containing an immigrant narrative or not. We use sentences that all or no humans classify as immigrant narratives as proxies for true and false positives. Overall accuracy depends on both true and false positives, so those two measures provide a more nuanced explanation for differences in overall accuracy.

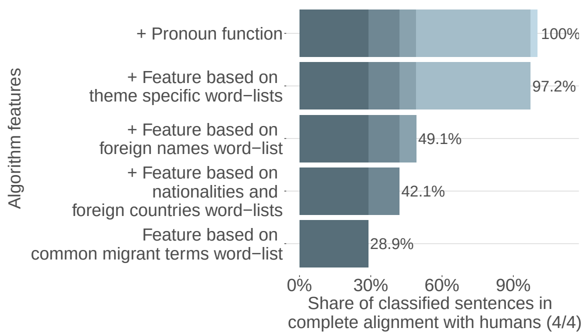
Some aspects stand out in Figure 10c. First, the overall accuracy rate of our algorithm is 96.6%. This is marginally higher than simple matching with all dictionaries, and about 10% higher than matching with only immigrant dictionaries. Using all dictionaries, however, under-performs by more than 10% for the sentences that all humans classify compared with our algorithm, at the same time as it also under-performs by classifying a higher fraction of sentences that none of the humans classifies. Relying on immigrant dictionaries without theme dictionaries performs badly in terms of avoiding false positives: it classifies about 11.9% of sentences that not a single human considers to contain immigrant narratives.

We next evaluate how large differences are in sentiment evaluation when comparing our algorithm to simpler alternatives, conditional on being classified as an immigrant narrative sentence. The main alternative that has been used in prior papers are so-called sentiment dictionaries that are based on human annotation and sometimes machine learning. The terms in the dictionaries are assigned a positive or negative sentiment score and then matched to the words in a sentence. We compare our results to the two most common German dictionaries: *SentimentWortschatz (SentiWS)* (Remus,

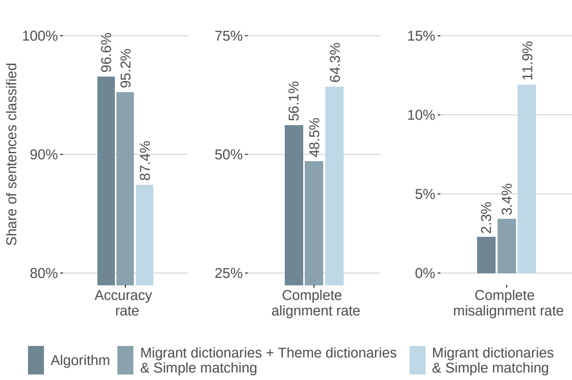
Figure 10: Performance of Algorithm



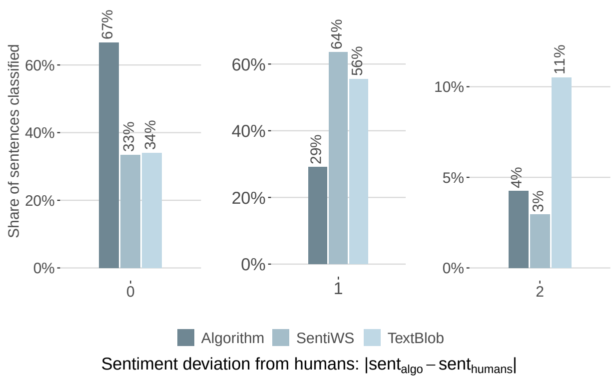
(a) Algorithm performance as a function of human alignment



(b) Marginal contribution to classification of each algorithmic feature



(c) Classification performance compared to alternatives



(d) Sentiment assignment performance compared to alternatives

Panel (a): Each bar represents the share of sentences classified by our algorithm to contain immigrant narratives conditional on how many human coders also classified the sentence. Each sentence is evaluated by four human coders. If there would be perfect agreement among human coders, all sentences would be classified by either 0 or 4 human coders. 2.29 percent of sentences not classified by any human get classified by the algorithm to be about immigrants. 56.1 percent of sentences classified to contain immigrant narratives by all four human coders also get classified by our algorithm.
Panel (b): Each bar shows the marginal contribution of each additional element of the algorithm in classifying sentences for which human coders and the algorithm are in complete alignment. In other words, the sentence is classified by all four human coders and our algorithm. The bottom bar shows the share of sentences that gets classified to contain immigrant narratives when only using the most basic feature of the algorithm that relies on a common immigrant terms dictionary. The bar on top of it shows the share of additional sentences that gets classified when adding the algorithm's feature based on nationalities and foreign countries dictionaries to the previous feature. The other bars are to be understood with the same logic.
Panel (c): The graph shows the accuracy of the algorithm compared to simpler alternative methods, in terms of agreement with unanimous evaluations by human coders. Therefore, we only consider the sample of sentences classified by either zero or four human coders (60,395 sentences). The bar chart on the left shows the share of sentences for which the algorithm (dark blue), simple matching with theme assignment (medium blue) and simple matching (light blue) are in line with human coders at classifying or not classifying a sentence. It is the probability that a sentence gets correctly classified by the algorithm (simple matching) when human coders all agree with each other. The second (third) bar chart restricts the attention to sentences that all (none of the) humans classify to be about immigrants and shows the share of sentences that get classified by the algorithm (dark blue), by simple matching with theme assignment (medium blue), and simple matching (light blue).
Panel (d): The graph shows how the sentiments assigned by the algorithm and alternative sentiment assignment methods differ from the sentiment assigned by human coders. The sample of sentences is restricted to the 1,030 sentences that get classified by all human coders and by the algorithm, and on which all human coders agree on the sentiment. Human coders all agree on the sentiment for 56 percent of sentences that get unanimously classified by all human coders. Sentiment deviation is measured as the absolute difference between the sentiment assigned by the algorithm (alternative sentiment assignment methods) and the consensual human sentiment. A deviation of 0 indicates perfect alignment between the algorithm (alternative sentiment assignment methods) and the human coders, and a deviation of 2 indicates diametrical opposition between the sentiment assigned by the algorithm (alternative sentiment assignment methods) and human coders.

Quasthoff and Heyer, 2010) and *Textblob* (Loria, 2018). *SentiWS* lists positive and negative polarity-bearing words weighted within the interval of [-1; 1].[25] *TextBlob* is a publicly available Python library that includes sentiment polarity scoring in German.

Figure 10d reveals striking differences between our algorithm and the two alternatives. For comparison, we assign to each sentence a positive, neutral or negative sentiment based on the average sentiment across theme-specific sentiment in our algorithm and across all matched words for the sentiment dictionaries. We evaluate – using all sentences where humans agree on classification – whether the assignment by a method was (a) in line with human assessment, (b) differs by one notch, e.g. neutral instead of negative, or (c) differs by two notches, e.g. positive instead of negative. Our algorithm captures sentiment correctly in about two thirds of cases, and for the remainder mostly deviates by one notch. Both alternatives fully align with humans only in about one third of cases, and deviate by at least one notch for the remainder. *TextBlob* even deviates by two notches for 11% of all sentences.

There are several potential reasons why our algorithm clearly outperforms the alternatives. First, it needs to be noted that sentiment can be quite topic-specific. Our algorithm is specifically tailored to immigrant narratives, while existing dictionaries try to capture general sentiment. Second, actual sentences in newspapers are often quite long and complex, and the NLP tools help to assess whether a sentiment-carrying word is really associated with immigrants within a sentence. Third, there might be more nuanced ways to convey a positive or negative picture of immigrants that are not linked to strong sentiment words. Fourth, our evaluator, qualifier and negation functions aim to adjust sentiment if required.

6 Immigrant Narratives over Time

6.1 Descriptive Evidence from a Balanced Sample over the 2013-2019 period

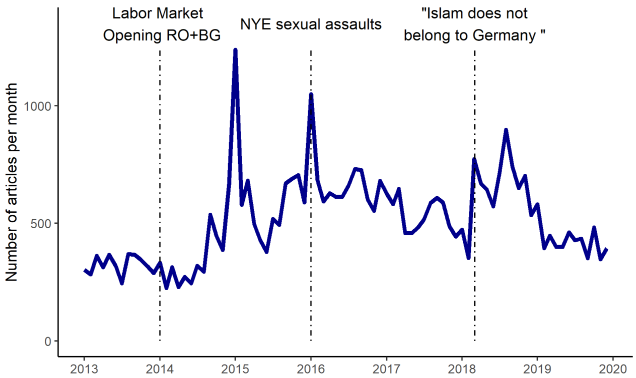
This section aims to provide some examples of how our new dataset can be used to better understand the creation and spread of narratives. For that purpose, we focus on three distinctive events in recent German history that plausibly have the potential to affect the perception of immigrants. To assess the impact of events over time, we create a balanced panel of 41 newspapers (2 national daily and 39 regional newspapers). We restrict our sample to the years from 2013 to 2019 because this allows us to have a broader and representative spatial newspaper coverage. Those years include sufficiently many relevant and interesting events that provide a good starting point for further analysis.

Figure 11a shows the development of the number of articles about immigrants in Germany for this balanced sample. The vertical lines indicate three events that relate to "shocks" in different areas ranging from economy over crime to religion.[26] First, the opening of labor markets for Romanians and Bulgarians is associated with the increased inflow of largely (but not solely) lower-skilled workers, and hence potentially related specifically to *Work* and *Welfare* narratives. Second, the incidence during the 2016 New Year's Eve celebrations in Cologne, featuring a large number of sexual harassment cases attributed to young male refugees from Arab countries might plausibly be related to *Immigrant Criminality* narratives in particular. The third and final event is an (in-)famous statement

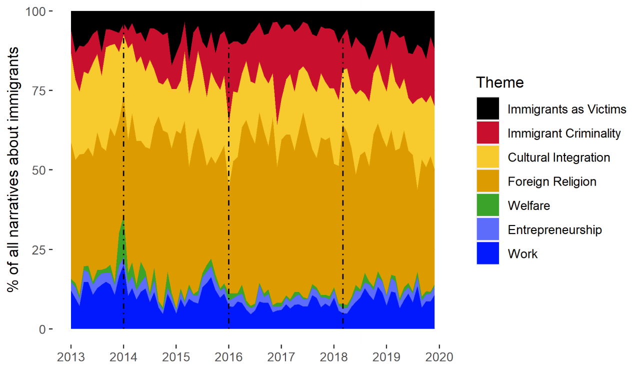
[25] The version of *SentiWS* used in this paper (v1.8b) contains 1,650 positive and 1,818 negative words, which gives 15,649 positive and 15,632 negative word forms including their inflections, respectively.

[26] The number of articles is highest in January 2015. However, there is no singular obvious event linked to this, but rather the combination of German domestic discussions originating from the terrorist attacks on Charlie Hebdo in France combined with the rise of the anti-Islam movement PEGIDA and the right-wing AFD party.

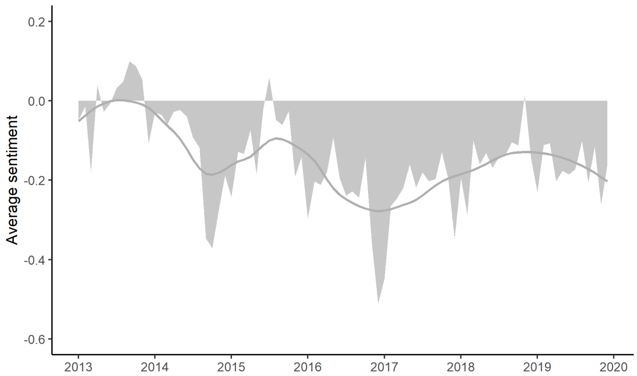
Figure 11: Balanced Panel of 41 Newspapers between 2013 and 2019



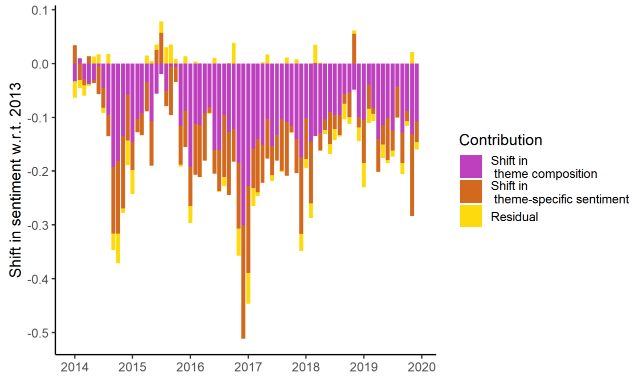
(a) Number of articles



(b) Theme shares



(c) Aggregated sentiment



(d) Decomposition of the shift in sentiment

Salience of articles about immigrants, relative prevalence of narrative themes, overall sentiment, and decomposition of sentiment between 2013 and 2019 for all 41 newspapers included in Factiva from January 1, 2013 onward, by month. **Panel (a)**: The number of articles about immigrants in Germany by month is a measure of the relative prevalence of immigrants in the printed media. The dot-dashed lines indicate three distinctive events further discussed in the next section. **Panel (b)**: The relative prevalence of narrative themes. **Panel (c)**: Average aggregate sentence-level sentiment and the locally smoothened trend. The Loess parameter $\alpha = 0.4$. **Panel (d)**: Decomposition of average aggregated sentiment between 2014 and 2019 with respect to 2013. In particular, we consider what extent of changes (which have been predominantly negative after 2013) can be attributed to changes in which narrative themes are salient and to changes in sentiments of the separate narrative themes. Appendix Appendix D discusses this procedure in detail.

by the back-then German minister of the interior that Islam does not constitute a part of German culture, which might relate in particular to *Foreign Religion* narratives.

Before investigating those events in more detail, it is helpful to examine the data and our possibilities in analyzing it for this balanced sample. Figure 11b shows the composition of immigrant narratives across themes over time, using a monthly aggregation of all newspapers in the sample. The overall picture shows that over that period *Foreign Religion* and *Cultural Integration* also dominate quantitatively, followed by the *Economy* themes. Some interesting longer-term developments are visible over time, though. The prevalence of *Immigrant Criminality* increases strongly, possibly related to the refugee crisis episode starting in 2015. An interesting observation is that while the refugee inflow declined strongly after 2017, the *Immigrant Criminality* remains persistently higher.[27]

Comparing *Foreign Religion* with Cultural Integration, it seems that if anything religion becomes more prevalent compared to general issues of *Cultural Integration*. It is important to note that there is an overlap between both categories, and the algorithm would assign a discussion of integration challenges related to Islamic faith to the *Foreign Religion* theme. Hence, the relative differences do

[27] We cannot rule out that this is related to an actual increase in the number of crimes committed by immigrants.

not indicate necessarily that *Cultural Integration* narratives were less relevant, but rather that that discussion probably centered more around integration barriers related to faith.

The *Economy* themes remain relevant throughout, but with a share between 15 and 20 % they are far from dominating. There is an increase around the initial months of the refugee crisis, but the share is overall higher in the earlier years. This suggests that narratives about the economic impact of the sudden inflow of a large number of refugees are relevant, but the change is rather short-lived and overall dominated by the change in other themes. This reinforces the notion from studying the theme composition over the full sample period that *Economy* matter, but economists tend to over-estimate their importance compared to other themes.

The *Immigrants-as-Victims* themes, capturing narratives about discrimination and crimes against immigrants, tend to be the smallest share of all themes on average. However, there are considerable spikes that are likely to be related to specific events. Moreover, there is a clear increase in reporting *Immigrants-as-Victims* narratives beyond individual spikes in the last two years in our sample. This might reflect an increased sensitivity about these issues and the role of host society and could also provide an interesting avenue for future research.

Figure 11c depicts changes in average sentiment over time. We can observe that sentiment became more negative over time, but with considerable variation. This variation seems broadly in line with specific movements and events during that brief period of recent German history. While average sentiment is important, a crucial advantage of our narrative dataset is that we can infer the underlying mechanisms behind such aggregate changes. Figure 11d provides a decomposition of the changes in sentiment in changes caused by (i) shifts in the share of narrative themes and (ii) shifts of sentiment within narrative-themes. Shifts in the composition of immigrant narratives can affect sentiment because the average sentiment differs strongly between themes, as was shown above. Shifts within themes indicate that journalists report differently, temporarily or permanently, about a particular immigrant narrative theme.[28] Overall, changes in the composition are on average most important, but there also seem to be considerable changes in theme-specific sentiments, especially for *Foreign Religion*, as visible in Figure 8.

As Factiva records which section of the newspaper an article is placed in for more than 60% of articles, we can study how themes and sentiments develop in different sections of the newspaper. In Appendix C.2 we show that the composition of narrative themes over time in sections labeled "news" and "politics" are comparable, but very different from the "local" sections. News and politics sections show many pronounced peaks, like in November 2012 for *Immigrants-as-Victims* after charges for murder were pressed on five members of far-right terrorist organization NSU and *Welfare* in politics sections in December 2014 in connection with the opening of the labor markets for Romanians and Bulgarians. The average shares of different themes are quite similar, although *Foreign Religion* is somewhat more pronounced in politics (47 percent) than in news (40 percent), while *Cultural Integration*, *Work* and *Immigrants-as-Victims* are slightly more pronounced in news.

In local news articles, *Cultural Integration* is much more important than in news and politics sections, and *Immigrant Criminality* and *Immigrants-as-Victims* receive much less coverage. The variability of theme shares over time is much lower for the local sections than for news and politics. Furthermore, the aggregated sentiment for "local" sections alternates somehow, but hovers around zero. In the news and politics sections, instead, aggregated sentiment is most of the time clearly

[28] All changes and the decomposition are relative to a reference period that can be flexibly defined. In this case, the full year of 2013 serves as the reference period.

negative. Also, the theme-specific sentiment for *Cultural Integration* is more positive and for *Foreign Religion* less negative in local news.

6.2 Event Studies

In this section, we focus on the three salient events concerning immigrants in Germany: (i) the opening of labor markets for Romanians and Bulgarians, (ii) the incidents during the 2016 New Year's Eve celebrations in the German city of Cologne, and (iii) the statement by the back-then conservative German minister of the interior that Islam is not a part of German culture.

For each event, captured in one of the columns of Figure 12, we focus on five types of results. First, to put things into perspective, we use the number of articles containing at least one immigrant narrative to proxy for the general salience of the topic. Second, the composition of immigrant narrative themes shows the relative prevalence of narrative themes at a particular point in time. Third and fourth, we show the frequency of narrative fragments from two of the themes that we deemed a priori most likely to be linked to the specific event. Fifth, we decompose changes in sentiment relative to the fortnight before the event.

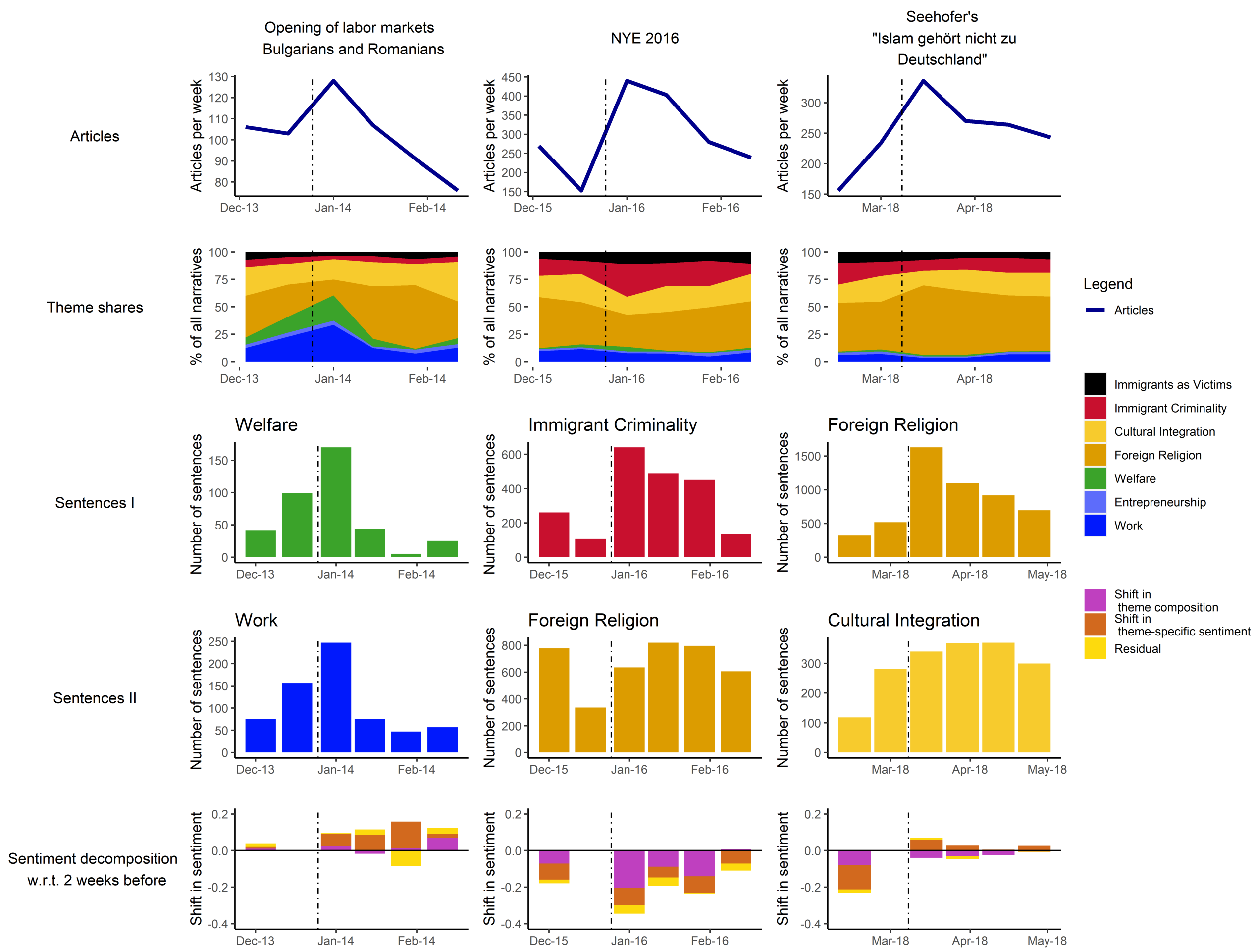
The events differ in their type and nature. The labor market opening is a planned event and thus anticipated. Nonetheless, the policy change plausibly creates attention for the topic and the actual effect of the opening – e.g. the extent of immigration for those two member states – starts to become visible. The incident on New Year's Eve can be considered an unexpected event. The speech of the German minister of the interior was part of an ongoing political controversy, but the content and timing was unknown prior to the event. Hence, we expect more of a run-up for the first event, and more sudden changes for the other two.

We begin with the opening of labor markets for Romanians and Bulgarians on January 1st, 2014, in the first column of Figure 12. There is a clear increase around the event, signaling that even with anticipation the realization of the law was an important topic for public discussion. Figure 12 shows the composition of themes as a share of all immigrant narrative fragment sentences. We observe a clear and drastic increase in the share of *Economy* narratives, driven by *Welfare* and *Work*.

Looking at the absolute changes for those two themes, we find a strong increase in both. What is interesting is that there is a build-up with those narrative fragments already increasing since mid of December, fitting to this being an anticipated event. It is also interesting just how short-lived this change is. After a little more than two weeks both narrative fragments seem to convert back to their initial levels.

One of the most interesting insights comes from analyzing sentiment and its decomposition. We decompose the change in sentiment with respect to the fortnight before the event. After the event, we see a clear improvement in sentiment, which is persistent at least over the following 8 weeks. This shift is mostly driven by within-theme changes in sentiment. This means it is not only the fact that journalists write more about themes that feature more positive immigrant narratives, but they also write about more positive narratives within a particular theme.

Figure 12: Event-study Analysis for Three Salient Events



Close-up of of four weeks before and 8 weeks after three salient events using the same sample of articles as Figure 11: the opening of the German labor market to Bulgarians and Romanians on first of January 2014, the series of sexual assaults during New Year's Eve in several cities in Germany and the statement "Islam gehört zu Deutschland" (Islam does not belong to Germany) by the minister of the interior, Horst Seehofer. All data is aggregated at the biweekly level. The first row shows the number of articles per two weeks, the second the narrative theme shares, as in Figure 7ab, the third and fourth the total number of sentences for selected relevant themes and the fifth shows the sentiment decomposition with respect to the two weeks prior to the event date, as in Figure 11d.

We continue with the New Year's Eve sexual assaults in the second column of Figure 12. This is a particularly noteworthy event, which is believed to have contributed to a more negative perception of refugees after the initial more positive "welcome culture" attitude. We focus specifically on the*Immigrant Criminality* and *Foreign Religion* themes given that many incidents were considered potential criminal offense and most perpetrators being from Muslim states in the Maghreb region.

The first row, focusing on the number of articles, validates the salience of the event in German media. There is an increase of almost 100% in articles that we classify as containing at least one immigrant narrative. The theme composition in the second row suggests in particular an increase in *Immigrant Criminality*, with some more minor shifts in other themes.

Row three and four provide more details. Indeed, while there is also an increase in *Foreign Religion* narratives compared to the two weeks before, this increase is not exceptional compared to earlier weeks. In contrast, there is a drastic increase in *Immigrant Criminality* narratives, which more than triple. Hence the absolute increase in immigrant narratives seen in row one reflects mostly the focus on this theme. With this simple analysis, we cannot yet say with certainty to what extent this change reflects mostly reporting about the event itself, or also a more general increase in reporting about immigrants as a risk with regard to crime. A qualitative review of some articles suggests that both happens.

In row five, we again look at changes in sentiment compared to the two weeks before and decompose those changes. As to be expected, we observe sentiment conveyed in immigrant narratives to become more negative. The change is driven by the shift in theme composition towards *Immigrant Criminality*, but there is also a negative effect on within-theme sentiment. Accordingly, it is not only the increase in reporting about crime, but also that journalists use more negative language when writing about the same theme. The worsening of sentiment persists for several weeks. After six weeks, the shift towards more crime is over, however, there is seems to be a persistent negative change in theme-specific sentiment.

The third and final event is a bit more subtle and complex event that requires a short backstory. On October 3rd, 2010, back-then German president Christian Wulff holds his first major speech during the celebration of German re-unification. The more memorable part of this speech, however, was not about German re-unification. Instead, he stated that "Islam belongs to Germany." To understand the significance of this statement, one needs to know that for decades in particular conservative German politicians were keeping the illusion alive that millions of former guest workers would finally return to their home countries. Whether Germany was an "Einwanderungsland" (broadly meaning "immigration destination country") was one of the big controversies in German politics over many years.

The fact that a conservative president would declare that Islam belongs to Germany was thus a powerful statement, marking a softer stance on Islam by the mainstream of the main conservative party. Those parts of German society more sceptical about immigrants never fully accepted the statement, and the shift likely contributed some voters shifting to support the radical right Alternative for Germany (AFD). Against this background, the conservative minister of the interior Horst Seehofer pursued an agenda of appealing to conservative voters. One important step was publicly declaring that for him "Islam does not belong to Germany" on March 15th, 2018. Rather than overwhelming support, however, this sparked an open debate in German media. In the end, German chancellor Angela Merkel took sides against her own minister and said that she is the chancellor of all Germans, including of course Muslims.

Our analysis in the third column of Figure 12 helps to understand how this event influenced immigrant narratives. We observe an increase in articles containing immigrant narratives overall, and the stacked composition graph in row two suggests an uptake in particular in *Foreign Religion* and *Cultural Integration* narratives. Row three and four support this in absolute numbers. Looking at *Cultural Integration* reveals an increase already before the statement, signaling that the statement was issued at a time of general discussions about immigrant integration. The development of *Foreign Religion* still shows that the statement itself had a large impact on the media discourse, with religion narratives more than tripling in absolute terms.

The most interesting question is then to understand how sentiment shifted in this debate. Was the conservative minister successful in facilitating a more sceptical discussion of Islam? What we observe is that narratives in the run-up to this event were actually more negative compared to the two weeks directly before. Sentiment after the statement, however, tend to improve rather than become more negative. Most notably, this is is case even though the decomposition reveals a shift towards more negative themes. Nonetheless, within-theme sentiment improves so much that the overall development is positive. In that sense, the conservative attempt to roll back a more progressive position on immigration backfired.[29]

Overall, the three event studies reveal the potential of our detailed immigrant narrative dataset and provide some very interesting insights that can form the basis of more in-depth analysis. Michalopoulos and Xue (2021) indicate that folklore is influenced by regional conditions, and we show that immigrant narratives are also influenced by local conditions, but also adapt to important related events. Events increase the salience of immigrant narratives that are associated with those events. This increase can at least persist for several weeks beyond the specific event. It remains an open question to what extent certain events have the potential to permanently change the discourse about immigration.

We find that the changes in narrative theme composition are substantial. While *Economy* themes are generally not a dominant narrative in newspapers, their importance seems to grow in line with events that specifically relate to the role of immigrants for the economy. Hence, it is not that journalists ignore *Economy* narratives, but rather that they are usually crowded out by other narrative themes.

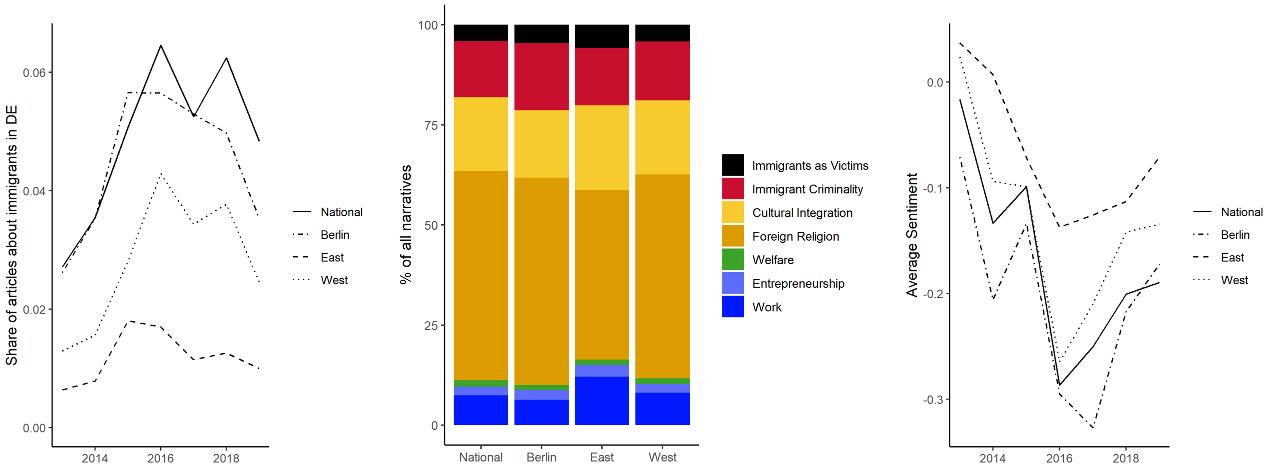
Finally, we demonstrate the analytical options enabled by being able to decompose sentiment changes. Average sentiment allows us to assess the effect of an event immediately and over time. Decomposing allows us to understand how and why the event influenced sentiment towards immigrants. In two cases, average sentiment moves as expected, in one the net effect was open and slightly surprising. In one case shifts towards more negative themes dominate the changes, but in two other we also document shifts in sentiment within themes. This is an important analytical distinction. Journalists decide about what to write, but also about how to write about it.

7 Immigrant Narratives Across Space

In the previous sections, we focused on aggregate changes in newspaper coverage over time. However, newspapers differ in various observable and unobserved characteristics, which could affect coverage about immigrants. As one exploratory analysis in this direction, we divide newspapers based on their spatial coverage in four groups: national newspapers, regional newspapers from Berlin, regional newspapers from former East Germany, and regional newspapers from former West Germany. Figure 13 shows the share of articles that contain at least one immigrant narrative, the composition of themes, and the average overall sentiment. In the left panel, we find that national and Berlin-focused regional newspapers write considerably more articles that contain an immigrant narrative than regional newspapers, especially from former East Germany. For national newspapers, the share of articles writing about immigrants in Germany exceeds 6 percent in 2016, whereas this figure is less than 2 percent in regional newspapers in former East Germany.

[29] Analyzing variation across space and partisan position of newspapers is an interesting endeavor that could reveal more details.

Figure 13: Heterogeneity in Immigrant Narratives across National and Regional Newspapers



(a) Share of articles about immigration

(b) Theme composition of narratives

(c) Average aggregated sentiment

For the sample of newspapers present since at least 2013, we identify four main groups of spatial distribution: two national newspapers, four regional newspapers from Berlin, 26 regional newspapers from former West Germany and nine regional newspapers from former East Germany. We exclude weekly national newspapers. **Panel (a)** shows the share of all articles in Factiva that contain at least one immigrant narrative, by year and by newspaper group. **Panel (b)** shows the relative prevalence of narrative themes over the whole period from 2013 to 2019. See notes to Table 7b for further explanation. **Panel (c)** shows the unweighted average sentiment, by year and newspaper group. See notes to Table 8 for further explanation.

The middle panel shows the composition of themes across the four newspaper groups. There are some interesting differences: (1) newspapers focused on Berlin contain more *Immigrant Criminality* sentences, (2) regional newspapers from former East Germany contain fewer *Foreign Religion* sentences and more about *Work*, and (3) national newspapers and regional newspapers from former West Germany are very similar in terms of theme composition. As for the frequency of articles, there are many potential reasons for observed differences, such as differences in regional newspaper market structure and the local presence of immigrants.

The rightmost panel finally shows the average sentiment for the four newspaper groups. For each of the years, regional newspapers in former East Germany are less negative about immigrants than the other three groups. Also, the decrease in the average sentiment from 2013 until 2016 was considerably steeper in regional newspapers in former West Germany than in regional newspapers in former East Germany. Although also the improvement in average sentiment after 2016 was steeper in former West Germany, the average sentiment there remained below the average sentiment in regional newspapers in former East Germany throughout the period we analyze. Furthermore, average sentiment in all four newspaper groups remained in negative territory below their 2013 level in all subsequent years.

Both the overall more positive sentiment and the smaller decline in the East can be seen as surprising. The far-right parties like the AFD are much more popular in former East Germany and general attitudes towards immigrants are more negative there. This does not seem to be reflected in the newspapers in our sample. Our findings suggest that negative sentiment towards immigrants spreads mostly in the social media and through personal connections, with newspapers rather providing a potential counter-balance against anti-immigrant sentiments in the former East Germany. The initially quite neutral stance towards immigrants may help to explain widespread hostility towards traditional media among far-right supporters who label it as Lügenpresse (lying press).

We provide a more detailed analysis for the different spatial newspaper groups in Appendix C.3. Appendix Figure C.10 shows that the shares of different narrative themes develop quite similarly over the whole time period in national newspapers, in Berlin-based newspapers, and in newspapers in former West Germany. In regional newspapers in former East Germany, *Work* is more pronounced and *Foreign Religion* less pronounced than in other groups. Appendix Figure C.11 presents quarterly sentiment in regional newspapers in former West and former East Germany. Theme-specific sentiments are quite similar, suggesting that the difference in aggregate sentiment is driven more by theme composition than by differences in theme-specific sentiments between former West and former East Germany.

Finally, Appendix Figure C.12 decomposes the change in aggregate sentiment into shifts in theme composition and and shifts in theme-specific sentiments, compared with year 2013. In former West Germany, theme composition has shifted to themes with more negative average sentiment and there has been also a negative shift in theme-specific average sentiments, with both changes being strong from 2015 onward. In former East Germany, theme composition and theme-specific average sentiment became more positive in 2014. From 2015 onward, theme composition changed to themes with more negative average sentiment as in the former West Germany, but changes in theme-specific average sentiment remained modest. This suggests that the increase in far-right voting in former East Germany cannot be explained by more negative newspaper coverage of immigrants than in former West Germany. On the contrary, regional newspapers in former East Germany have remained less negative in their immigrant narratives than newspapers in former West Germany, and even had positive aggregate sentiment in 2014 when the sentiment was already negative in former West Germany.

8 Conclusion

We propose a novel way to study narratives using text-as-data methods, and apply it to media narratives about immigrants. This contributes to better understanding of the political and societal consequences of immigration (Alesina and Tabellini, Forthcoming). We focus on Germany as the largest member state of the European Union that has a large and diverse immigrant population and is the main destination country of asylum seekers. Germany also features a rich and diverse landscape of regional newspapers, opening up the possibility to link immigrant narratives to specific local conditions. We analyze 107,428 articles about immigrants in Germany over the period from 2000 to 2019, stemming from five national and 65 regional newspapers.

To provide a comprehensive dataset capturing immigrant narratives, we combine more traditional dictionary-based approaches with the possibilities of modern Natural Language Processing (NLP) packages that allow detecting linguistic features like grammar, word types, and dependencies. For each sentence, our method aims to detect (i) whether the sentence is about immigrants; (ii) if it fits into one or more narrative themes we have identified; and (iii) if it has a (theme-specific) negative, neutral, or positive sentiment. We classify narratives into the following pre-selected themes: *Economy* (subdivided in *Work*, *Welfare*, and *Entrepreneurship*), *Foreign Religion*, *Cultural Integration*, *Immigrant Criminality*, and *Immigrants-as-Victims*.

Our dataset provides several interesting insights. First, we find that *Economy* and *Immigrant Criminality* narratives - the subject of many studies - are dominated by the themes *Foreign Religion* and *Cultural Integration*. Second, we find that themes differ a lot in their sentiment, some being generally very positive, others usually negative. For instance, *Foreign Religion* narratives are mostly negative, whereas *Cultural Integration* tends to be a positive theme.

Thirdly, we analyze three important immigration-related events in recent German history and their influence on immigrant narratives. The event studies provide some interesting insights into how narratives in the media work. Our approach allows us to decompose sentiment shifts in theme composition and within-theme shifts. Changes in narrative theme composition are substantial. In two cases, average sentiment moves as expected, in one the net outcome was surprising. In one case shifts towards more negative themes dominate the changes, but in two other we also document shifts in sentiment within themes.

Our analysis provides general methodological insights that can be applied in other research using text-as-data. We developed functions to correct negations and qualifications that are common in newspaper texts, moving beyond standard dictionary-based approaches. Those functions can be reused by anyone working with German text. Our aim to provide a very precise and detailed account of immigrant narratives in Germany also required a high initial investment in gathering and fine-tuning the theme-specific dictionaries. These dictionaries can now be reused to study immigrants in any texts in German.

Our validation analysis using a large set of human coders uncovers that there are also significant differences in how humans classify and assign sentiment for the same sentence. Nonetheless, we can show that when focusing on those sentences where human coders agree, our algorithm performs remarkably well. This shows the potential of combining dictionaries with NLP tools to both increase true positives and decrease false positives in sentence classification into themes. The advantage of accounting for linguistic features in our NLP functions turns out to be even larger for sentiment assignment compared to simple general dictionaries. Which approach is the best fit depends on the goal of a specific study, but based on our results we encourage other researchers to use the tools we provide and combine dictionaries with modern NLP functionality.

References

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics*, 115(3): 715–753.

Alesina, Alberto, and Marco Tabellini. Forthcoming. "The Political Effects of Immigration: Culture or Economics?" *Journal of Economic Literature*.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. Forthcoming. "Immigration and Redistribution." *Review of Economic Studies*.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart. 2021. "Narratives about the Macroeconomy." *CEBI Working Paper 18/21, Available at SSRN 3974993*.

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2021. "Text Semantics Capture Political and Economic Narratives." *Unpublished Working Paper, arXiv preprint arXiv:2108.01720*.

Ash, Elliott, Ruben Durante, Maria Grebenshchikova, and Carlo Schwarz. 2022. "Visual representation and stereotypes in news media." *CESifo Working Paper No. 9686*.

Ash, Elliott, Sergio Galletta, Dominik Hangartner, Yotam Margalit, and Matteo Pinna. 2020. "The Effect of Fox News on Health Behavior During COVID-19." *Available at SSRN 3636762*.

Battisti, Michele, Gabriel Felbermayr, Giovanni Peri, and Panu Poutvaara. 2018. "Immigration, Search and Redistribution: A Quantitative Assessment of Native Welfare." *Journal of the European Economic Association*, 16(4): 1137–1188.

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. "Narratives, Imperatives, and Moral Reasoning." *NBER Working Paper No. w24798, Available at SSRN 3214341*.

Besley, Timothy, and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence From India." *The Quarterly Journal of Economics*, 117(4): 1415–1451.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. "Identity, Beliefs, and Political Conflict." *The Quarterly Journal of Economics*, 136(4): 2371–2411.

Borjas, George J. 2003. "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *The Quarterly Journal of Economics*, 118(4): 1335–1374.

Bursztyn, Leonardo, Aakaash Rao, Christopher P Roth, and David H Yanagizawa-Drott. Forthcoming. "Opinions as Facts." *Review of Economic Studies*.

Cagé, Julia, Nicolas Hervé, and Béatrice Mazoyer. 2022. "Social Media Influence Mainstream Media: Evidence from Two Billion Tweets." *CEPR Discussion Paper No. 17358*.

Cagé, Julia, Nicolas Hervé, and Marie-Luce Viaud. 2020. "The Production of Information in an Online World." *The Review of Economic Studies*, 87(5): 2126–2164.

Campante, Filipe, Ruben Durante, and Francesco Sobbrio. 2018. "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation." *Journal of the European Economic Association*, 16(4): 1094–1136.

Card, Dallas, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. "Computational Analysis of 140 Years of US Political Speeches Reveals More Positive but Increasingly Polarized Framing of Immigration." *Proceedings of the National Academy of Sciences*, 119(31): e2120510119.

Card, David, Christian Dustmann, and Ian Preston. 2012. "Immigration, Wages, and Compositional Amenities." *Journal of the European Economic Association*, 10(1): 78–119.

Couttenier, Mathieu, Sophie Hatte, Mathias Thoenig, and Stephanos Vlachos. Forthcoming. "The Logic of Fear - Populism and Media Coverage of Immigrant Crimes." *Review of Economics and Statistics*.

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist. 2012. "Ethnic Diversity and Preferences for Redistribution." *Journal of Political Economy*, 120(1): 41–76.

Dehdari, Sirus H, and Kai Gehring. 2022. "The Origins of Common Identity: Evidence from Alsace-Lorraine." *American Economic Journal: Applied Economics*, 14(1): 261–92.

Depetris-Chauvin, Emilio, Ruben Durante, and Filipe R. Campante. 2020. "Building Nations Through Shared Experiences: Evidence from African Football." *American Economic Review*, 110(5): 1572–1602.

Djourelova, Milena. 2022. "Media persuasion through slanted language: Evidence from the coverage of immigration." *Unpublished Working Paper*.

Durante, Ruben, and Ekaterina Zhuravskaya. 2018. "Attack When the World is not Watching? US News and the Israeli-Palestinian Conflict." *Journal of Political Economy*, 126(3): 1085–1133.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. "The Political Legacy of Entertainment TV." *American Economic Review*, 109(7): 2497–2530.

Edo, Anthony, Yvonne Giesing, Jonathan Öztunc, and Panu Poutvaara. 2019. "Immigration and Electoral Support for the Far-Left and the Far-Right." *European Economic Review*, 115: 99–143.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–85.

Enke, Benjamin. 2020. "Moral Values and Voting." *Journal of Political Economy*, 128(10): 3679–3729.

Esposito, Elena, Tiziano Rotesi, Alessandro Saia, and Mathias Thoenig. 2021. "Reconciliation Narratives: The Birth of a Nation after the US Civil War." *CEPR Discussion Paper No. 15938, Available at SSRN 3816811*.

Facchini, Giovanni, and Anna Maria Mayda. 2009. "Does the Welfare State Affect Individual Attitudes Toward Immigrants? Evidence Across Countries." *The Review of Economics and Statistics*, 91(2): 295–314.

Fouka, Vasiliki. 2019. "How do Immigrants Respond to Discrimination? The Case of Germans in the US During World War I." *American Political Science Review*, 113(2): 405–422.

Galletta, Sergio, and Elliott Ash. 2019. "How Cable News Reshaped Local Government." *Unpublished Working Paper*.

Gehring, Kai. 2021. "Overcoming History Through Exit or Integration: Deep-Rooted Sources of Support for the European Union." *American Political Science Review*, 115(1): 199–217.

Gehring, Kai. 2022. "Can External Threats Foster a European Union Identity? Evidence from Russia's Invasion of Ukraine." *The Economic Journal*, 132(644): 1489–1516.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature*, 57(3): 535–74.

Gentzkow, Matthew, Jesse M Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review*, 101(7): 2980–3018.

German Audit Bureau of Circulation (IVW). 2020. "IVW VA2020."

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty. 2022. "Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020." *The Quarterly Journal of Economics*, 137(1): 1–48.

GLES. 2019. "Nachwahl-Querschnitt (GLES 2017)." *GESIS Datenarchiv, Köln. ZA6801 Datenfile Version 4.0.1, https://doi.org/10.4232/1.13235*.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-term Persistence." *Journal of the European Economic Association*, 14(6): 1401–1436.

Halla, Martin, Alexander F Wagner, and Josef Zweimüller. 2017. "Immigration and Voting for the Far Right." *Journal of the European Economic Association*, 15(6): 1341–1385.

Hangartner, Dominik, Elias Dinas, Moritz Marbach, Konstantinos Matakos, and Dimitrios Xefteris. 2019. "Does Exposure to the Refugee Crisis Make Natives More Hostile?" *American Political Science Review*, 113(2): 442–455.

Harari, Yuval Noah. 2015. *Sapiens: A Brief History of Humankind.* New York, Harper.

Hatton, Timothy J. 2017. "Refugees and Asylum Seekers, the Crisis in Europe and the Future of Policy." *Economic Policy*, 32(91): 447–496.

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. "spaCy: Industrial-strength Natural Language Processing in Python." https://github.com/explosion/spaCy.

Hudson, Richard Paul. 2022. "Coreferee: Coreference Resolution for Multiple Languages." github.com/explosion/coreferee.

Hyndman, Rob J, and Yanan Fan. 1996. "Sample Quantiles in Statistical Packages." *The American Statistician*, 50(4): 361–365.

Kearney, Melissa S, and Phillip B Levine. 2015. "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing." *American Economic Review*, 105(12): 3597–3632.

Keita, Sekou, Thomas Renault, and Jérôme Valette. 2021. "The Usual Suspects: Offender Origin, Media Reporting and Natives' Attitudes Towards Immigration." *Documents de travail du Centre d'Économie de la Sorbonne*.

Loria, Steven. 2018. "Textblob Documentation." github.com/sloria/textblob.

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. "Folklore." *The Quarterly Journal of Economics*, 136(4): 1993–2046.

Norris, Pippa, and Ronald Inglehart. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism.* Cambridge University Press.

Osnabrügge, Moritz, Elliott Ash, and Massimo Morelli. 2021. "Cross-Domain Topic Classification for Political Texts." *Political Analysis*, 1–22.

Ottaviano, Gianmarco IP, and Giovanni Peri. 2012. "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association*, 10(1): 152–197.

Otto, Alkis Henri, and Max Friedrich Steinhardt. 2014. "Immigration and Election Outcomes—Evidence from City Districts in Hamburg." *Regional Science and Urban Economics*, 45: 67–79.

Remus, Robert, Uwe Quasthoff, and Gerhard Heyer. 2010. "SentiWS - A Publicly Available German-language Resource for Sentiment Analysis." *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), available for spaCy at* github.com/Liebeck/spacy-sentiws.

Scheve, Kenneth F, and Matthew J Slaughter. 2001. "Labor Market Competition and Individual Preferences Over Immigration Policy." *Review of Economics and Statistics*, 83(1): 133–145.

Shiller, Robert J. 2017. "Narrative Economics." *American Economic Review*, 107(4): 967–1004.

Snyder, James M., and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy*, 118(2): 355–408.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics*, 12: 415–438.

Appendix A Methodology

Dictionaries

Figure A.1: Dictionaries Used in Article Selection, Sentence Selection, Theme Assignment, and Sentiment Assignment

- **A1 - Immigrant dictionary** terms related to immigration and immigrants. Examples are ”Asylunterkunft” (”Asylum shelter”) and ”EU-Angehörige” (”EU Citizen”). N = 64.
- **A2 - Foreign and dual nationalities dictionary** terms related to nationalities and dual nationalities, both adjectives and nouns. Examples are ”Ungarinnen” (”Hungarians” - female) and ”Deutsch-Afghanisch*” (”German-Afghan”). N = 5467.
- **A3 - Foreign names dictionary** terms related to immigration and immigrants. Examples are ”Mohammed” and ”Janowski”. N = 7651.
- **A4 - Countries dictionary**. N = 203.
- **A5 - Theme specific dictionary** terms related to predefined themes and sub-themes. Each sub-theme has its own respective dictionary: Economy - Work, Economy - Welfare, Economy - Entrepreneurship, Foreign Religion, Cultural Integration, Immigrant Criminality, and Immigrants-as-Victims. Total = 1014.
  - **A5.1 - Foreign Religion dictionary** terms related to immigration and immigrants manually compiled. Examples are ”Muslim” (”Muslim”) and ”Moschee” (”Mosque”). N = 129.
  - **A5.2 - Cultural Integration dictionary** terms related to cultural integration, including education. Examples are ”integrationsbereit” (”willing to integrate”) and ”Schüler” (”high school student”). N = 185.
  - **A5.3 - Economy dictionaries** terms related to the economy, manually compiled. Subdivided in work, entrepreneurship, and welfare state. Respective examples are ”arbeitsunwillig” (”unwilling to work”), ”Unternehmer” (”entrepreneur”) and ”Sozialhilfe” (”social assistance”). N = 280.
  - **A5.4 - Crime dictionary** terms related to crimes, manually compiled. Examples are ”Mord” (”Murder”) and ”Dealer” (”dealer”). N = 230.
  - **A5.5 - Anti-immigrant dictionary** terms related to crimes and negative attitudes towards immigrants, manually compiled. Examples are ”Muslim-Mörder” (”Muslim killer”) and ”fremdenfeindlich” (”xenophobic”). N = 144.
  - **A5.6 - Theme-specific n-gram dictionary** terms related to various themes, manually compiled. Examples are ”Bedarf an” (”demand for”), related to economy and ”kulturelle Vielfalt” (”cultural diversity”), related to cultural integration. N = 46.
- **A6 - German/foreign locations dictionary** terms used for German and foreign locations. N = 16504

Factiva filter

We queried relevant newspaper articles from Factiva by newspaper and time period using the following logical filter:

(einwander OR eingewandert* OR migra* OR ausländer* OR islam* or muslim*)* ***AND*** *(deutsch* OR schleswig-holstein OR mecklenburg-vorpommern OR hamburg OR bremen OR niedersachsen OR nordrhein-westfalen OR sachsen OR sachsen-anhalt OR berlin OR brandenburg OR thüringen OR hessen OR baden-württemberg OR saarland OR rheinland-pfalz OR bayern OR münchen OR köln OR Frankfurt OR Stuttgart OR düsseldorf OR dortmund OR essen OR leipzig OR dresden OR hannover OR nürnberg OR duisburg OR bochum OR wuppertal OR bielefeld OR bonn OR münster OR karlsruhe OR mannheim OR augsburg OR wiesbaden OR gelsenkirchen OR mönchengladbach OR braunschweig OR kiel OR aachen OR magdeburg OR*

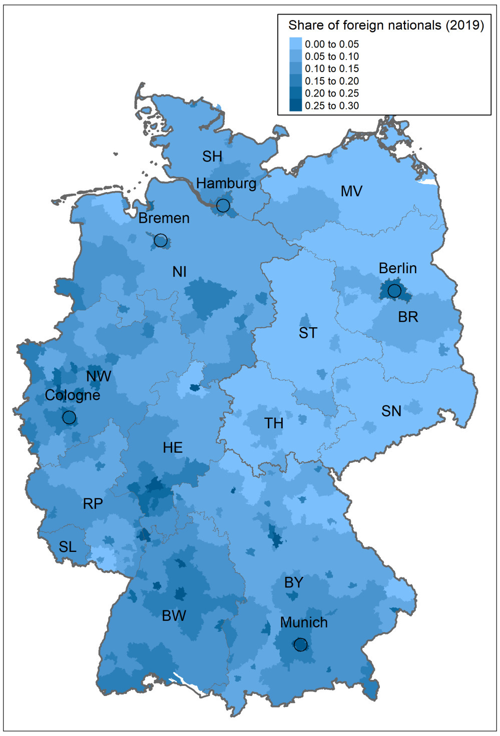
freiburg OR krefeld OR lübeck OR oberhausen OR erfurt OR mainz OR hagen OR hamm OR saarbrücken OR mülheim OR potsdam OR leverkusen OR osnabrück OR solingen OR ludwigshafen OR oldenburg OR hanau OR kassel OR halle OR rostock OR ruhrgebiet) ***NOT*** *(covid* OR corona* OR "Ausstellungen im" OR "Was diese Woche bringt" OR amerika* OR US OR USA OR "Theater und Tanz*" OR "veranstaltung in" OR "Kein Titel" OR "THEATER - OPER - TANZ - SHOW" OR "ROCK & POP - KLASSIK - JAZZ & BLUES" OR "ROCK & POP - JAZZ & BLUES - WELTMUSIK")*

Article location

We determine whether an article is probably in Germany using the entities (subdivided in locations, organizations, and miscellaneous entities such as events and nationalities). If any of the first 30 words contains only a location in Germany (Using locations from SPACY), the article is probably about immigrants in Germany. If it only mentions a location abroad, it is not about Germany. If neither is the case, we register all entities in the article and assign the article to be probably not about Germany when foreign entities are mentioned at least twice as often as German entities. If this is not the case, we register the first mentioned entity in the article. If the first mentioned entity is German, the article is probably about Germany. If it is foreign, the article is probably about foreign. If the entity's location is ambiguous or if there is no entity mentioned in the article, we set this variable to missing. On our full sample between 2000 and 2019, we identify 82.5% of articles to be about immigrants in Germany, 17.1% of articles to be about immigrants outside Germany and 0.5% could not be assigned any geographic location.

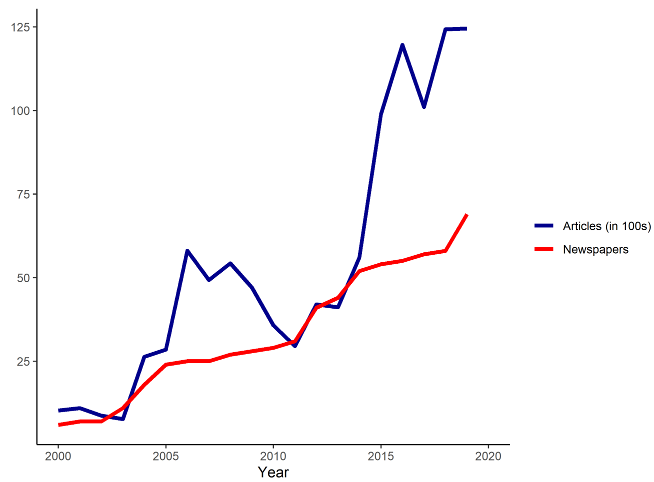
Appendix B Descriptive Statistics

Figure B.1: Share of Foreigners by County in 2019



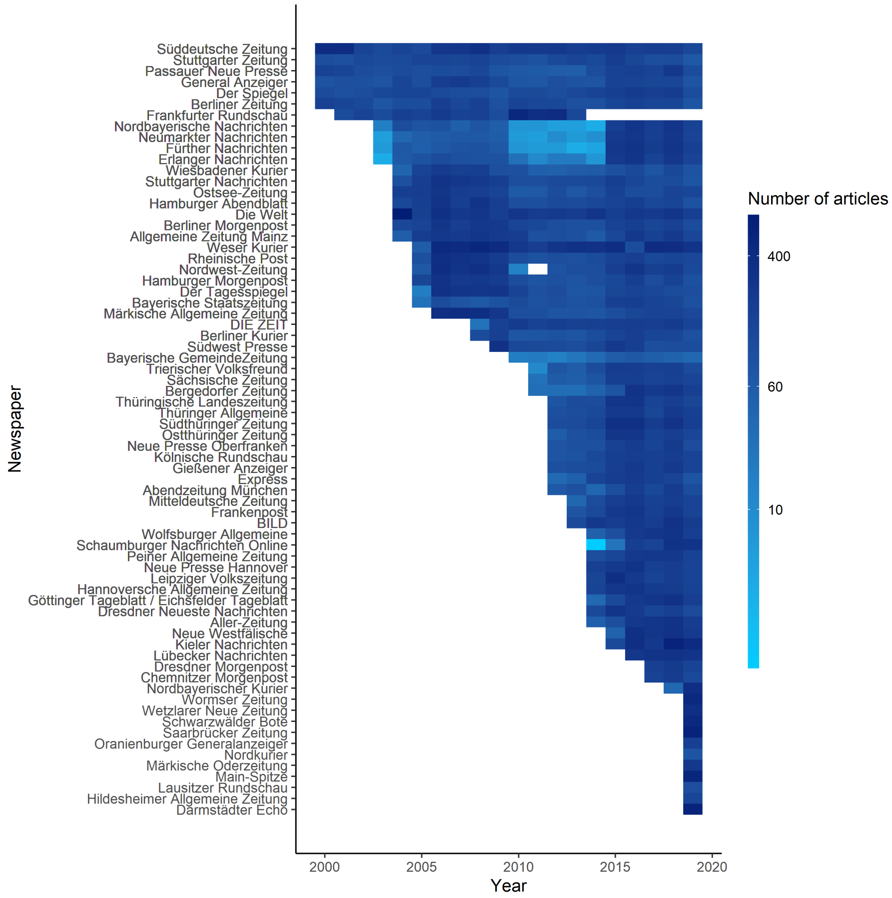
This figure shows the share of immigrants based on nationality, by county (*Landkreis* in German). State borders are drawn in grey, the border between former Eastern (GDR) and Western (FRG) Germany is drawn as a thicker grey line. The city states of Bremen, Hamburg and Berlin are indicated explicitly, as well as the cities of Cologne and Munich (which have a population exceeding 1 million). The other states are abbreviated in the following way (clockwise order): SH = Schleswig-Holstein, MV = Mecklenburg-Vorpommern, BR = Brandenburg, ST = Saxony-Anhalt, SN = Saxony, TH = Thuringia, BY = Bavaria, BW = Baden-Württemberg, SL = Saarland, RP = Rhineland-Palatinate, HE = Hesse, NW = Nordrhein-Westfalen, NI = Lower Saxony.

Figure B.2: Number of Newspapers and Articles from 2000 to 2019



This figure shows the number of newspapers in our dataset based on *Factiva* and the total number of articles downloaded, by year.

Figure B.3: Newspaper Coverage from 2000 to 2019



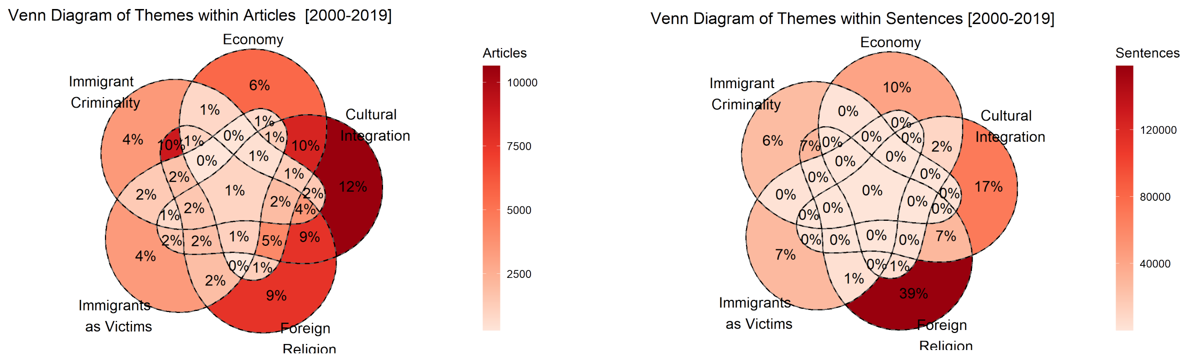
This figure shows the number of articles in our dataset per newspaper per year on a logarithmic color scale.

Table B.1: Summary Statistics of National and Regional Newspapers

	Newspaper	Articles	About immigrants	From agency	Average sentiment
1	Weser Kurier	4703	0.80	0.02	-0.05
2	**Süddeutsche Zeitung**	4480	0.87	0.04	-0.06
3	**Die Welt**	4242	0.83	0.09	-0.15
4	Frankfurter Rundschau	2741	0.86	0.11	0.00
5	*Der Spiegel*	2705	0.83	0.01	-0.18
6	Allgemeine Zeitung Mainz	2672	0.85	0.19	-0.05
7	Berliner Zeitung	2632	0.86	0.09	-0.09
8	Nordwest-Zeitung	2610	0.75	0.12	-0.10
9	Hamburger Abendblatt	2596	0.83	0.23	-0.13
10	Berliner Morgenpost	2547	0.82	0.17	-0.14
11	Rheinische Post	2545	0.81	0.02	-0.05
12	Stuttgarter Zeitung	2488	0.87	0.12	0.01
13	Stuttgarter Nachrichten	2466	0.85	0.22	-0.07
14	Märkische Allgemeine Zeitung	2459	0.73	0.27	-0.03
15	General Anzeiger	2445	0.84	0.24	-0.06
16	Passauer Neue Presse	2403	0.76	0.00	0.01
17	Der Tagesspiegel	2326	0.87	0.13	-0.07
18	Ostsee-Zeitung	2302	0.69	0.05	-0.03
19	Hamburger Morgenpost	2289	0.77	0.02	-0.20
20	*DIE ZEIT*	2085	0.85	0.00	-0.05
21	Südthüringer Zeitung	1750	0.79	0.35	-0.11
22	Wiesbadener Kurier	1718	0.85	0.29	-0.12
23	Erlanger Nachrichten	1711	0.79	0.01	-0.05
24	Kieler Nachrichten	1709	0.76	0.01	-0.05
25	Südwest Presse	1702	0.82	0.33	-0.10
26	Fürther Nachrichten	1679	0.80	0.01	-0.10
27	Bayerische Staatszeitung	1631	0.69	0.05	-0.01
28	Nordbayerische Nachrichten	1616	0.79	0.01	-0.07
29	**BILD**	1541	0.76	0.00	-0.31
30	Neumarkter Nachrichten	1501	0.77	0.01	-0.11
31	Berliner Kurier	1462	0.78	0.32	-0.29
32	Thüringische Landeszeitung	1348	0.79	0.21	-0.03
33	Thüringer Allgemeine	1331	0.79	0.18	-0.07
34	Gießener Anzeiger	1321	0.85	0.29	-0.21
35	Ostthüringer Zeitung	1296	0.75	0.32	-0.08
36	Frankenpost	1287	0.83	0.10	-0.19
37	Hannoversche Allgemeine Zeitung	1286	0.86	0.04	-0.25
38	Leipziger Volkszeitung	1249	0.81	0.09	-0.18
39	Neue Presse Hannover	1231	0.82	0.01	-0.24
40	Kölnische Rundschau	1220	0.82	0.41	-0.21
41	Bergedorfer Zeitung	1196	0.76	0.23	-0.23
42	Neue Westfälische	1183	0.82	0.10	-0.15
43	Aller-Zeitung	1177	0.75	0.03	-0.32
44	Göttinger Tageblatt / Eichsfelder Tageblatt	1173	0.81	0.07	-0.21
45	Peiner Allgemeine Zeitung	1147	0.78	0.05	-0.30
46	Neue Presse Oberfranken	1146	0.83	0.10	-0.20
47	Lübecker Nachrichten	1143	0.64	0.13	-0.06
48	Trierischer Volksfreund	1126	0.80	0.25	-0.17
49	Wolfsburger Allgemeine	1120	0.75	0.03	-0.32
50	Mitteldeutsche Zeitung	1110	0.81	0.03	-0.15
51	Express	1072	0.80	0.13	-0.37
52	Dresdner Neueste Nachrichten	1067	0.83	0.10	-0.14
53	Sächsische Zeitung	1064	0.87	0.19	-0.05
54	Schaumburger Nachrichten Online	966	0.85	0.77	-0.38
55	Abendzeitung München	955	0.77	0.37	-0.15
56	Darmstädter Echo	562	0.68	0.23	-0.14
57	Saarbrücker Zeitung	561	0.64	0.32	-0.07
58	Dresdner Morgenpost	553	0.70	0.00	-0.25
59	Main-Spitze	502	0.70	0.26	-0.16
60	Chemnitzer Morgenpost	500	0.71	0.00	-0.25
61	Schwarzwälder Bote	457	0.66	0.00	-0.04
62	Wormser Zeitung	445	0.69	0.29	-0.13
63	Nordbayerischer Kurier	403	0.64	0.36	-0.08
64	Bayerische GemeindeZeitung	378	0.74	0.00	0.14
65	Wetzlarer Neue Zeitung	351	0.70	0.30	-0.17
66	Märkische Oderzeitung	242	0.62	0.13	-0.04
67	Oranienburger Generalanzeiger	180	0.65	0.17	-0.13
68	Hildesheimer Allgemeine Zeitung	147	0.65	0.01	-0.19
69	Lausitzer Rundschau	98	0.67	0.46	-0.09
70	Nordkurier	79	0.67	0.00	-0.21

Number of articles, share of articles about immigrants, share of articles identified to originate from newspaper agencies and the aggregated average sentiment by newspaper. National daily newspapers are bold-faced, national weekly newspapers italic.

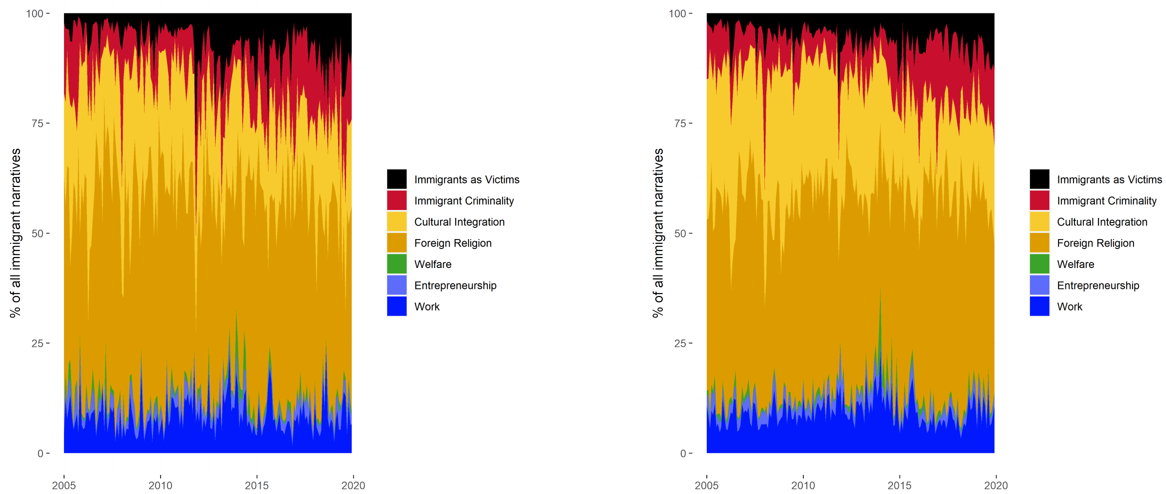
Figure C.1: Venn diagram of 5 Main Themes, on the Article and Sentence Level



(a) Article level **(b)** Sentence level

Panel (a) shows a Venn diagram of the five main themes (*Economy* includes *Work*,*Welfare*, and *Entrepreneurship*) on the article level. Each area represents a mutually exclusive region where the article contains between one and five of the themes. An example of how to read this diagram: 5 percent of articles containing at least one immigrant narrative contain *Cultural Integration*, *Foreign Religion* and *Economy*, but not *Immigrant Criminality* and *Immigrants-as-Victims*. **Panel (b)** shows a similar diagram, but instead of the article level, it concerns the sentence level.

Figure C.2: Shares of Themes between 2005 and 2019, for National and Regional Newspapers



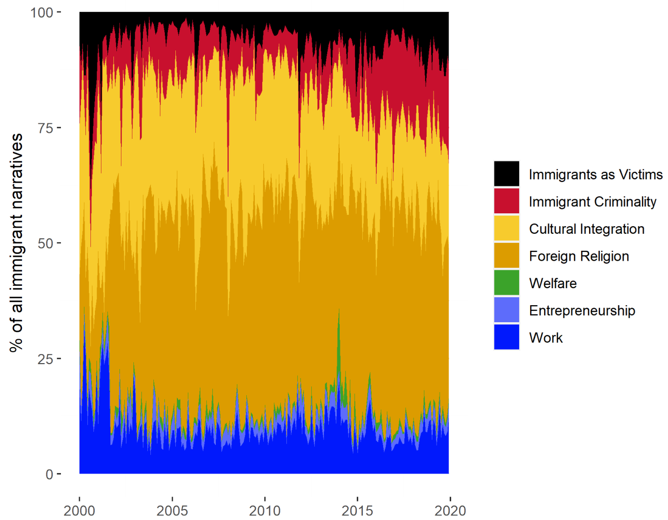
(a) National **(b)** Regional

Panel (a) shows the share of immigrant narratives about any of the seven themes over time, for a balanced panel of three national newspapers, by month. The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative theme this may exceed the number of sentences carrying narratives. **Panel (b)** shows a similar picture for a balanced panel of 24 regional newspapers.

Appendix C Additional Results

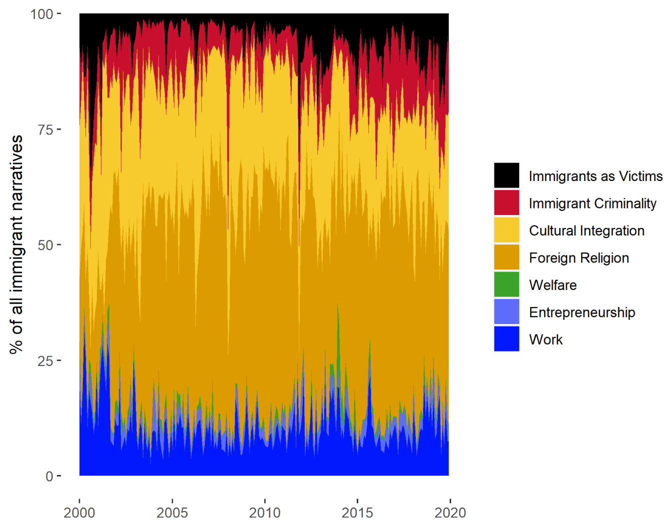
C.1 Theme Shares

Figure C.3: Shares of Themes between 2000 and 2019, Unbalanced



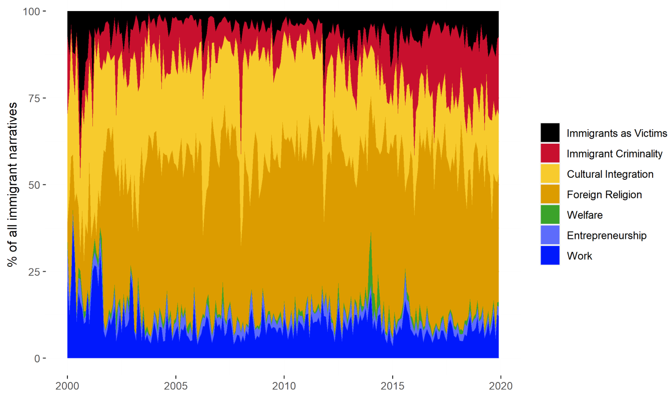
This figure shows the share of immigrant narratives about any of the seven themes over time, for all 70 newspapers by month. The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative theme this may exceed the number of sentences carrying narratives.

Figure C.4: Shares of Themes between 2000 and 2019, Fully Balanced



This figure shows the share of immigrant narratives about any of the seven themes over time, for all six newspapers that are covered uninterruptedly since 2000, by month. The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative theme this may exceed the number of sentences carrying narratives.

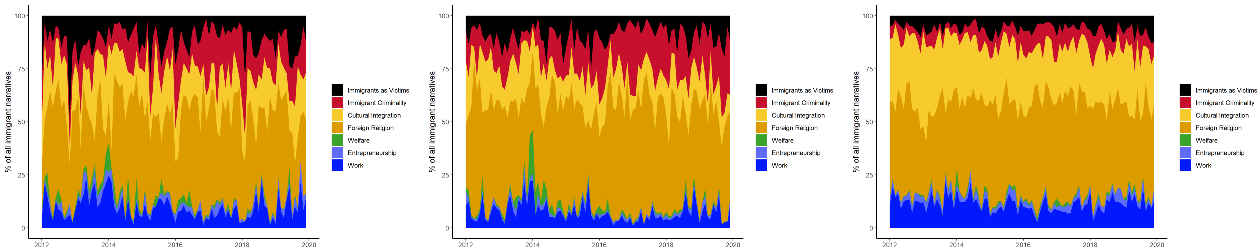
Figure C.5: Shares of Themes, weighted by sales



This figure shows the share of immigrant narratives about any of the seven themes over time, for all 70 newspapers in our dataset, by month. Each sentence is weighted by the sales from IVW of the respective newspaper in 2019. The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative theme this may exceed the number of sentences carrying narratives.

C.2 Heterogeneity across Newspaper Sections

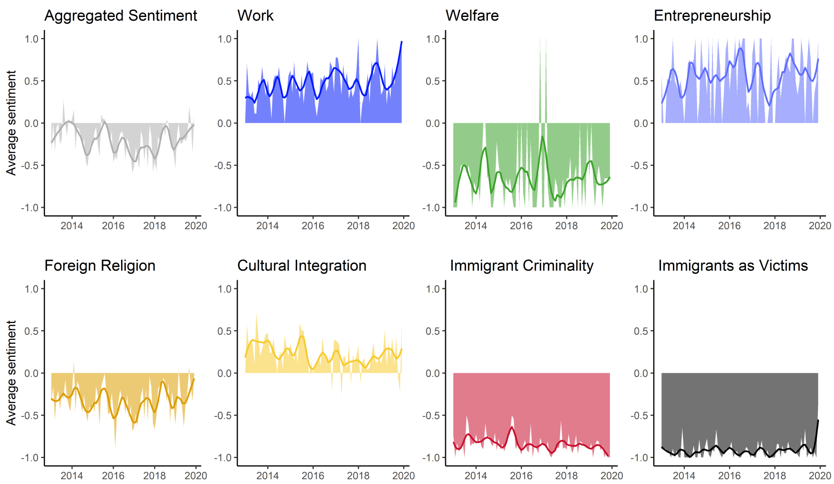
Figure C.6: Heterogeneity in Theme Shares across Newspaper Sections between 2013 and 2019



(a) Section ”news” **(b)** Section ”politics” **(c)** Section ”local”

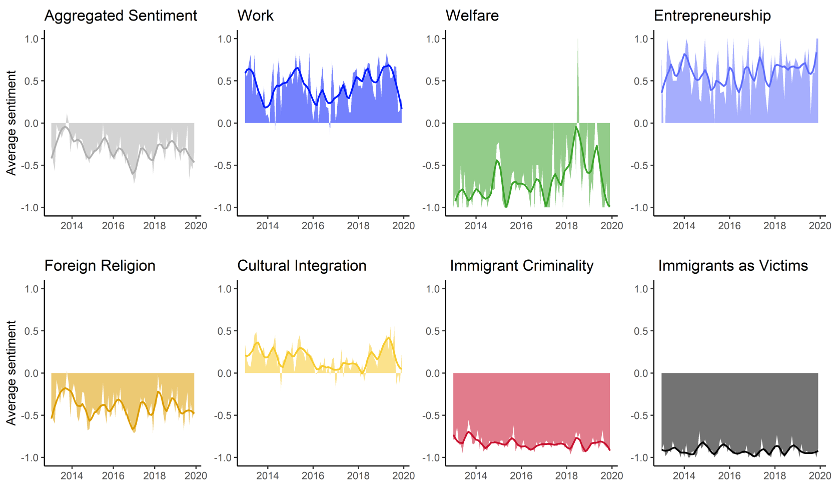
This figure shows the share of immigrant narratives about any of the seven themes over time, for all newspapers by month, subdivided by newspaper section. Newspaper sections are supplied by *Factiva* for 64 percent of all downloaded newspaper articles and are specific to newspapers. We classified sections into various broad categories, including news, politics, and local sections. **Panel (a)** concerns news sections (such as ”Nachrichten” (short news in German)), **Panel (b)** concerns news sections (such as ”Politik” (politics in German)), and **Panel (c)** concerns local news sections (identified by the name of a municipality or terms such as ”Lokal” (local in German)). The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative theme this may exceed the number of sentences containing immigrant narratives.

Figure C.7: Theme-specific Sentiments between 2013 and 2019: Section "News"



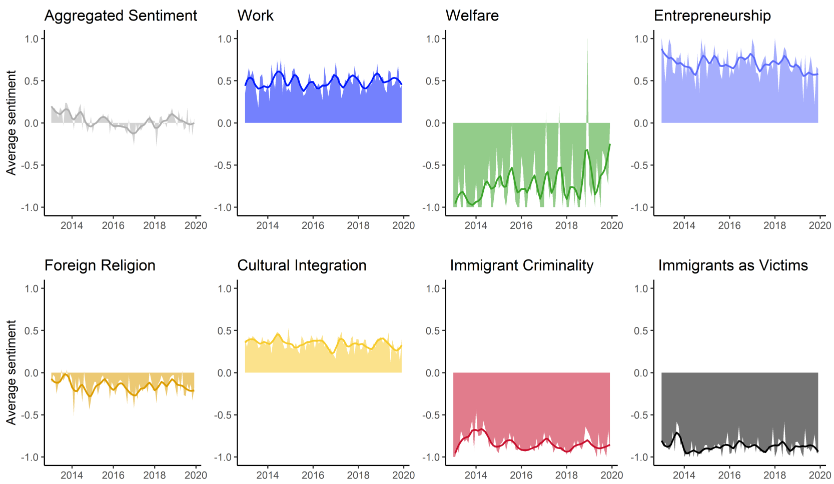
This figure shows the aggregated and theme-specific sentiments in the "news" sections of newspapers. For details on construction, please see notes to Table 8.

Figure C.8: Theme-specific Sentiments between 2013 and 2019: Section "Politics"



This figure shows the aggregated and theme-specific sentiments in the "politics" sections of newspapers. For details on construction, please see notes to Table 8.

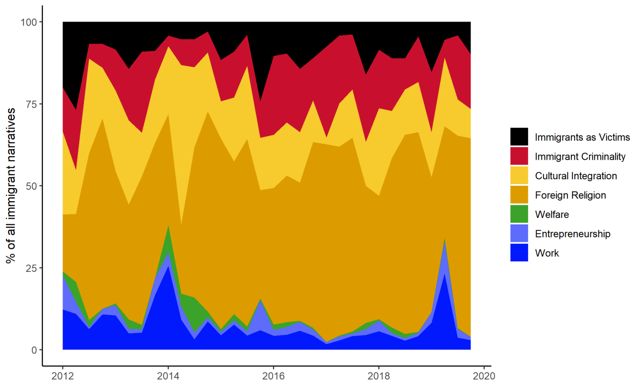
Figure C.9: Theme-specific Sentiments between 2013 and 2019: Section "Local"



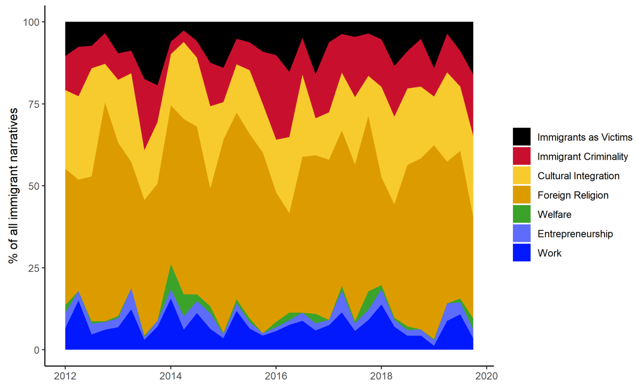
This figure shows the aggregated and theme-specific sentiments in the "local" sections of newspapers. For details on construction, please see notes to Table 8.

C.3 Heterogeneity across Space

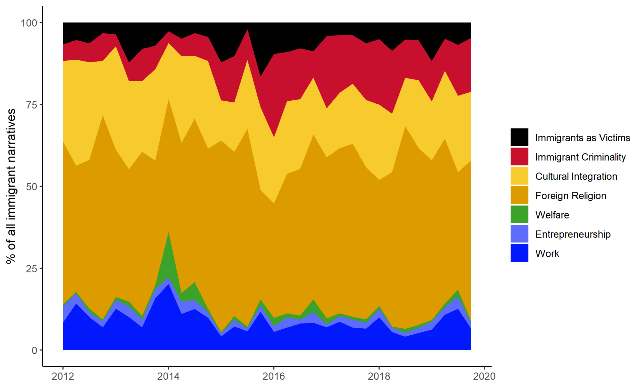
Figure C.10: Theme Shares between 2013 and 2019 across National and Regional Newspapers



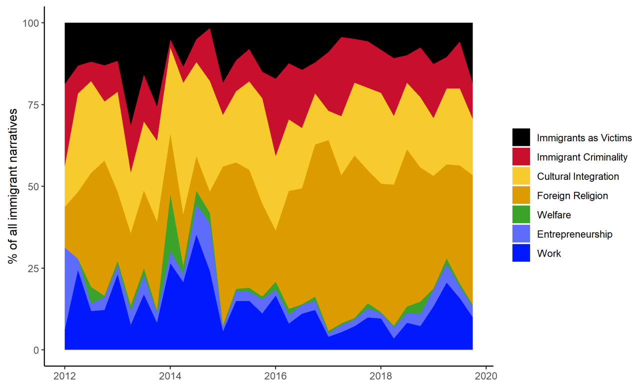
(a) Berlin



(b) National



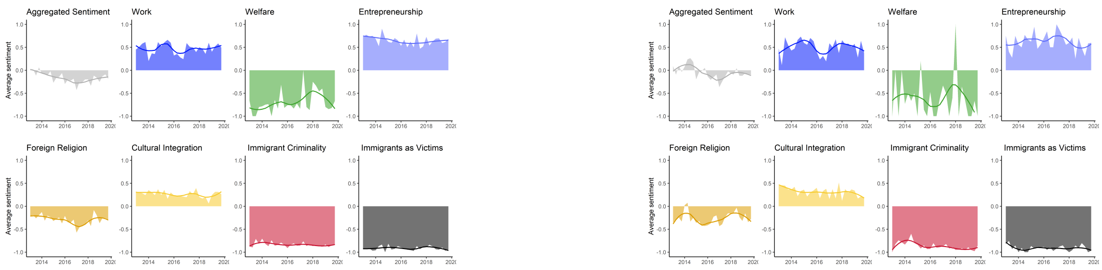
(c) West



(d) East

This figure shows the share of immigrant narratives about any of the seven themes over time, for all newspapers, by quarter for each of the four regional newspaper groups. **Panel (a)** concerns newspapers distributed predominantly in Berlin, **Panel (b)** national newspapers, **Panel (c)** newspapers from former West Germany and **Panel (d)** newspapers from former East Germany. The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative theme this may exceed the number of sentences carrying narratives.

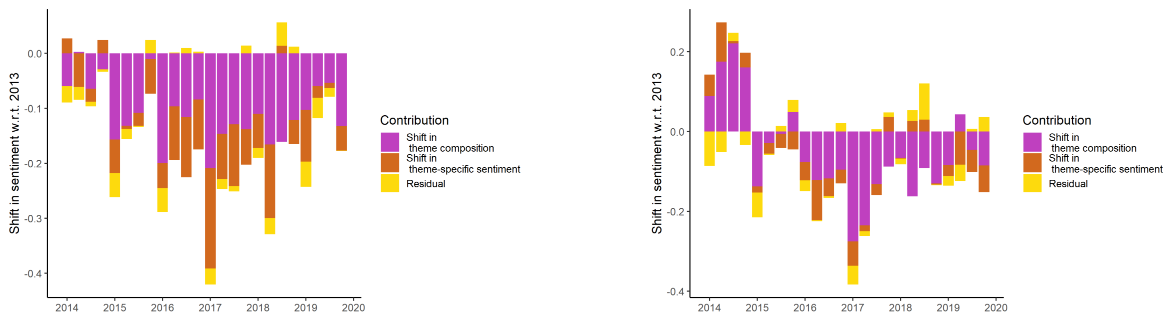
Figure C.11: Sentiment between 2013 and 2019 across Newspapers from East and West



(a) West **(b)** East

This figure shows the aggregated and theme-specific sentiments for a balanced panel of newspapers between 2013 and 2019 from former West Germany in **Panel (a)** and from former East Germany in **Panel (b)**. For details on construction, please see notes to Table 8.

Figure C.12: Decomposition of Sentiment Shift between 2014 and 2019 across Newspapers from East and West



(a) West **(b)** East

This figure shows the decomposition of shifts in sentiments by quarter with respect to 2013 into shifts in theme prevalence, theme-specific sentiments and a residual term for a balanced panel of newspapers between 2013 and 2019 from former West Germany in **Panel (a)** and from former East Germany in **Panel (b)**. For details on the decomposition procedure, please see notes to Table 11d.

Appendix D Decomposition of Sentiment

In the following, we provide details on how we decompose changes in sentiment into three distinct contributions (as shown in Figure 11d): a change in sentiment due to shifts in salience of different themes (which have distinct sentiments), a change in theme-specific sentiments, and a change in sentiment unexplained by these two shifts. The total sentiment change is the change in the overall sentiment between t and 2013:

$$\Delta S_t = S_t - S_{2013}. \tag{1}$$

The contribution of the shift between themes is calculated by keeping sentiment fixed in 2013 (S^{th}_{2013}), and isolating the contribution of changes in theme shares (f^{th}):

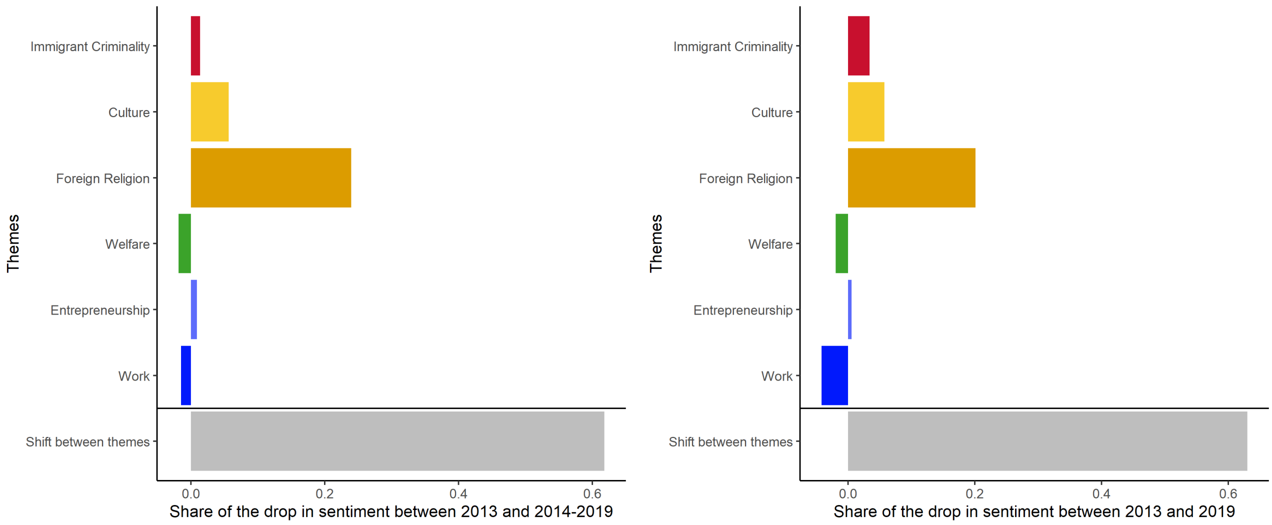
$$\Delta S^{between}_t = \sum_{th} S^{th}_{2013} \times \left(f^{th}_t - f^{th}_{2013} \right) \tag{2}$$

The contribution of the shift within themes is calculated by keeping theme shares fixed in 2013, and isolating the contribution of changes in theme-specific sentiments:

$$\Delta S^{within}_t = \sum_{th} f^{th}_{2013} \times \left(S^{th}_t - S^{th}_{2013} \right) \tag{3}$$

Following, $S_t - S_t^{between} - S_t^{within}$ is the residual component of the sentiment shift, which is driven by correlations between shifts between and within themes and by the fact that we truncate theme-specific and average sentiments to negative, neutral, or positive. However, in most time periods this contribution is small, as shown in Figure 11d. Figure D.1 shows the decomposition of the worsening of sentiment during and after the refugee crisis. Panel a) takes t as the average over all years 2014-2019 (which roughly coincides with the refugee crisis and its aftermath) and Panel b) the year 2019. We find a similar picture in the decomposition for both time periods: shifts between themes explain more than 60 percent of the drop in sentiment, whereas the worsening of sentiment of the theme *Foreign Religion* explains somewhat more than 20 percent. Moreover, comparing 2019 to 2013, sentiments regarding *Welfare* and *Work* improved, counteracting the deterioration in overall sentiment.

Figure D.1: Decomposition of Sentiment Changes



(a) Between 2013 and 2014-2019

(b) Between 2013 and 2019

This figure shows a detailed breakdown of the sentiment decomposition shown in Figure 11d, where shifts in theme-specific sentiment shown by theme. **Panel (a)** shows the decomposition between the whole period between 2014 and 2019 with respect to 2013. An example of how to interpret the diagram: more than 60 percent of the drop in sentiment is explained by shift in theme prevalence, whereas more than 20 percent is explained by a decrease in *Foreign Religion* sentiment. **Panel (b)** shows the same for 2019 with respect to 2013.

Appendix E Human Coder Instructions

Coder Instructions

We ask you to classify and rate sentences coming from German newspaper articles that cover the time-span 2000-2020; articles concern migration and migrants in Germany. You were provided with an excel file in which every row is dedicated to a sentence derived from a newspaper article. The sentences are displayed in the same order as in the article , e.g. if *article 1* is constituted of *20 sentences*, the first 20 rows in the excel will contain *article 1*'s sentences, in order. Each row includes the following information:
– the article and sentence ID;
– the text of the sentence;
– one column for each migrant category and sub-category to which you have to assign the sentence during the rating process,
– one column for sentiment rating that you will fill in during the rating process.

For each sentence, you should follow these three steps:
– identify whether the sentence is about immigration and/or immigrants;
– if so, assign the sentence to one or more categories explained in detail below;
– rate the sentence as being negative (-1), neutral (0) or positive (1).

The rating process should follow these steps and be iterated for every row of the file provided:
1. in order, read the first sentence in the first row of the excel file;
2. if the text is about immigrants or immigration, fill in the cells relative to the categories:
   - write 1 if you think the sentence's content belongs to the specific category;
   - leave the cell *blank* if the content is not about the specific category or the sentence is not clear to you;
3. fill in the cell relative to the sentiment score:
   - write -1 if you think the sentence's content is negative, 0 if neutral, 1 if positive;
   - leave the cell *blank* if the sentence is not clear to you.

The categories to which you have to assign the sentences are:

1.1 **Economy - work:**
– immigrants' role on the job-market - interested in work or not;
– portrait of immigrants' qualifications for the job-market.

1.2 **Economy - entrepreneurship:**
– immigrants as entrepreneurs (creating jobs);
– establishment of businesses/companies by immigrants (including restaurants).

1.3 **Economy - welfare:**
– immigrants as a burden or benefit for the German Welfare State;
– immigrants' ability to make their own living.

2.1 **Society - foreign religion:**
– mentioning of religion/religious symbols related to any religion except Christianity and Judaism;
– mentioning of actors/groups related to any religion except Christianity and Judaism (Salaafi, Islamic State, etc.).

2.2 **Society - culture:**
– immigrants' integration in the social and cultural life in Germany (language, school, etc.);
– immigrants' contribution to art, culture or associations/clubs.

3 **Migrant Criminality:**
– crime committed by immigrants or immigrant groups;
– immigrants role as criminals.

4.1 **Anti-migrant - Migrants as victims of crimes:**
– crimes committed against immigrants;
– racist crimes.

4.2 **Anti-migrant - Other anti-migrant acts:**
– discrimination against immigrants;
– discrimination against foreign religions;
– racism.

General Guidelines

1. **The aim is to capture how immigrants are described in a sentence, including certain narratives and stereotypes; we do not want you to judge or assess whether a certain narrative is correct or justified**;

2. Categorize each sentence as a **stand-alone**; **do not** take the broader context of the article into account, e.g. if a sentence portraits a negative picture of immigrants, and the next sentence argues why this portrait is wrong, we still want the first sentence to be coded as negative;

3. A sentence can portray a migrant using:
   - a **direct term** used in relation with migrants, migration or ethnicity, such as "Ausländer, Einwanderer, Migrant, Türke, kurdische Frau, Mann aus Frankreich, etc.",
   - a first or last **name** you would perceive as belonging to an immigrant,
   - any **"placeholder term"** referring to a migrant mentioned before, such as "er, sie, ihr, ihnen", or "Frau, Mann" or "40-Jähriger" etc.
     Note that ethnic German expellees from Russia who immigrated to Germany right after WW2 or before 1960s should not be considered as immigrants. Russian-Germans should be considered as immigrants if they immigrated more recently or there is no information about the immigration period.

4. Regarding the **placeholder terms**, sentences are reported following the order of the article; if the subject of the sentence is a placeholder term you can use sentences reported in the preceding cells to conclude whether the term refers to a migrant; e.g. sentence 1: "*Auch gegen **Sinti** und **Roma** richten sich ausgeprägte Aggressionen.*", sentence 2: "***Sie** neigten zur Kriminalität, meinen 58,5 Prozent der Deutschen*" [code sentence 2 as: migrant criminality][sentiment: -1]

5. **Do not** distinguish between regular immigrants, refugees, asylum seekers and irregular immigrants.

6. Sentences for which the sentiment conveyed is ambiguous should be rated as **neutral**.

7. A sentence can be classified under different categories as well as into more than one sub-category within the selected category. Be careful to only apply this if you are certain that a sentence belongs to more than one category and/or sub-category (**do not** assign a sentence to two categories because you are not sure which category fits exactly).

8. The category **Anti-migrant** is an exception to the above rule: a sentence can be either classified as **Anti-migrant** only, or **Anti-migrant** and **Migrant criminality** in certain cases.

9. **Do not** consider sentences that mention **only** a (probably) German actor or institution related to migration, e.g. "Integrationsbeauftragter", "Migrationsexperte".

Categories

The following reports instructions specific to each category, as well as examples of sentences that we already classified and scored. Read through each box very carefully.

Economy

Specific instructions:

- **Do not** classify sentences about **education** under the economy category, unless they are specifically about apprenticeships or they tie education to labor market access.
- If a sentence portrays immigrants as working/searching work this should be interpreted as **positive**; if a sentence refers to unemployment among immigrants that should be interpreted as **negative**.
- If a sentence portrays immigrants as receiving welfare benefits, this should be interpreted as **negative**..

Examples:
Economy - work:

– *'Muharrem E. lebt wie ein besserer Deutscher, mit Arbeit, Kindern, Eigentumswohnung."* [S: +1]
– *"Wir wollen, dass Menschen zu uns kommen , auch wegen des Fachkräftemangels.."* [S: +1]
– *"Ali arbeitet seit 2 Jahren bei einer Firma in der Region.."* [Sentiment: +1]
– *"40% der Jugendlichen mit Migrationshintergrund haben keine berufliche Qualifizierung."* [S: -1]
– *"Die Arbeitslosigkeit unter Menschen mit Migrationshintergrund beträgt über 10%."* [S: -1]

Economy - entrepreneurship:

– *"Der 50-jährige Türke betreibt ein Restaurant"* [S: +1]
– *"Sie gründete ein Unternehmen mit inzwischen 5 Mitarbeitern"* [S: +1]
– *"Unternehmen mit ausländischen Inhabern wollen zusätzliche Ausbildungsplätze schaffen."* [S: +1]
– *"Sein Unternehmen ging letztes Jahr bankrott"* [S: -1]

Economy - welfare:

– *"Die Zuwanderung in das deutsche Sozialsystem darf politisch keine Unterstützung erfahren"* [S: -1]
– *"In dieser Gruppe sind rund 35 Prozent durch Armut gefährdet, bei Personen ohne Migrationshintergrund sind es lediglich 10,7 Prozent."* [S: -1]
– *"Sie kann seitdem selbständig für ihren Lebensunterhalt aufkommen"* [S: +1]
– *"Sie arbeitet zwar, aber es ist ein prekärer Job"* [S: 0]

Society

Specific instructions:
In the society dimension, we capture how migrants are integrated and their efforts to integrate. We are therefore interested in their actions and their behavior.

- **Positive** examples are that migrants take efforts to learn German or speak the language, go to school or contribute to the cultural life. We are not interested in political demands and discussions about whether migrants should *get* something, i.e. rights or privileges, as opinions about that can vary according to political attitudes. *"Migranten sollen das Wahlrecht bekommen"* is a political demand and does not describe a migrant. This sentence should thus be scored **neutral**.
- What we *do* want to capture is what is *demanded* of migrants, as well as requests for migrants to take efforts to integrate themselves, as these imply poor integration and request the migrants to act. Sentences such as *"Migranten mussen Deutschkurse besuchen"* are therefore to be scored **negative**.
- If a sentence portrays a foreign religion as secular, liberal, open-minded this should be interpreted as **positive**; if the foreign religion is portrayed as anti-modern, fundamentalist, influenced by foreign countries or state power (e.g. DITIB, Erdogan etc.) and trying to substitute the rule of law with religious rules, this should be interpreted as **negative**.

Examples:
Society - religion:
- *" Muslime bleiben oft unter sich, und halten ihr eigenes Fest."* [S: -1]
- *"Berlin wird immer mehr zur Hochburg von Salafisten."* [S: -1]
- *"Sie hielt die Abhängigkeit des Verbands Ditib von der türkischen Religionsbehörde Diyanet für problematisch."* [S: -1]
- *"Ob der Islam zu Deutschland gehöre, beantwortete er (...) mit einem klaren Ja."* [S: +1]
- *"Am Tag der offenen Moschee gestern haben Berliner Muslime für den Umweltschutz geworben und Besucher in ihre Gotteshäuser geladen."* [S: +1]

Society - culture:
- *"Er spricht fließend Deutsch."* [S: +1]
- *"Die Künstlerin mit indischen Wurzeln"* [S: +1]
- *"Ausländische Kinder haben öfter Probleme in der Schule"* [S: -1]
- *"Anstatt Integration entwickeln sich Parallelgesellschaften"* [S: -1]

Migrant Criminality

Specific instructions:

- In all examples migrants are placed into a context where they are associated with crimes.
- The sentiment is negative when migrants are described as perpetrators in that context.
- If a sentence argues that criminality among migrants has decreased over the last years for instance, then this may be associated with a positive sentiment.

Examples:

- *"Die Hamburger Polizei ermittelt gegen Dmitri Kowtun wegen des illegalen Schmuggels und Gebrauchs von radioaktiven Substanzen. "* [Sentiment S: -1]
- *"Clankriminalität ist ein ernsthaftes Problem."* [Sentiment S: -1]
- *"Die meisten Migranten sind nicht kriminell"* [Sentiment: 0]
- *"Der Ehrenmord der jungen Muslima durch ihren Bruder ist tragisch."* [S: -1]
- *"Im Falle eines Terroranschlages werden die Muslime in Wiesbaden nicht in Generalverdacht genommen."* [S: 1]
- *"Nach einem terroristischen Attentat durch einen syrischen Flüchtling nahm die Hetze gegen Ausländer zu'* [S: -1]
- *"Ein syrischer Flüchtling zeigte Zivilcourage und stoppte den Dieb'* [S: 1]

Anti-migrant

Specific instructions:
This dimension captures the discourse around migrants and migration relative to the context of discrimination and crimes against immigrants: "Are immigrants victims of: crime, racism, discrimination, etc.?"

- A sentence that mentions a right-wing party or group should only be coded in the *Anti-migrant* category if it is clear that it is against immigrants.
- If a sentence is classified under **Anti-migrant**, it **cannot** be coded into another category. An exception is the **Migrant Criminality** category. A sentence is coded both Anti-migrant and Migrant Criminality if migrants are actors of crimes in reaction to anti-migrant acts, e.g. *"Auslöser für den Terroranschlag war ein Anstieg der Anti-Migrations-Demonstrationen in der Region. "*
- A sentence that portrays migrants as victims of crimes or discrimination is **negative**.
- A sentence that portrays a reduction in crimes or discrimination against migrants is **positive**.
- A sentence that highlights that migrants are not specifically victims of crimes or discrimination is **neutral**.

Examples:
Anti-migrant - Victims of crimes:

- *"Patrick E. soll auf 2 Afghanen eingeschlagen und sie rassistisch beschimpft haben."* [S: -1]
- *"Die NSU- Morde haben für viel Verunsicherung gesorgt'* [S: -1]
- *"Nach einem terroristischen Attentat durch einen syrischen Flüchtling nahm die Hetze gegen Ausländer zu'* [S: -1]

Anti-migrant - Other:

- *"Frauen mit Koptuch werden diskriminiert'* [S: -1]
- *"Ausländer werden auf dem Arbeitsmarkt nicht benachteiligt'* [S: 0]
- *"Die Autoren der Mitte-Studie" konstatieren, dass die Demokratie weiter auf einem festen Fundament steht, dass Ausländerfeindlichkeit abgenommen hat.'* [S: 1]
- *"Ausländer werden auf dem Arbeitsmarkt diskriminiert'* [S: -1]

Submission of your work

You will receive a folder containing four Excel files, organized in a specific order. You are asked to code through the files following the order of the Excel files (filename_._1 first, filename_._2 second, etc.). This is very important. **DO NOT**:

- **change the name of the Excel files;**
- **change the column names or order in the Excel files.**
- **change the content in pre-filled rows (article id, sent_i and sentence content)**
- **change the order of the rows**

Your finished work should be sent back to us **before January 5**. The University cannot pay you if something is missing or if the work has been done inconstantly (quality control).

You need to report your working hours in the given table. This is not to evaluate you; speed does not matter.

Send your finished work in a zipfile to [redacted].[redacted]@[redacted].com and [redacted].[redacted]@[redacted].de (in CC), with the e-mail subject 'Narrative workshop - Final output - {insert your given ID}'.
Your zip file should be named 'final_{your given id}.zip'

Optional intermediate submission

You can send Jonathan and Charotte the finished *first* Excel file if you want feedback on your work. **This is completely optional**.