Immigrant narratives*

Kai Gehring (University of Bern, Wyss Academy, CESifo)
kai.gehring@unibe.ch

Panu Poutvaara (University of Munich, ifo Institute, CESifo, CReAM, IZA, UCFS)
poutvaara@ifo.de

Joop Adema (ifo Institute)
adema@ifo.de

Charlotte Robert (University of Heidelberg)
charlotte.robert@awi.uni-heidelberg.de

August 23, 2022

Abstract

Narratives have been identified as a key mechanism influencing human behavior. Migration is a topic where competing narratives are particularly influential and divisive. This paper analyzes narratives about immigrants in Germany over the 2000 to 2019 period, drawing on 107,428 newspaper articles from 70 regional and national newspapers. Using a combination of newly developed theme-specific dictionaries with advanced natural language processing tools allows us to create a comprehensive data-set of immigration narratives in 7 distinct themes with a theme-specific sentiment.

To evaluate the quality of our algorithm as a method to measure narratives, we then use 16 human coders who manually code more than 1,700 articles containing around 75,000 sentences. One important insight is that among humans there is considerable disagreement how to classify a sentence and how to assess its sentiment. Using sentences where humans agree to evaluate the quality of our algorithm, it clearly outperforms simple word-matching methods and sentiment dictionaries in sentence classification and sentiment assignment. The final data-set allows several interesting conclusions.

The final data-set reveals that while economists often focus on the economic consequences of immigration, our results highlight that narratives regarding foreign religion and cultural integration are much more frequent. Narratives related to immigrant criminality and foreign religion are more negative, whereas economy and cultural integration are predominantly positive. We further highlight differences across space and examine how important events influence the composition and sentiment of narratives.

Keywords: *Narrative Economics; Immigration; Media; Newspapers; Voting.*
JEL Classification: *F22; J15; C81; Z13; D72.*

1 Introduction

Recent studies reveal how individual narratives and their spread can crucially affect personal, economic and political decisions and outcomes (Shiller, 2017). Those outcomes range from inflation expectations (Andre et al., 2021) to racism (Esposito et al., 2021) and behavior during the COVID-19 pandemic (Bursztyn et al., 2020). Mass media both reflects and shapes societies, forming and spreading the narratives about a particular topic (Ash et al., 2022; Durante, Pinotti and Tesei, 2019; Durante and Zhuravskaya, 2018; Zhuravskaya, Petrova and Enikolopov, 2020). Media slant affect people's voting decisions (DellaVigna and Kaplan, 2007; Gerber, Karlan and Bergan, 2009), and newspapers specifically have been show to influence turnout in American elections (Gentzkow, Shapiro and Sinkinson, 2011).

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

Immigration is one of the most controversial topics in the media, and a particularly suitable example where various competing narratives influence how people interact with immigrants and vote on specific policies. Worries about immigration constitute a key political cleavage in all Western democracies (Gethin, Martínez-Toledano and Piketty (2022)) that has fueled the rise of far-right parties across Europe (Otto and Steinhardt, 2014; Halla, Wagner and Zweimüller, 2017; Edo et al., 2019; Hangartner et al., 2019) and played a major role in the Brexit referendum and in Donald Trump's election as US president (Norris and Inglehart, 2019). Examples of competing narratives centering around immigrants reflect for instance them being a burden to the welfare state in contrast to their stimulating role as entrepreneurs. Immigrant narratives are diverse, but can be assigned to clear themes related to areas like the economy or crime. Existing papers usually focus on a single crucial event or individual competing narratives. Given the importance of the topic and the crucial role of narratives, it is important to measure immigrant narratives more comprehensively and understand their relationship with societal conditions and role in the media.

In this paper, we propose a novel way to study narratives using text-as-data, and apply it to media narratives about immigrants. 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.

To provide a comprehensive data-set 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 of five main narrative themes 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 the following main themes: *Economy* (subdivided in Work, Welfare and Entrepreneurship), *Foreign Religion*, *Cultural Integration*, *Immigrant Criminality* and *Anti-immigrant*. We apply our method to newspaper articles from five national and 65 regional newspapers available in the *Factiva* database. After an initial pre-selection, we downloaded 107,428 articles about immigrants in Germany over the period from 2000 to 2019.

Our current main results are the following. First of all, we find that descriptively *Economy* was the subject of only 13 percent of narrative sentences, and about 12 percent relate to *Immigrant Criminality*. Instead, a vast majority of immigrant narrative sentences relates to the themes *Foreign Religion* and *Cultural Integration*, (respectively 45 and 22 percent of all sentences in articles published from 2000 to 2019). 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. Our data-set allows us to propose a method to decompose sentiment shifts in compositional effects and within-theme shifts. We then, thirdly, analyze three important immigration-related events in recent German history and their influence on immigrant narratives.

Usually economic studies using text-as-data have no systematic assessment of the quality of the output and measures. One of the innovations of our paper is to also investigate heterogeneity in the human understanding of narratives and to compare the quality of our method both with simpler solely dictionary-based approaches. Our method is somewhere between traditional simple word-matching and black-box supervised machine learning approaches. Developing specific functions and combining NLP tools with dictionaries require effort, hence it is necessary to evaluate if and which parts of the effort are justified by improved performance. While our data-set can be used and applied to many research questions within the specific German context, this investigation helps to understand the general potential of such an approach better.

For this purpose, we use 16 human coders recruited among native German-speaking university students from different parts of Germany and trained them intensively for the task. Each of them coded a batch of 437 articles, which equals around 18,000 to 20,000 sentences. We use that sample to study heterogeneity among human coders and 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 newly created immigrant 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.

Regarding heterogeneity in the human understanding of narratives, we also discover some interesting insights. We document a large heterogeneity across human coders for classification and sentiment assignment, which differs between themes. Complete alignment among humans is the 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. Nonetheless, it highlights a potential problem for supervised machine learning approaches that require a ground truth to learn from. Our algorithm tends to be well within the range of human assessments, usually classifying slightly more conservatively below the human average. Using a dictionary-based approach in combination with modern NLP features can thus be a good compromise for applications like this, allowing transparency and an assessment that is close to the average human.

Our main contributions are thus the following. First, from a methodological point of view we develop new immigration-specific dictionaries and test and implement a novel combination of tools that delivers promising results. Second, 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 our data and methodology. Third, we provide a large data-set of immigrant narratives in seven themes and their sentiment for German national and regional newspapers. Fourth, we demonstrate how that data-set can be used to understand both larger trends over time and across space, as well as analyze individual events in detail.

Our paper contributes to three different strands of literature. First, it contributes to the strongly growing literature on the economics of narratives, sparked among others by Shiller (2017). Narratives are simple stories or fragments relating to such stories that do not just transport neutral, objective information, but evoke an emotion, opinion or sentiment. They can be passed on over many generations as part of religion, mythology, fairy tales or more generally folklore (Michalopoulos and Xue, 2021). Narratives can have a crucial role in shaping human incentives (Bénabou, Falk and Tirole, 2018). Bursztyn et al. (2020), 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 particular narrative spread via a movie crucially affects public opinion and race relations in the US. In contrast to those papers, we can measure narratives comprehensively across seven immigrant themes and decompose the effect of important events in shifts across themes and within-theme sentiment changes.

One aim of this paper is to combine tools in a novel way to create the most comprehensive data-set and measures of immigrant narratives and their sentiment. Hence, we also relate to an important and growing methodological literature aiming at defining and measuring narratives. Like Ash, Gauthier and Widmer (2021) we analyze individual sentences that contain shorter narrative fragments 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 and we define narratives in a broader way beyond simple cause and effect statements (in line with Shiller (2017)).[1] We do find that narratives do reflect societal conditions and are influenced by important events, corresponding to the results from Michalopoulos and Xue (2021) that folklore stories reflects the respective environment groups of humans reside in. [2]

Second, studying narratives is part of the more general attempt in economics to integrate group-level behavioral phenomena like culture (e.g., Guiso, Sapienza and Zingales, 2016), moral values (e.g., Enke, 2020), and group identity Akerlof and Kranton (2000) into economics. While behavioral economics and lab experiments highlighted the importance of individual-level psychological biases, many important psychological mechanisms can only be understood at the group level. For instance, economists started to study the origins of group identity (Dehdari and Gehring, 2022), its influence on political preferences (Gehring, 2021; Fouka, 2019) and how current events can shape it (Depetris-Chauvin, Durante and Campante, 2020; Gehring, 2022). Narratives, like propaganda, can be understood as a technology that shapes group identity, moral values and culture. While this is beyond the scope of this paper, our data-set provides the basis for investigating the relationship between narratives and political and economic outcomes in more detail.

Third, we contribute to the literature on media economics. While Gentzkow, Shapiro and Sinkinson (2011) studies newspaper markets, other studies also look at the relationship between different types of media. (Cagé, Hervé and Viaud, 2020). 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., 2020; Campante, Durante and Sobbrio, 2018; Galletta and Ash, 2019; Kearney and Levine, 2015). Some important papers examine social media specifically (Enikolopov, Petrova and Zhuravskaya, 2011; Cagé, Hervé and Mazoyer, 2020; Zhuravskaya, Petrova and Enikolopov, 2020), others focus on newspapers (Besley and Burgess, 2002; Snyder and Strömberg, 2010). 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. (2022), who focus on specific coverage of immigrant criminality using a dictionary approach, by capturing immigrant narratives comprehensively across 7 themes.

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. Our results also highlight that at least in mainstream media, still a major source of information for voters, non-economic topics play a crucial role when discussing immigration. This links our

[1] Given the rapidly evolving nature of the field, it is impossible to give a fair overview of all new development. There are many exciting developments that integrate machine-learning. While we are sceptical about the possibilities to fully automate the detection of narratives in a way that is transparent and replicable, supervised machine learning using neural networks has the potential to provide measures in large data-sets (e.g. Card et al., 2022).

[2] Narratives naturally relate to the culture of a society. As Giuliano and Nunn (2021) show, transporting culture across generations is the more valuable, the more stable the environment. This is in line with our results of narratives reacting to certain salient events, which they need to to act as a technology to communicate an updated view of the world.

research to papers studying the effects of immigration on voting. Alesina and Tabellini (2022) whose synthesis of existing results also indicate that culture and ethnic background play an important role in determining the policial reaction to immigrants.[3]

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 book and this initial definition, the study of narratives has been gaining attention in the popular press, bestseller books like Harari's "Sapiens" (2016), as well as in the economics literature. Since the work by Shiller on narrative economics, both theoretical (Bénabou, Falk and Tirole, 2018) and empirical (Esposito et al., 2021; Bursztyn et al., 2020) 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 complete causal statements as one type of narratives.

The reason for choosing a broader definition of narratives is based on our aim to study migration 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' perception and decisions. 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 sentiment. Instead, journalists often shape narratives by selecting which facts and news to report how intensively, often using more subtle distinctions to frame them as positive or negative. 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 of immigrant reality.

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.[4] 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 advan-

[3] Important studies looking at the effects of immigration on voting are (Otto and Steinhardt, 2014; Barone et al., 2016; Halla, Wagner and Zweimüller, 2017; Dustmann, Vasiljeva and Piil Damm, 2019; Edo et al., 2019; Hangartner et al., 2019; Steinmayr, 2021) and for general support for redistribution (Dahlberg, Edmark and Lundqvist, 2012; Alesina, Miano and Stantcheva, Forthcoming).

[4] 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.

tage of defining a few clear themes is their transparency and clear relationship with existing quantitative and qualitative research strands, while automated topic-models suffer from the sensitivity to set parameters and provide useful interpretations of generated topics. We outline and explain our seven themes in more detail below.

We refer to *narrative fragments* - building on Ash et al. (2020) - because narratives can be 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 use both words interchangeably in the remainder of the paper.

Technically, two main approaches to measure narratives and their sentiment in texts can be distinguished, manual dictionary methods and machine learning (ML). 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 (topic,...) and sentiment is 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 tasks like text classification out-of-sample. ML also empowers new NLP packages (like the Python package *Spacy*) that can automatically understand and code linguistic features of words and sentences.

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 number 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 to measure immigration narratives, our own initial manual coding of a substantial amount of articles and sentences suggests a large heterogeneity among humans in classifying a sentence and assigning a sentiment. This is one main reason for initially opting against supervised ML. The second is that supervised ML clearly excels at tasks like general sentiment and classification of binary categories, but we were not sure whether our goal to classify sentences into seven narrative schemes and assign a theme-specific sentiment was too demanding. Instead, we opt in favor of a more simple approach that combines 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.[5]

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 precisely than simple dictionary methods. Figure A.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 these steps:

1. Identify which articles are about immigrants in Germany by using a comprehensive filter on a newspaper database, retrieving articles and removing remaining irrelevant articles using NLP tools and lists of locations in Germany (see details in the data section 3.1). Within those articles, identify which sentences are about immigrants based on keywords from custom dictionaries and NLP tools. (see section 2.2.2).

2. Categorize sentences into themes and assigning a negative, neutral or positive sentiment to each theme-sentence using dictionaries and NLP tools. NLP tools and self-constructed evaluator/negation/qualification functions (see section 2.2.3).

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. "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 1000 articles, plus additional selective articles for specific lists.

  Examples of dictionaries are common migrant 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 lists, e.g., *find/lose jobs,benefit from/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 100 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 sample sentences, we validated with three human coders that reversing or adjusting brings the sentiment closer to human judgement.

[5] 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/`.

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 2, 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 by simply matching with a dictionary of immigrant terms (”Ausländer/foreigners, Muslime/muslims”). These are assigned to be about ”cultural integration” as it mentions integration and a lack of education. The negation tool ensures that they are classified as negative instead of positive, and as neutral instead of negative in the second case.

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.

Such constructions are quite common in newspapers. 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 these final two examples:

"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 have increased dramatically.

Recognizing this sentence to be about refugees is easy, and could be achieved by having a simple immigrant term dictionary that contains the word. Assigning it to be about crime is also relatively straightforward, as it contains the word for "criminal offense" ("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. Hard-coding now means 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 the anti-migrant theme 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 approach theme classification and sentiment assignment.

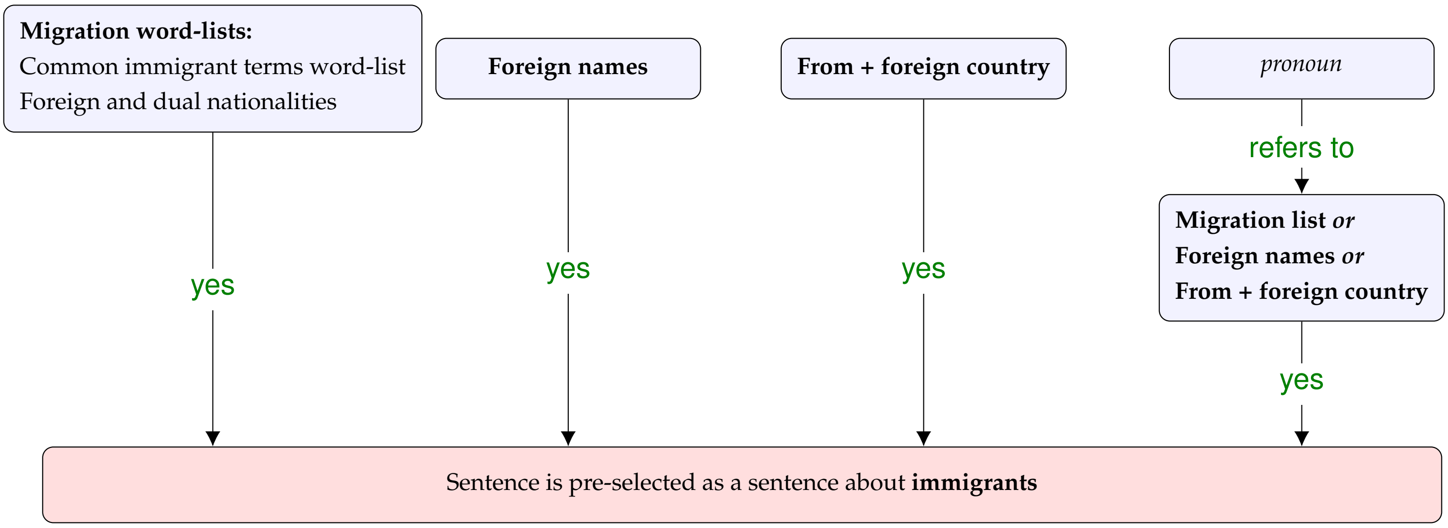
2.2.2 Identifying immigrant sentences

To capture narratives fragments at the sentence-level, the first general step in methodology is to select sentences relating to immigrants in Germany. As outlined above, a 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.[6] 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.[7]

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

[7] Foreign names are identified using *spaCy*'s *entity recognition* tool to identify *person entities* in the sentence. If these entities do not intersect with a manually collected list of German last and first names gathered from *www.behindthename.com*, they are considered as foreign names. We exclude publicly known German and foreign individuals.

Figure 2: Immigrant Sentence Selector



2.2.3 Classifying into themes and assigning sentiment

We classify sentences in four broad themes: Economy, Crime, Society and Anti-Immigrant. The economy theme is further divided in three sub-themes (labor market, welfare state and entrepreneurship) and the society theme into migrant crime and cultural integration. 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 somehow, but illustrates sufficiently how the different dictionaries and tools from A.1 are combined for classification and sentiment. 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 intensive 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 that describe how the host society is treating immigrants. Those can be covering *Cultural Integration*, *Economy* or *Immigrant Criminality*, with the distinct feature that immigrants are not the subject, but rather the object of a narrative. To capture that we use various features, maybe most importantly the Part-of-Speech (POS)-tags that allow recognizing exactly that. To give one example, we distinguish if immigrants are the victims or the perpetrator of a crime or discriminatory act.

Figure 3: Themes: Economy, Foreign Religion, Cultural Integration

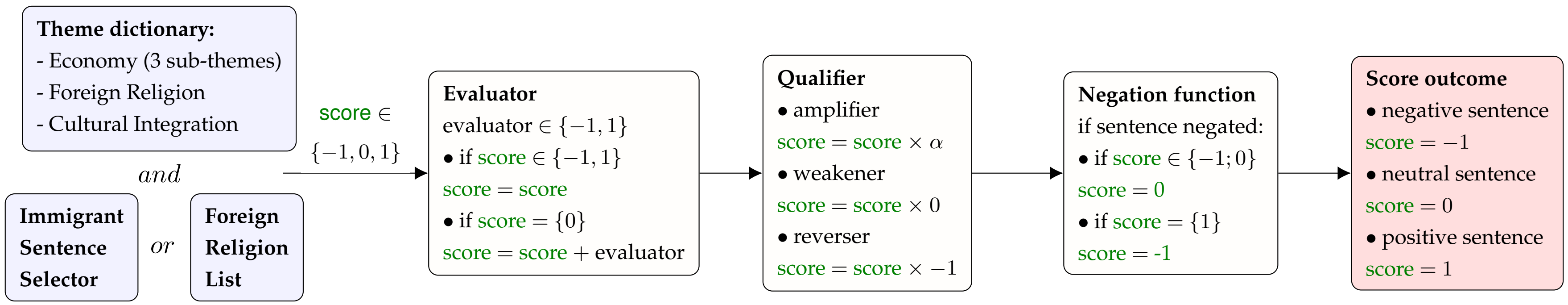
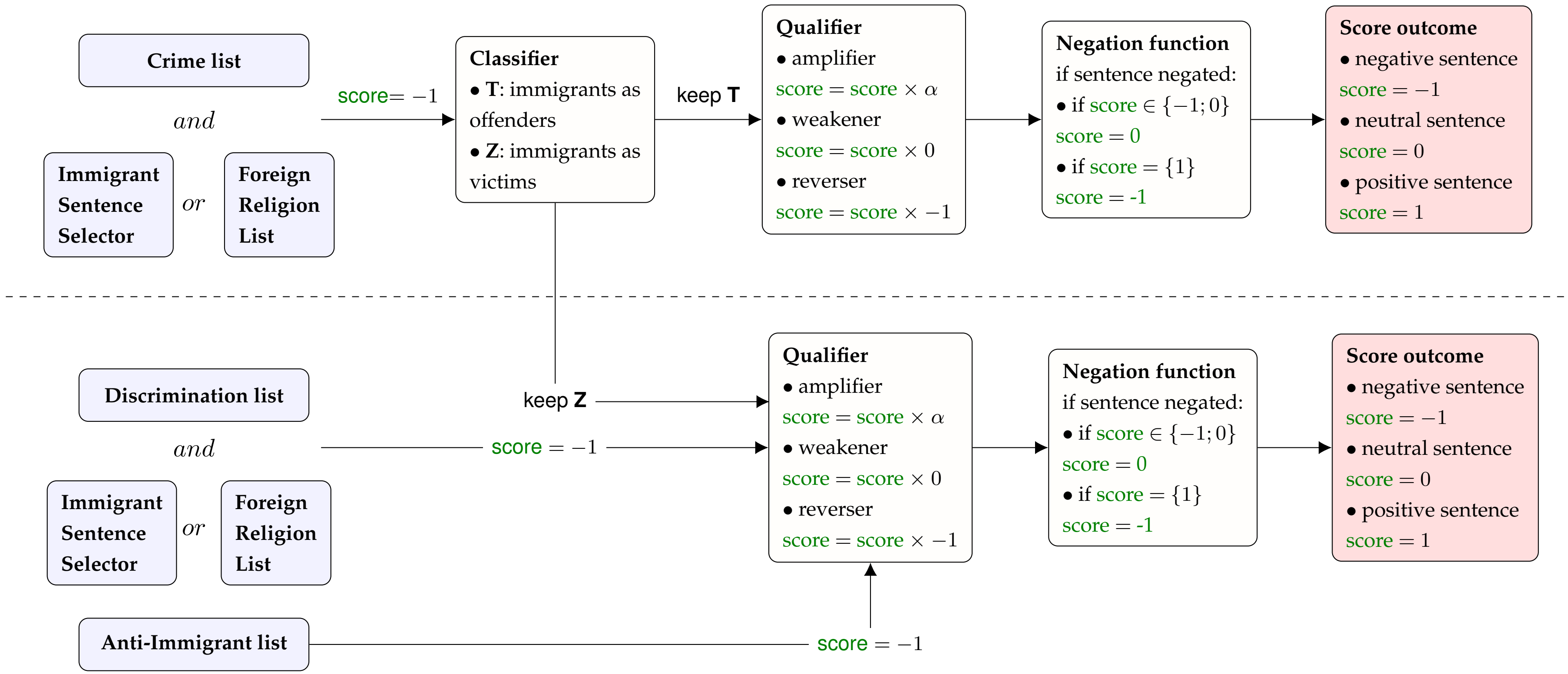


Figure 4: Crime and Anti-Immigrant



3 Data

3.1 Source of newspaper articles: Factiva

We obtain individual articles about immigrants published by German national and regional newspapers from Factiva, an international newspaper database (, 2020).[8] Figure 5 outlines the steps taken to obtain relevant articles. We begin by querying articles using a Boolean search filter (logical AND, OR and NOT operators) that combines immigrant-specific search terms with a geographic location within Germany consisting (logical AND, OR and NOT operators) by newspaper.[9] It excludes certain terms that seem to create many false positives in a careful manual test.[10] Using this approach, we obtain 136,170 articles from 88 newspapers. This includes four national newspapers, that occupy different positions along the political spectrum, and two major national weekly magazines.

We constrain the sample of articles in three ways: we omit articles from newspapers with less than 100 articles, articles from 2020 (because of differential delay of articles by newspaper in Factiva and the COVID-19 pandemic) and articles that are probably not about Germany using an article-level procedure[11]. This procedure identifies 17 percent of the articles not to be about Germany. The resulting data set comprises 107,428 newspaper articles from 5 national (Süddeutsche Zeitung, Die Welt, Der Spiegel (weekly), Die Zeit (weekly), and BILD) and 65 regional newspapers. These 107,428 articles constitute our core sample on which we apply the process described earlier, aiming to capture how prominent the main themes we defined are in newspaper reporting about immigrants.

For an overview of newspapers in scope, the total number of articles, what share of articles have at least one sentence related to any of our themes, the share of articles obtained from a newspaper agency such as Reuters or the German Press Agency (DPA), see Table B.1. On average, 80 percent of

[8] Factiva includes 5 national and 65 regional German newspapers, listed in the appendix. Although Factiva contains identifiers for 89 newspapers, we do not consider each of these as individual newspapers. We define a newspaper as an entity that has at least a partial editorial office. First of all, some newspapers are included under multiple identifiers, which are combined. Secondly, some newspapers are published under different local names but largely contain the same content and do not have independent editorial offices. To determine whether we analyze a newspaper separately or not, we collected information about editorial offices from the official websites of the newspapers to determine. For example, the *Passauer Neue Presse* comprises 20 local editions, published under different names throughout Eastern Bavaria. We treat it as a single newspaper, although the local editions may vary. In addition, many newspapers fall under a single publishing house and there is strong co-operation between newspapers, sharing a ”mantel” (English: cloak) for extra-regional news. However, they may have independent editorial offices for the local parts of the example. As another example, we treat *Fürther Nachrichten* and *Erlanger Nachrichten* because of independent editorial offices as two separate newspapers, although they clearly cooperate on non-local content.

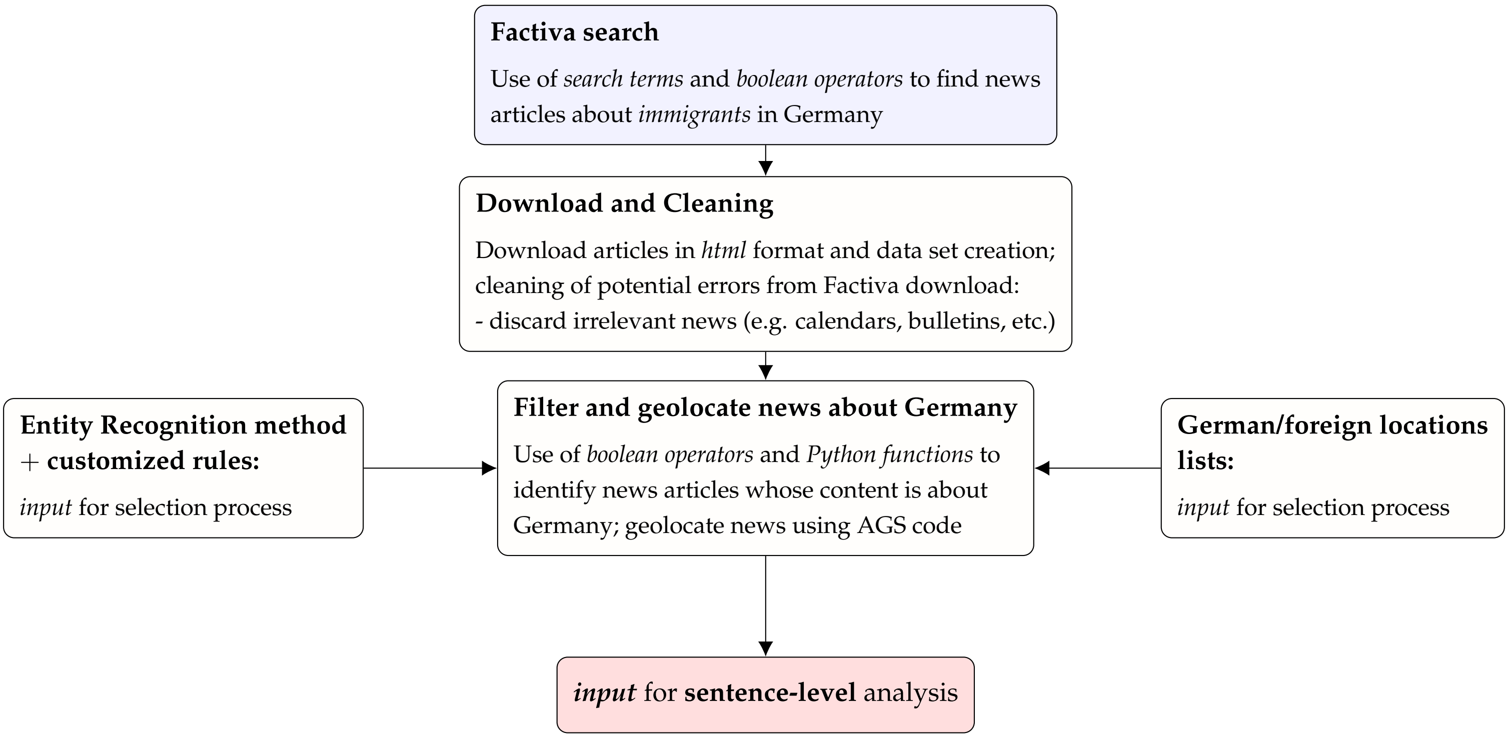
[9] This includes all words starting with: Einwanderer (immigrant), eingewandert (immigrated), Migra (starting with migra), or Ausländer (foreigner). Location terms are Germany, all 16 German federal states (German: *Bundesländer*), the 50 biggest cities in Germany and “Ruhrgebiet”, the term for a large conglomerate of cities in the West of Germany, home of many immigrants. For the full filter, see Appendix A.

[10] These exceptions include search terms that by coincidence also correspond to other terms that yield many false positives, related to event-calendars, sport results, or COVID-19. Due to limitations of the Factiva portal, we obtain a maximum of 1,500 newspaper articles per query. Factiva does not allow systematic webscraping and limits manual access to search results to a specific number per newspaper. This threshold exceeded 1,500 for most newspapers. We therefore manually download up to 1,500 articles per newspaper per time period. The articles are sorted according to relevance, which is based on a sorting algorithm in Factiva that is the same for each newspaper. For all national newspapers, we subdivided the period 2000-2020 in 2 periods to obtain up to 3,000 articles per regional newspaper and in 4 periods to obtain up to 6,000 articles per national newspaper.

[11] 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.

all articles have at least one sentence containing one of the themes and 14 percent of all articles are written by newspaper agencies. Across newspapers, the share of articles with at least one sentence about any of the themes do not differ much, but are somewhat higher for newspapers with more articles due to the Factiva relevance filter. As newspapers with more articles are more likely impacted by the download limit of 1,500 articles, these contain more relevant articles and more sentences are classified. The share of articles from agencies varies much more across newspapers and time.

Figure 5: Steps to construct data set



3.2 Source of newspaper subscriptions and background information: IVW

We use municipality-level data by the *Informationsgesellschaft zur Feststellung der Verbreitung von Werbeträgern e.V.* (IVW), by the German Audit Bureau of Circulation (IVW), an independent auditing organization that records and certifies circulation German newspaper data.[12] Beyond general distribution zones, we reached a bilateral agreement to get access to municipality-level sales data (*Gemeinde*) for the year 2019, covering 98 percent of German municipalities.[13] 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.[14] Having only 2019 data is less problematic for the geographical reach of regional newspapers, which changes little over time. The precise market share of newspapers is more likely to fluctuate over time, which could be more or less problematic depending on the type and source of endogeneity.

[12] 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.

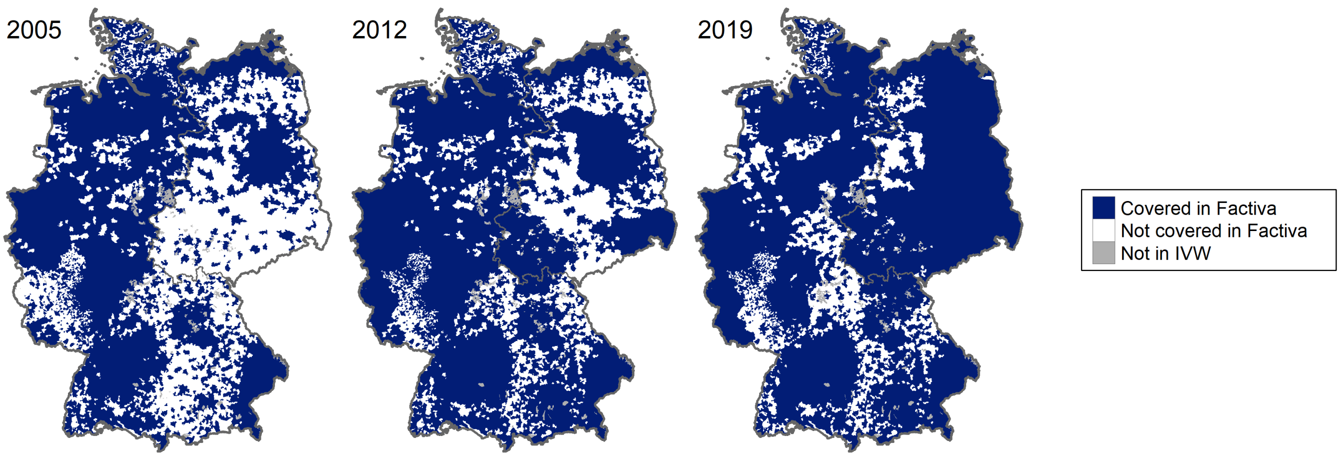
[13] IVW measured sales in the reference week of 4th to 10th of November 2019. This includes subscriptions and single copy sales. It includes electronic sales only for dew selected newspapers.

[14] 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.

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 good already in 2005, but initial coverage is limited in the Eastern states of Sachsen and Thüringen. 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 onward. If we include national newspapers, over 90% of municipalities are covered from 2012 onward. Appendix Figure B.3 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.4 displays the coverage for all individual newspapers over time.

Figure 6: Geographical coverage of Factiva articles from regional newspapers over time



The maps above show the municipalities for which our Factiva data set contains articles from regional newspapers in the years 2005, 2012 and 2019. Our data set covers 47 percent of German municipalities in 2005, 66 percent in 2012 and 73 percent in 2019. Those municipality 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 data-set. 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 about the migrant share, unenmployment or the share of moslems in a municipality.[15]

[15] 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 Theme composition and sentiment of immigrant narratives

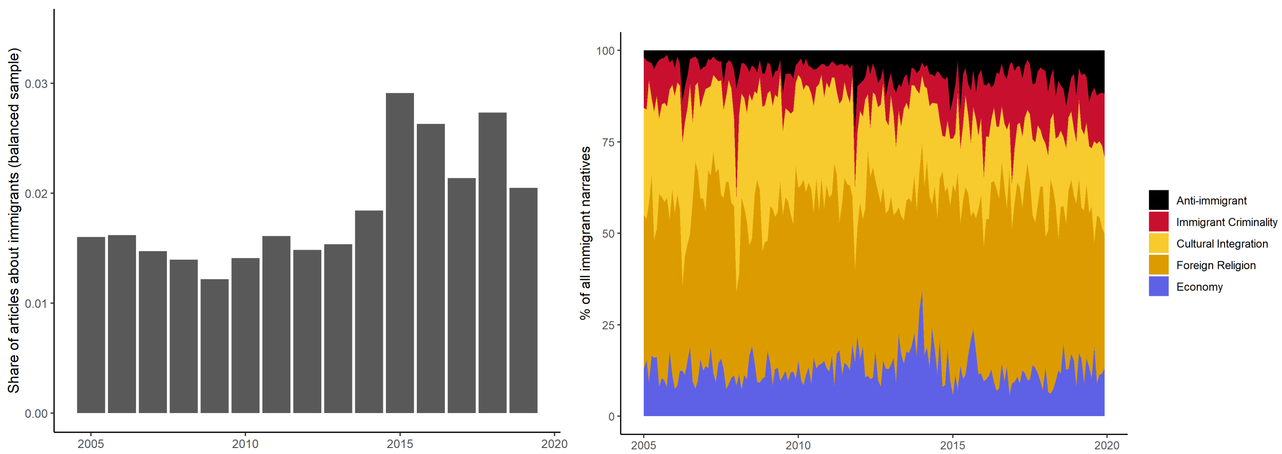
ALTERNATIVE VERSION (ONE INTRODUCTORY PARAGRAPH AFTER SECTION TITLE):

Our new data-set 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, immigration is a key political issue.

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.[16] The first panel of Figure 7a shows the share of articles about immigrants by year. The salience of immigration remains relatively flat from 2005 to 2014, averaging at ADD NUMBER, 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 ADD NUMBER from 2016 to 2019.

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



(a) Share of articles about immigration

(b) Shares of narratives by theme

Panel (a) shows the share of all articles in Factiva that contain at least one immigrant narrative within one of our predefined themes, by year. To ensure comparability over time the only newspapers used to construct this figure are the six newspapers that are available in Factiva since 2000.
Panel (b) shows the share of immigrant narratives about any of the five main themes over time, for all 70 newspapers. The total number of immigrant narratives is the sum of all narratives identified. As multiple sentences can contain more than one narrative, this is not the same as the number of sentences carrying narratives.

Our second major 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 dif-

[16] 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 ADD NUMBER newspapers.

ferent 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 in different themes. 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. Composition and sentiment is of course in a direct sense that of the journalists writing the articles, but newspapers also plausibly reflect the concerns of their readers.

Figure 7b shows the composition of immigrant narratives by month for all 70 newspapers.[17] 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 two decades, out of all themes identified in our sample, 12 percent concerned the *Economy*, 45 percent *Foreign Religion*, 23 percent *Cultural Integration*, 12 percent *Immigrant Criminality* and 7 percent *Anti-Immigrant*.

4.2 Narrative sentiment by theme over time

Our third major descriptive question is how average sentiment and theme-specific sentiment changes 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, but whether it 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 economy, foreign religion, cultural integration, and immigrant criminality. We do not include anti-immigrant narratives in this aggregation as this is about how recipient society treats immigrants, rather than narrative about immigrants. From the year 2000 to 2014, the average sentiment is close to zero, ranging from -0.08 in year 2005 to 0.03 in year 2000. 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 IS terrorist attacks).

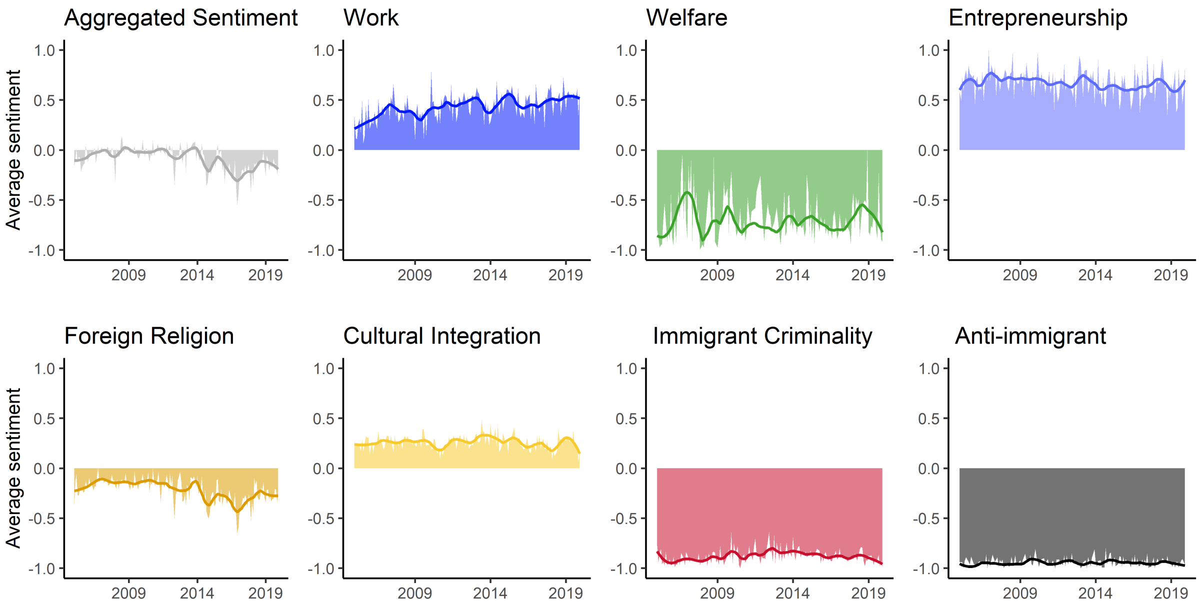
Next, we consider theme-specific sentiment. We begin with the *Economy* sub-themes *Work*, *Entrepreneurship* and *Welfare*. The average sentiment related to work and entrepreneurship is always positive, and that related to welfare always negative. Average weighted sentiment related to *Econ-*

[17] We also analyzed themes over time separately for national and regional newspapers. As shown in Appendix Figure C.2, themes develop quite similarly in national and regional newspapers over time. Appendix Figure C.2 shows the equivalent of Figure 7b for the five national in panel **a)** and 75 regional newspapers in panel **b)**. On average, the share of themes is comparable. However, the opening of labor markets to Bulgarian and Romanians in 2014 was relatively more important in regional newspapers. Furthermore, we analyzed themes by weighting each sentence by the sales of the newspaper it appeared in. Appendix Figure C.3 shows that the resulting pattern of themes over times is near indistinguishable from that shown in 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 to Figure 7b.

omy is predominantly positive.[18]

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 2000 and 2019 was -0.20. The average sentiment related to foreign religion ranged between years 2000 and 2014 between -0.22 and -0.07. From 2015 to 2019, the average sentiment ranged between -0.47 and -0.34. The average sentiment related to other aspects of *Cultural Integration* than *Religion*, instead, is positive in every month, with average annual sentiment between 2000 and 2019 being 0.24. Sentiment in sentences related to *Immigrant Criminality* and *Anti-immigrant* narratives are almost always negative, with average annual sentiment for *Immigrant Criminality* ranging between -0.91 and -0.73, and for *Anti-immigrant* between -0.95 and -0.84.

Figure 8: Theme-specific sentiments over time (monthly)



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). 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 smoothened local regression line (Loess) with $\alpha = 0.15$.

5 Human Validation

Our goal is to provide a data set 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 if the sole goal is to measure an average relationship in a regression (given sufficient observations). However, such

[18] 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.

methods may not be able to capture (the sentiment of) narratives over time and space.[19]

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 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 data set. The human-coded articles serve as a benchmark to estimate the performance of our algorithm.

2. We study the share of sentences classified to be about immigrants, 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 hired 20 students from 5 German universities to code more than 2,000 articles in a way comparable to our algorithm.[20] To understand the extent to which there is heterogeneity in human perception of narratives and sentiments, each article and sentence was assigned to four coders. 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 (ii) 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 unless an answer would have biased the coding. The ex-post verification suggests no coder used simple random patterns and that there was some fatigue, but with negligible impact.[21]

[19] 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.

[20] 20 out of initially 24 completed the task. The students were remunerated at a net hourly wage of €15. All but two of the 20 students who completed the task consented to provide personal background information. Out of the remaining 18 students, 9 are female, 9 are male, and their age ranges from 19 to 27, with a mean of 22.4. Four of the students have a migration background.

[21] The codebook and further instructions can be found in Appendix F. 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. We ran several additional quality checks to investigate fatigue and sub-batch consistency. We find evidence for some fatigue, but it translates only in negligible differences. For an assessment of the determinants of coding of their answers, see Appendix E.

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. As some students dropped out, not all six batches were evaluated by four students. Therefore, we focus on the four batches comprising 1,748 articles (1.5 percent of our data set) that were evaluated each by four students in subsequent analysis.

The first aspect we are interested in is to understand differences among humans in classifying and interpreting immigrant narratives. 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. As a note of caution, it is impossible to say what share of heterogeneity is measurement error and what share reflects real differences in understanding given sentences. While human classifications are imperfect, we find it important to at least explore heterogeneity in how humans interpret narratives that other studies usually ignore.

The second and main purpose is to use the human data as a basis for assessing the performance of our algorithm. Performance consist broadly of two aspects. First, whether a sentence is classified as an immigrant narrative aligned with one of our themes or not. Second, conditional on classification, what sentiment humans assign to the 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.

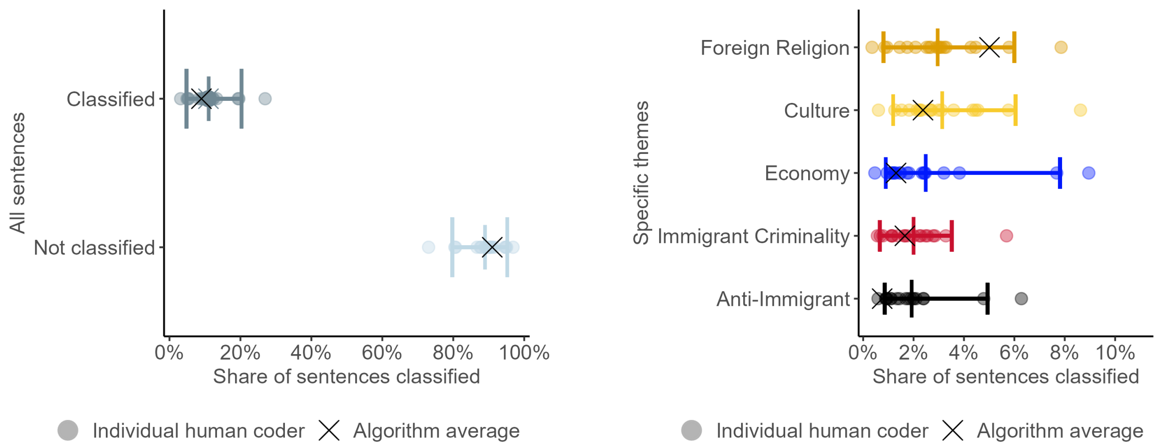
5.2 Classifying immigrant sentences and theme assignment

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 be about immigrants by human coders, the middle vertical line the coder average, and the cross 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 to be about immigrants, compared with 9.0 percent by the algorithm. However, there is considerable variation among coders which we indicate with a 95% 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 tended to be rather conservative in the construction of the dictionaries when it was unclear if a particular term added more true positives or false negatives. 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 *Anti-Immigrant* 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, while *Economy* and *Anti-Immigrant* are very broad themes where our approach can only cover a certain, but sizeable, share.

Figure 9: Human coders and algorithm classification rates



(a) Classification of migrant sentences

(b) Theme assignment

Panel (a): The graph shows the distribution of classification rates of sentences to be about immigrants or not, across human coders and the algorithm. A sentence is classified as a migrant narrative when it is assigned to at least one immigrant narrative theme. The average classification rate by human coders is reported with a constructed 95% confidence interval, representing the 5th and 95th percentile of the coders.
Panel (b): The graph shows the classification rate within 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 a 95% confidence interval. A sentence can be classified into more than one theme.

So how should we interpret the heterogeneity among human coders? It is striking how strong 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. Striking differences and interesting discussions did arise already when creating the dictionaries and algorithm among the authors of this paper and the initial research assistants. The same narrative fragment can be interpreted quite differently depending on what grander narrative scheme and perception of reality a reader has in mind. Nonetheless, it is surprising that such a large 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.

The large heterogeneity among humans in understanding immigrant narratives that we discover suggests that for this application using a transparent, dictionary-based approach like we do has some advantages over a machine-learning approach. While there is heterogeneity in interpretation, even with the intensive training and detailed codebook, 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 an immigrant narratives increases strictly in the share of human coders agreeing on classifying it that way. In the following, we use sentences that no coder classified as an immigrant narratives as a proxy for a false positive, and the ones all four coders classify that way as a proxy for a true positive.

Figure 10b uses the sentences that all humans and our algorithm classify to be about immigrants to assess which features of the immigrant sentence selector and the theme-specific approaches contribute how much to a good classification. We find that about half of the correct classifications are based on general immigrant dictionaries in combination with the NLP tools and our own functions (conditions for dictionary mapping, amplifier, qualifier, negation). 28.9% originate from the simple migrant term dictionary, 13.2% from our nationalities and foreign country dictionary, and 7% based on the foreign names dictionary. Almost half the correctly specified sentences 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.

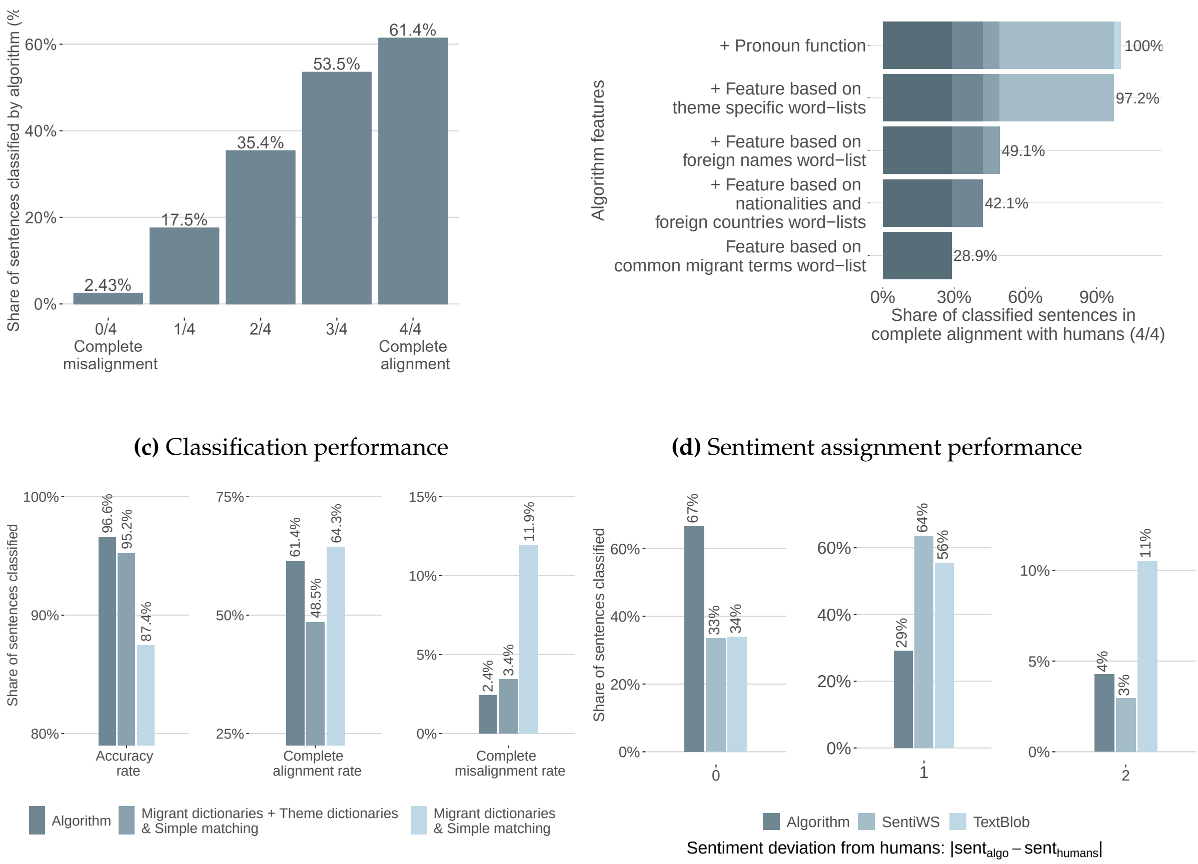
Figure 10: Performance of algorithm

(a) Algorithm performance improves with human alignment

(b) Marginal contribution of each algorithm feature for classification

(c) Classification performance

(d) Sentiment assignment performance



Panel (a): Each bar represents the share of sentences classified by our algorithm to be about immigrants conditional on how many human coders also classified the sentence to be about immigrants. Each sentence is evaluated by four human coders. If there would be perfect agreement among human coders, all sentences would be classified to be about immigrants by either 0 or 4 human coders. 2.43 percent of sentences not classified by any human to be about immigrants get classified by the algorithm to be about immigrants. 61.4 percent of sentences classified by all four human coders to be about immigrants also get classified by our algorithm to be about immigrants.
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 that the sentence is about immigrants. The bottom bar shows the share of sentences in complete alignment that gets classified to be about immigrants when using the most basic feature of the algorithm that relies on a common immigrant terms word-list. The bar on top of it shows the share of additional sentences in complete alignment that gets classified when adding the algorithm's feature based on nationalities and foreign countries word-lists 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. 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 to be about immigrants, within the restricted sample of sentences on which all human coders unanimously agree with each other in terms of classification (60395 sentences). 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) and 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 to be about immigrants 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 61 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.

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 the combining dictionaries with NLP tools and designing our own functions adds much additional value. Even though the resulting data-set 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 migrant dictionaries from the immigrant sentence selector and matching combining the migrant 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. In addition, we analyze the share of sentences that all humans classify as a migrant narrative, as well as the share of those that no human classifies that way. 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 migrant dictionaries. Using all dictionaries, however, under-performs by more than 10% for the sentences that all humans evaluate to be about immigrants 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 to be about immigrants to be about immigrants. Relying on migrant 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 be about immigrants as an immigrant narrative fragment.

We next evaluate how large differences there are in sentiment evaluation when comparing our algorithm to simpler alternatives, conditional on being classified as a migrant narrative sentence.[22] 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, Quasthoff and Heyer, 2010) and Textblob (Loria, 2018). SentiWS lists positive and negative polarity-bearing words weighted within the interval of [-1; 1].[23] 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 –

[22] About 1,300 sentences get classified to be about migration by all four human coders and our algorithm.

[23] 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.

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 performs so much better. First, 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 Event studies

6.1 Creating a balanced sample over the 2013-2019 period

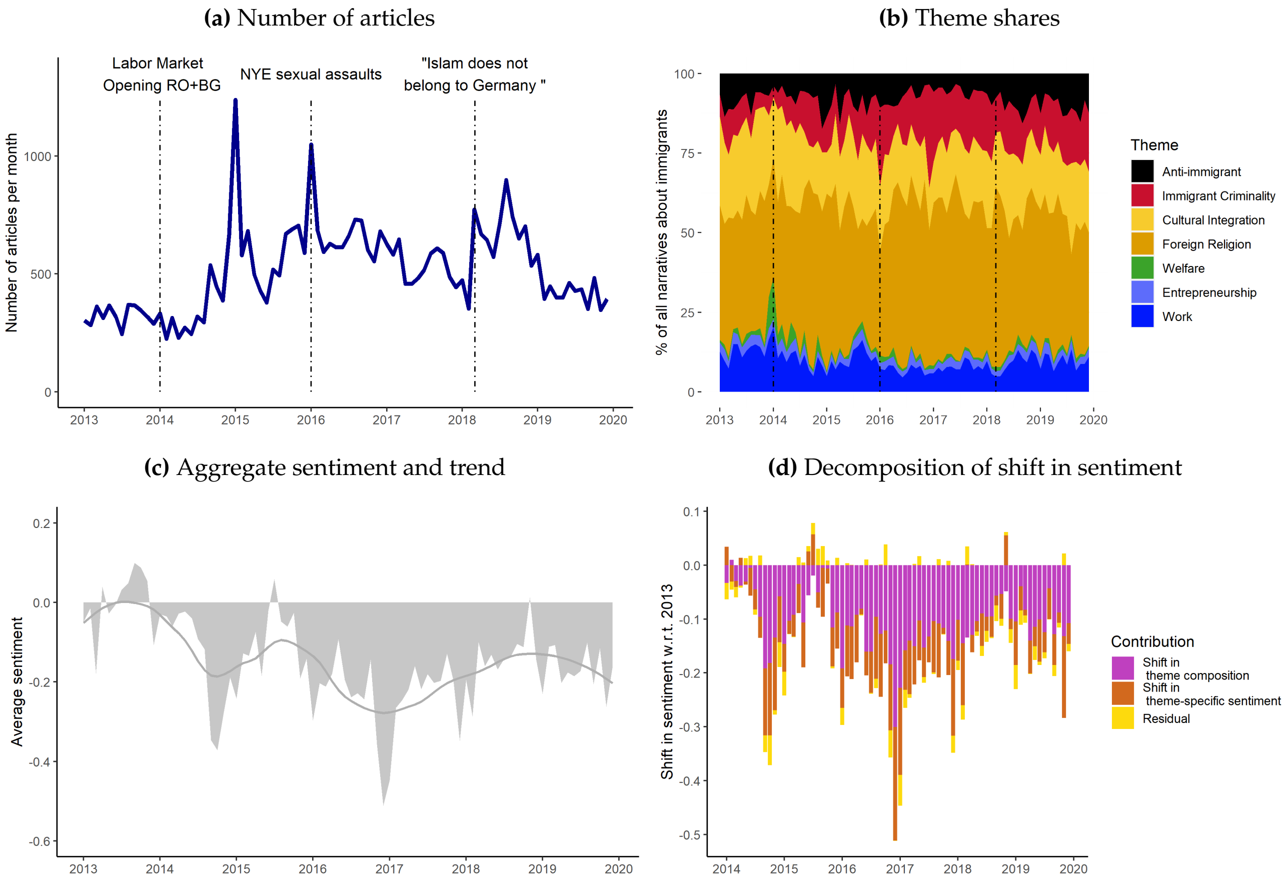
This section aims to provide some examples how our new data-set can be used to better understand the creation, spread and impact of narratives. For that purpose, we focus on three distinctive events in recent German history that are widely believed to have a large influence on the perception of immigrants. To assess the impact of events over time, we create a balanced panel of 41 newspapers (2 national daily, 2 national weekly and 37 regional newspapers). We restrict our sample to the years from 2013 to 2019 for this purpose for two reasons. First, this allows us to have a very broad and representative spatial coverage. Second, those years include many critical and interesting events that provide a good starting point for further analysis.

Figure 11a shows the development of the number of articles about immigrants for the balanced sample, together with three selected events that we want to analyze in more detail in the next subsection. Coverage of immigrants seems to increase on average, with some particular spikes standing out. The vertical lines indicate the three events that relate to "shocks" in different areas ranging from economy over crime to foreign religion.

The three events are the following. 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 Arabic states might plausibly be related to *Immigrant Crime* narratives in particular. The third and final event is an (in-)famous statement by the back-then German minister of the interior that Islam is not part of German culture, which might trigger in particular *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 data set. Figure 11b shows the composition of migrant narratives across themes over time, using a biweekly aggregation of all newspapers in the sample. The overall picture is not drastically different than we considering our maximum sample period, even though we cover a much larger share of all regional newspapers. *Foreign Religion* and *Cultural Integration* dom-

Figure 11: Monthly balanced panel of 41 newspapers between 2013 and 2019



inate quantitatively, followed by the *Economy* themes. Some interesting longer-term developments are visible over time, though. The prevalence of *Immigrant Crime* increases strongly, possibly related to the refugee crisis episode starting in 2015. An interesting observation is that while the refugee crisis ceased to be a major topic after 2017, the *Immigrant Crime* remains persistently higher.[24]

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 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 *Anti-Immigrant* 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

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

that are likely to be related to specific events. Moreover, there is a clear increase in reporting *Anti-Immigrant* 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 data-set 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.[25] Overall, changes in the composition are on average most important, but there also seem to be considerable changes in sentiment within themes that warrant further examination.

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 Years celebrations in the central German city of Cologne, and (iii)the (in-)famous 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 column of Figure **??**, we focus on five types of results. First, to put things into perspective, we report the numbers of immigrant articles published around the event. Second, the composition of immigrant narratives in a stacked chart shows the share of each of them at a particular point in time. Third and fourth, we show the frequency of narrative fragments from two of the themes that we find theoretically most likely to be linked to the specific event. Fifth, we decompose changes in sentiment.

The events differ in their type and nature. The labor market opening is planned event and thus anticipated. Nonetheless, the day plausibly should create attention for the topic, and journalists might update their priors after the law takes place. The incident on New Year's Eve can be considered an unexpected treatment. Similarly, the speech of the back then German minister of the interior was not or only partly anticipated by the German public. 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 **??**. 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. Actual newspapers articles about immigrants, based upon manual reading a significant amount of articles from about 130 per week to more than 150. Figure **??** show 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*.

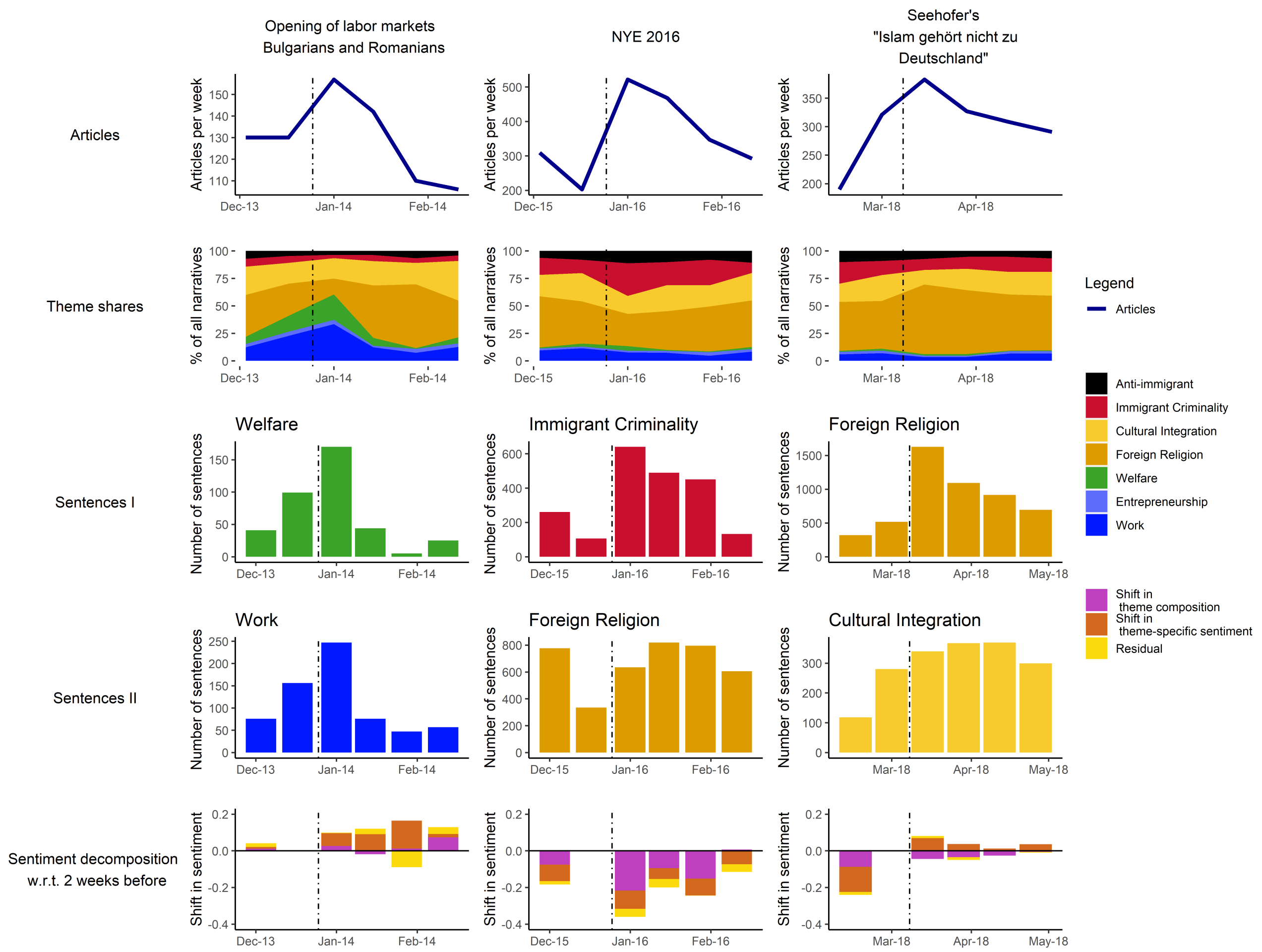
Column 3 and 4 show the absolute changes for those two themes. We find a strong increase in both. What is interesting is that there is a build-up with those narrative fragments already increasing

[25] All changes and the decomposition are relative to a reference period that can be flexibly defined.

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. This graph is normalized to the two weeks 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 with sentiment decomposition for three events



We continue with the 2016 New Year's events in Cologne in the second column of Figure ??. 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 Crime 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 ”the 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 immigrant 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 part of German society more sceptical about immigrant 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 ?? 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 role back a more progressive position on immigration backfired.[26]

Overall, the three events studies reveal the potential of our detailed immigrant narrative data-set and provide some very interesting insights that can form the basis of more in depth analysis. Augmenting the more long-term insights from Michalopoulos and Xue (2021), we find that narratives like folklore are not only 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 way beyond the specific event. An open question is 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 possibilities and importance of measuring sentiment changes and being able to decompose it. 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 outcome 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 report 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 (un)observable characteristics, which affects coverage about immigrants. We divide newspapers in four groups: national newspapers, regional newspapers from Berlin, regional newspapers from former East Germany, and regional newspapers from former West Germany. Figure **??** shows the share of articles that write about immigrants, the average overall sentiment and the polarization in sentiments. In the leftmost panel, we find that Berlin-based regional and national newspapers write considerably more articles about immigrants than the other regional newspapers. Furthermore, we find that for all newspaper groups the number of immigrant articles increases in 2014 and 2015. 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.

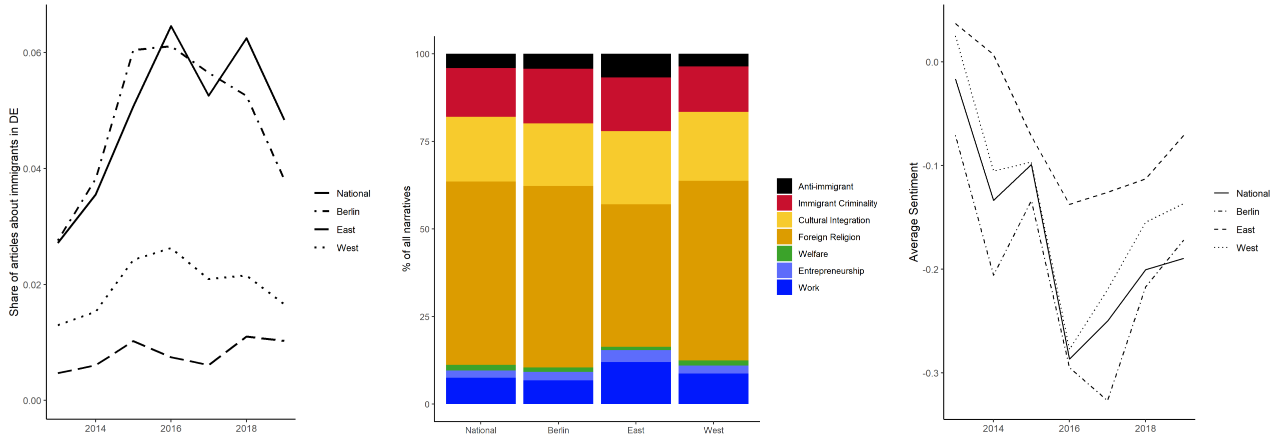
The middle panel shows the relative prevalence of themes across the four newspaper groups between 2013 and 2019. We find few striking differences: (1) newspapers based in Berlin are most likely to report about immigrant crime, (2) regional newspapers from eastern Germany are much less likely to cover foreign religion and (3) much more likely to cover work, and (4) national newspapers

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

and regional newspapers from the west are very similar in terms of theme composition.

The rightmost panel shows the average sentiment for the four groups. Although sentiments in 2013 are comparable between the groups, the pattern in subsequent years is very different. Sentiments worsen by only 0.13 (on a 2 point scale from -1 to 1) between 2013 for newspapers in former Eastern Germany, compared to 0.22 (0.26) [0.28] for Berlin (national) [Eastern regional] newspapers.

Figure 13: Heterogeneity in narratives about immigrants across spatial groups



(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 newspaper groups: two national newspapers, four regional newspapers from Berlin, 17 regional newspapers from Former Western Germany and eight regional newspapers from Former Eastern Germany. **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 within newspaper group. See notes to Table 8 for further explanation.

8 Conclusion

We propose a novel way to study narratives using text-as-data, and apply it to media narratives about immigrants. 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 classify narratives into the following main themes: *Economy* (subdivided in Work, Welfare and Entrepreneurship), *Foreign Religion*, *Cultural Integration*, *Immigrant Criminality* and *Anti-immigrant*. We apply our method to newspaper articles from five national and 65 regional newspapers available in the *Factiva* database. After an initial pre-selection, we downloaded 107,428 articles about immigrants in Germany over the period from 2000 to 2019.

Our aim was to provide a very precise and detailed account of immigrant narratives in Germany for five national and 65 regional newspapers over the 2000 to 2019 period. Following that aim, our initial investment in fine-tuning the theme-specific dictionaries was quite high. These dictionaries can now be reused to study immigrants in German language texts. Similarly, we invested considerable time in developing functions to correct negations and qualifications that are common in newspaper texts. Those function can easily be reused by anyone working on German text.

Methodologically, it depends on the goal of a specific study just how much effort should be invested in each part. Our evaluation using a large set of human coders 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 and our functions turns out to be even larger for sentiment assignment compared to simple general dictionaries. Of course, it is also possible to combine dictionaries, NLP tools and other possibilities offered by machine learning in creative ways.

To provide a comprehensive data-set 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 of five main narrative themes 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 the following main themes: *Economy* (subdivided in Work, Welfare and Entrepreneurship), *Foreign Religion*, *Cultural Integration*, *Immigrant Criminality* and *Anti-immigrant*. We apply our method to newspaper articles from five national and 65 regional newspapers available in the *Factiva* database. After an initial pre-selection, we downloaded 107,428 articles about immigrants in Germany over the period from 2000 to 2019.

Using our new data-set, we made several interesting discoveries. First, we find that descriptively *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 *Entrepreneurship* tends to be a very positive theme. There are, however, also some shifts within themes over time. Our data-set allows us to propose a method to decompose sentiment shifts in compositional effects and within-theme shifts. We then, thirdly, analyze three important immigration-related events in recent German history and their influence on immigrant narratives.

This event analysis provides some interesting insights into how narratives in the media work. Changes in narrative theme composition are substantial. Decomposing sentiment allows us to understand how and why an event influences sentiment towards immigrants. 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. One important insight is thus that media shapes narratives not only by picking what to write about, but also about how to write about it.

FIRST ALTERNATIVE, ENDING WITH RESULTS: We propose a novel way to study narratives using text-as-data, and apply it to media narratives about immigrants. 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 apply our method to newspaper articles from five national and 65 regional newspapers available in the *Factiva* database. After an initial pre-selection, we downloaded 107,428 articles about immigrants in Germany over the period from 2000 to 2019.

To provide a comprehensive data-set 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. Instead of following an unstructured topic-modelling approach, we classify narratives into the following main themes: *Economy* (subdivided in Work, Welfare and Entrepreneurship), *Foreign Religion*, *Cultural Integration*, *Immigrant Criminality* and *Anti-immigrant*.

Our aim was to provide a very precise and detailed account of immigrant narratives in Germany. This required a high initial investment in fine-tuning the theme-specific dictionaries. These dictionaries can now be reused to study immigrants in any texts in German. Similarly, we developed functions to correct negations and qualifications that are common in newspaper texts. Those functions can easily be reused by anyone working on German text.

Methodologically, it depends on the goal of a specific study just how much effort should be invested in each part. Our evaluation using a large set of human coders 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 and our functions turns out to be even larger for sentiment assignment compared to simple general dictionaries. Of course, it is also possible to combine dictionaries, NLP tools and other possibilities offered by machine learning in creative ways.

Using our new data-set, we made several interesting discoveries. First, we find that descriptively *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 *Entrepreneurship* tends to be a very positive theme. There are, however, also some shifts within themes over time. Our data-set allows us to propose a method to decompose sentiment shifts in compositional effects and within-theme shifts. We then, thirdly, analyze three important immigration-related events in recent German history and their influence on immigrant narratives.

This event analysis provides some interesting insights into how narratives in the media work. Changes in narrative theme composition are substantial. Decomposing sentiment allows us to understand how and why an event influences sentiment towards immigrants. 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. One important insight is thus that media shapes narratives not only by picking what to write about, but also about how to write about it.

SECOND (MY PREFERRED) ALTERNATIVE, WITH RESULTS EARLIER:

We propose a novel way to study narratives using text-as-data, and apply it to media narratives about immigrants. 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 apply our method to newspaper articles from five national and 65 regional newspapers available in the *Factiva* database. After an initial pre-selection, we downloaded 107,428 articles about immigrants in Germany over the period from 2000 to 2019.

To provide a comprehensive data-set 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. Instead of following an unstructured topic-modelling approach, we classify narratives into the following main themes: *Economy* (subdivided in Work, Welfare and Entrepreneurship), *Foreign Religion*, *Cultural Integration*, *Immigrant Criminality* and *Anti-immigrant*.

Our data-set provides several interesting insights. First, we find that descriptively *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. Our data-set also allows us to propose a method to decompose sentiment shifts in compositional effects and within-theme shifts. We then, thirdly, 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. 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 easily be reused by anyone working on German text. Our aim to provide a very precise and detailed account of immigrant narratives in Germany also required a high initial investment in 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 perceive same sentences. This raises a concern about machine-learning tools that use human-coded text to guide the algorithm: if different human coders perceive the same text differently, the outcome arising from machine-learning may depend more on coders used than previously realized. It is also unclear what training set should be used if humans perceive the same text differently. Nonetheless, focusing on those sentences where human coders agree, our algorithm performs much better in capturing human perceptions than previous dictionary-based approaches. 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 and our functions turns out to be even larger for sentiment assignment compared to simple general dictionaries. We invite other researchers to use and further develop the tools we provide.

.

References

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of*

Economics, 115(3): 715–753. 4

Alesina, Alberto, and Marco Tabellini. 2022. "The Political Effects of Immigration: Culture or Economics?" 5

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. Forthcoming. "Immigration and redistribution." *Review of Economic Studies*. 5

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart. 2021. "Narratives about the Macroeconomy." 1

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2021. "Text Semantics Capture Political and Economic Narratives." *arXiv preprint arXiv:2108.01720*. 4, 5

Ash, Elliott, Ruben Durante, Maria Grebenshchikova, and Carlo Schwarz. 2022. "Visual representation and stereotypes in news media." 1

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*. 4, 6

Barone, Guglielmo, Alessio D'Ignazio, Guido De Blasio, and Paolo Naticchioni. 2016. "Mr. Rossi, Mr. Hu and Politics. The role of immigration in shaping natives' voting behavior." *Journal of Public Economics*, 136: 1–13. 5

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. 4

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. "Narratives, imperatives, and moral reasoning." 3, 5

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. 4

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. 4

Bursztyn, Leonardo, Aakaash Rao, Christopher P Roth, and David H Yanagizawa-Drott. 2020. "Misinformation during a pandemic." National Bureau of Economic Research. 1, 3, 4, 5

Cagé, Julia, Nicolas Hervé, and Béatrice Mazoyer. 2020. "Social Media and Newsroom Production Decisions." *Available at SSRN 3663899*. 4

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. 4

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. 4

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. 4

Card, David, Christian Dustmann, and Ian Preston. 2012. "Immigration, wages, and compositional amenities." *Journal of the European Economic Association*, 10(1): 78–119. 4

Couttenier, Mathieu, Sophie Hatte, Mathias Thoenig, and Stephanos Vlachos. 2022. "The Logic of Fear - Populism and Media Coverage of Immigrant Crimes." *Review of Economics and Statistics, forthcoming*. 4

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist. 2012. "Ethnic diversity and preferences for redistribution." *Journal of Political Economy*, 120(1): 41–76. 5

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. 4

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122(3): 1187–1234. 1

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. 4

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. 1

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. "The political legacy of entertainment TV." *American Economic Review*, 109(7): 2497–2530. 1

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm. 2019. "Refugee migration and electoral outcomes." *The Review of Economic Studies*, 86(5): 2035–2091. 5

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. 2, 5

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–85. 4

Enke, Benjamin. 2020. "Moral values and voting." *Journal of Political Economy*, 128(10): 3679–3729. 4

Esposito, Elena, Tiziano Rotesi, Alessandro Saia, and Mathias Thoenig. 2021. "Reconciliation Narratives: The Birth of a Nation after the US Civil War." 1, 3, 5

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. 4

Factiva Database. 2020. 12

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. 4

Galletta, Sergio, and Elliott Ash. 2019. "How Cable News Reshaped Local Government." *Available at SSRN 3370908*. 4

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. 4

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. 4

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature*, 57(3): 535–74. 6

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. 1, 4

Gerber, Alan S, Dean Karlan, and Daniel Bergan. 2009. "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions." *American Economic Journal: Applied Economics*, 1(2): 35–52. 1

German Audit Bureau of Circulation (IVW). 2020. "IVW VA2020." 13

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. 2

Giuliano, Paola, and Nathan Nunn. 2021. "Understanding Cultural Persistence and Change." *The Review of Economic Studies*, 88(4): 1541–1581. 4

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-term Persistence." *Journal of the European Economic Association*, 14(6): 1401–1436. 4

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. 2, 5

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. 2, 5

Harari, Yuval Noah. 2016. *Sapiens.* Bazarforlag AS. 5

Hatton, Timothy J. 2017. "Refugees and asylum seekers, the crisis in Europe and the future of policy." *Economic Policy*, 32(91): 447–496. 4

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. "spaCy: Industrial-strength natural language processing in python." 6, 7

Hudson, Richard Paul. 2022. "Coreferee: Coreference resolution for multiple languages." 7

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. 4

Loria, Steven. 2018. “textblob Documentation.” *Release 0.15*, 2. 23

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. “Folklore.” *The Quarterly Journal of Economics*, 136(4): 1993–2046. 3, 4, 29

Norris, Pippa, and Ronald Inglehart. 2019. *Cultural backlash: Trump, Brexit, and authoritarian populism.* Cambridge University Press. 2

Osnabrügge, Moritz, Elliott Ash, and Massimo Morelli. 2021. “Cross-Domain Topic Classification for Political Texts.” *Political Analysis*, 1–22. 6

Ottaviano, Gianmarco IP, and Giovanni Peri. 2012. “Rethinking the effect of immigration on wages.” *Journal of the European Economic Association*, 10(1): 152–197. 4

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. 2, 5

Remus, Robert, Uwe Quasthoff, and Gerhard Heyer. 2010. “SentiWS - A Publicly Available German-language Resource for Sentiment Analysis.” Valletta, Malta:European Language Resources Association (ELRA). 23

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. 4

Shiller, Robert J. 2017. “Narrative economics.” *American Economic Review*, 107(4): 967–1004. 1, 3, 4, 5

Snyder, James M., and David Strömberg. 2010. “Press coverage and political accountability.” *Journal of Political Economy*, 118(2): 355–408. 4

Steinmayr, Andreas. 2021. “Contact versus exposure: Refugee presence and voting for the far right.” *Review of Economics and Statistics*, 103(2): 310–327. 5

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. “Political Effects of the Internet and Social Media.” *Annual Review of Economics*, 12: 415–438. 1, 4

Appendix A Methodology

Dictionaries

Figure A.1: All dictionaries used in the article selection, sentence selection, theme assignment and sentiment assignment

- **A1 - Immigrant list** terms related to migration and migrants. Examples are ”Asylunterkünft” (”Asylum shelter”) and ”EU-Angehörige” (”EU Citizen”) N = 64.
- **A2 - Foreign and dual nationalities list** terms related to nationalities and dual nationalities, both adjectives and nouns. Examples are ”Ungarinnen” (”Hungarians” - female) and ”deutsch-Afghanistan*” (”German-Afghani*”). N = 5467.
- **A3 - Foreign names list** terms related to migration and migrants. Examples are ”Mohammed” and ”EU-Angehörige”.
- **A4 - Countries list**. N = 203.
- **A5 - Theme specific lists** terms related to predefined themes and sub-themes. Each sub-theme has its own respective dictionary: Economy - Work, Economy - Entrepreneurship, Foreign religion, Cultural integration, Immigrant criminality Anti-immigrant. Total = 1014.
  - **A5.1 - Foreign Religion list** terms related to migration and migrants manually compiled. Examples are ”Asylunterkünft” (”Asylum shelter”) and ”EU-Angehörige” (”EU Citizen”) N = 129. item **A5.2 - Cultural Integration list** terms related to cultural integration, including education. Examples are ”Überfremdung” (”infiltration”) and ”Toleranz” (”tolerance”) N = 185.
  - **A5.3 - Economy list** terms related to the economy, manually compiled. Subdivided in work, entrepreneurship and welfare state. Examples are ”Unternehmer” (”entrepreneur”) and ”arbeitsunwillig” (”unwilling to work”) N = 280.
  - **A5.4 - Crime list** terms related to crimes and attitudes towards immigrants, manually compiled. Examples are ”Tötung” (”Murder”) and ”Dealer” (”dealer”) N = 230.
  - **A5.5 - Anti-immigrant list** terms related to crimes and attitudes towards immigrants, manually compiled. Examples are ”Dönermorde” (”Döner murders”) and ”fremdenfeindlich” (”foreigner unfriendly-”) N = 144.
  - **A5.6 - Theme-specific n-gram list** 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 list** terms used for German and foreign locations.

Factiva filter

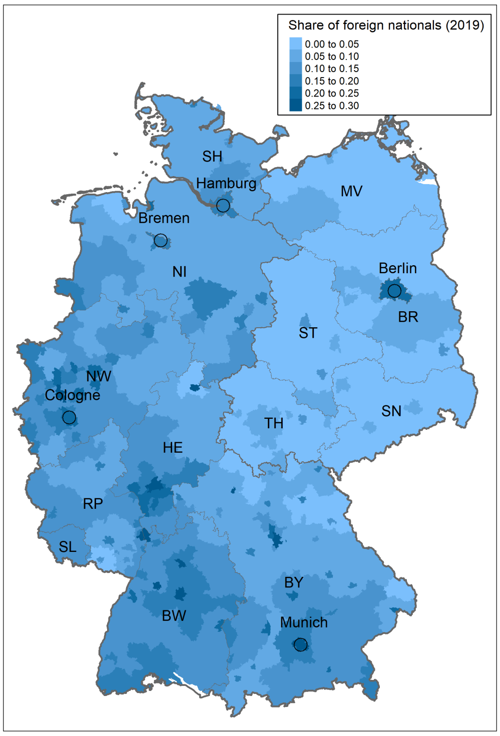
We queried relevant newspaper articles from Factiva by newspaper and time period using the following logical filter:

(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")*

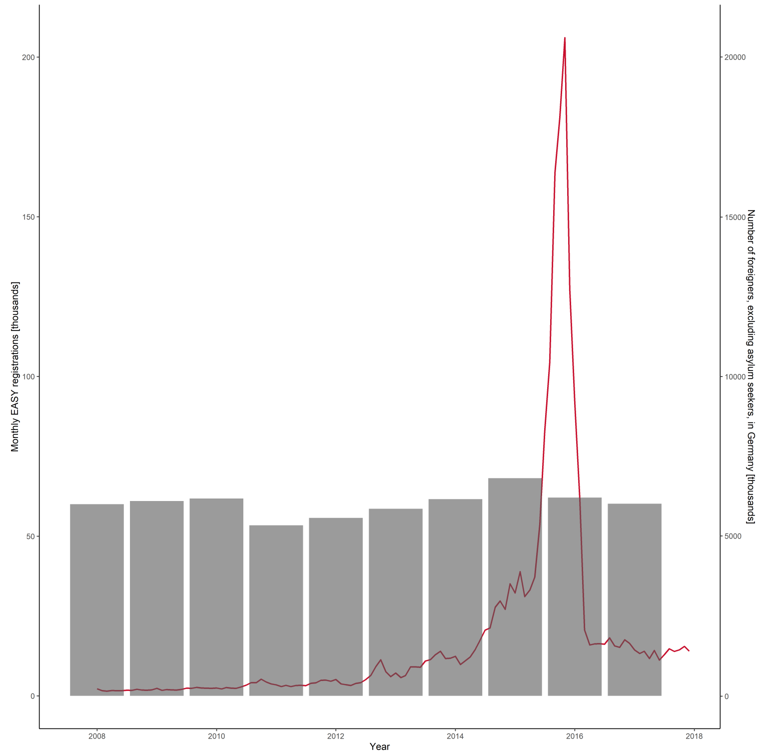
Appendix B Descriptive statistics

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 = Sachsen-Anhalt, SN = Sachsen, TH = Thüringen, BY = Bayern, BW = Baden-Wurttemberg, SL = Saarland, RP = Theinland-Pfalz, HE = Hessen, NW = Nordrhein-Westfalen, NI = Niedersachsen.

Figure B.2: Asylum seeker registrations and immigrants in Germany



??

Figure B.3: Number of Newspapers and Articles from 2000 to 2019

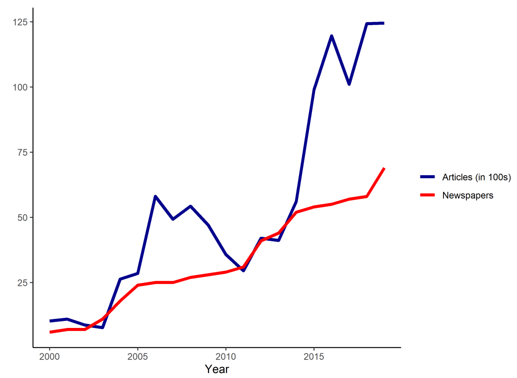


Figure B.4: Newspaper Coverage from 2000 to 2019

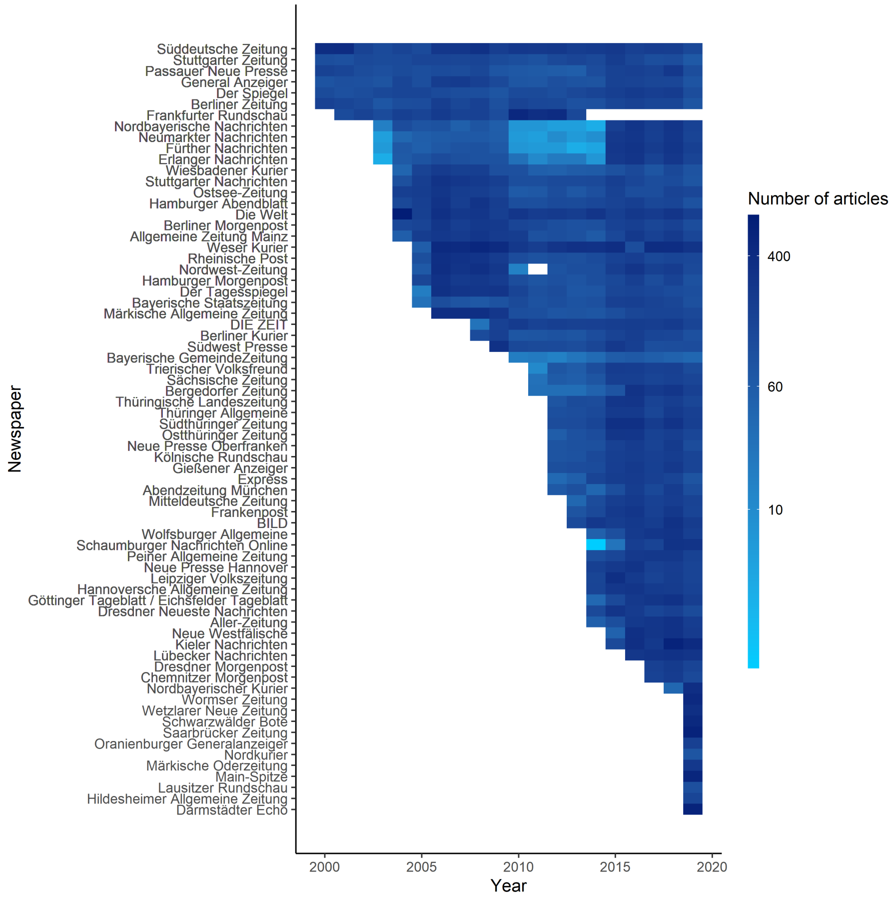
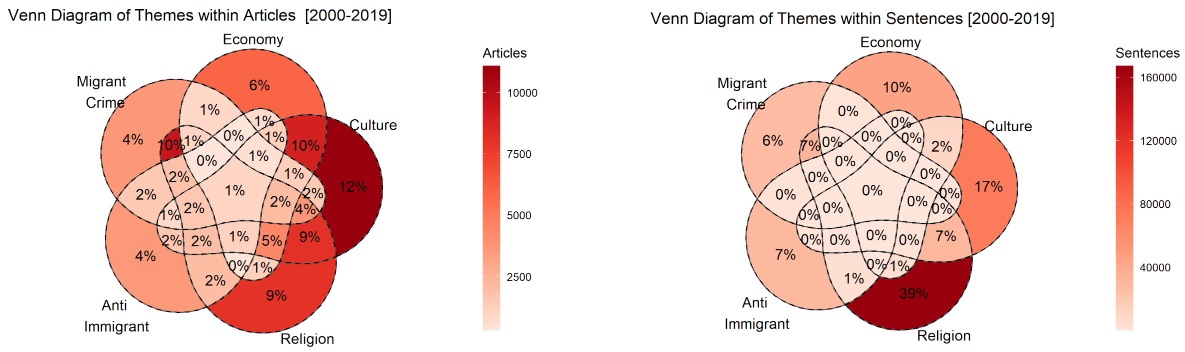


Table B.1: Summary statistics of national and regional newspapers

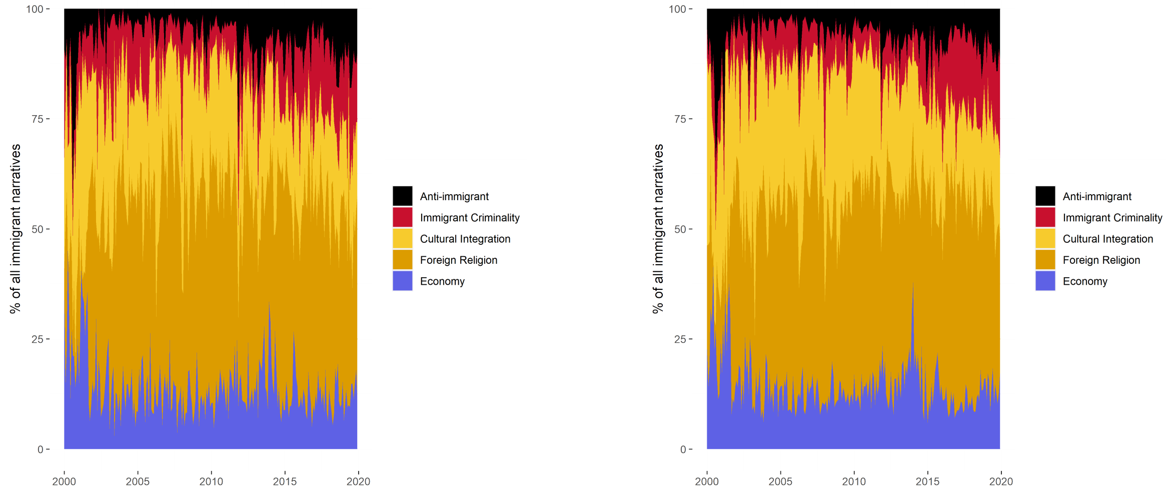
	Newspaper	Articles	About immigrants	From agency	Aggregated sentiment
1	Weser Kurier	4703	0.80	0.02	-0.05
2	Süddeutsche Zeitung	4480	0.87	0.04	-0.06
3	Die Welt	4242	0.83	0.09	-0.15
4	Frankfurter Rundschau	2741	0.86	0.11	0.00
5	Der Spiegel	2705	0.83	0.01	-0.18
6	Allgemeine Zeitung Mainz	2672	0.85	0.19	-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

Figure C.1: Venn diagram of 5 main themes, on the article and sentence level



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

Figure C.2: Shares of the Main Themes, for National and Regional Newspapers



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

Appendix C Additional results

C.1 Main themes for various subsamples

Figure C.3: Shares of the Main Themes, weighted by sales

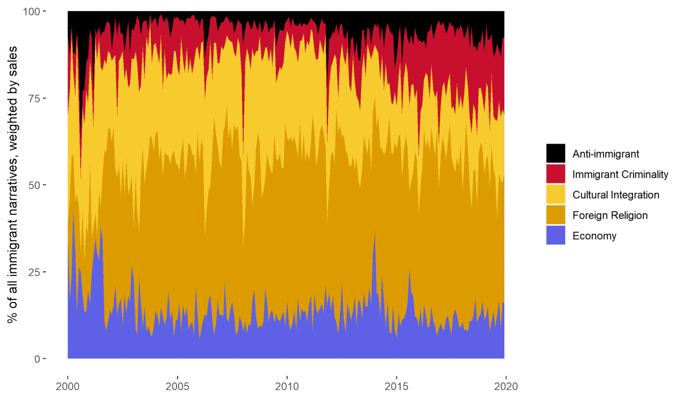
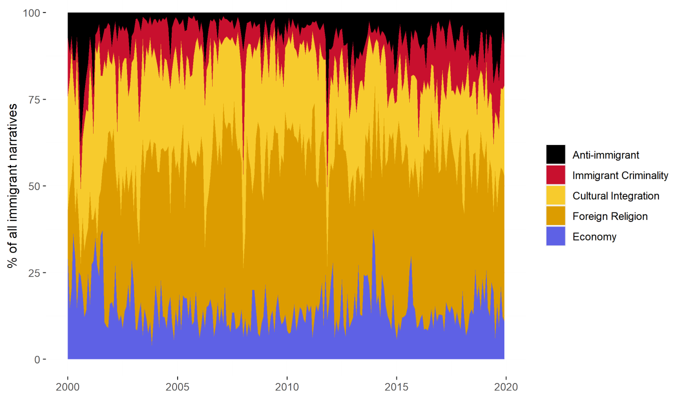


Figure C.4: Shares of the Main Themes, balanced



C.2 Sub-themes

Figure C.5: Shares of Economy Sub-themes from 2000 to 2019

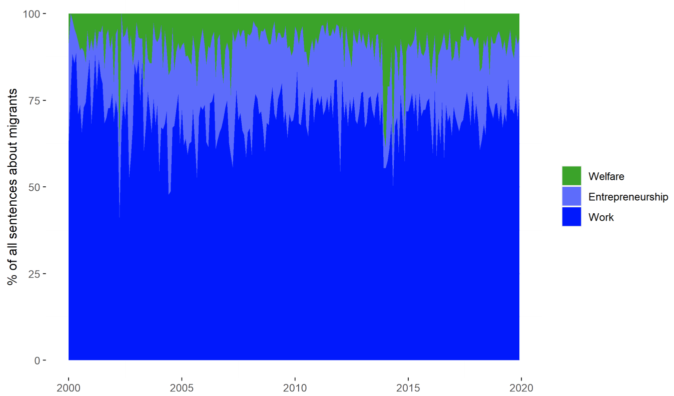


Figure C.6: Most salient theme per municipality - Regional and national newspapers - 2019

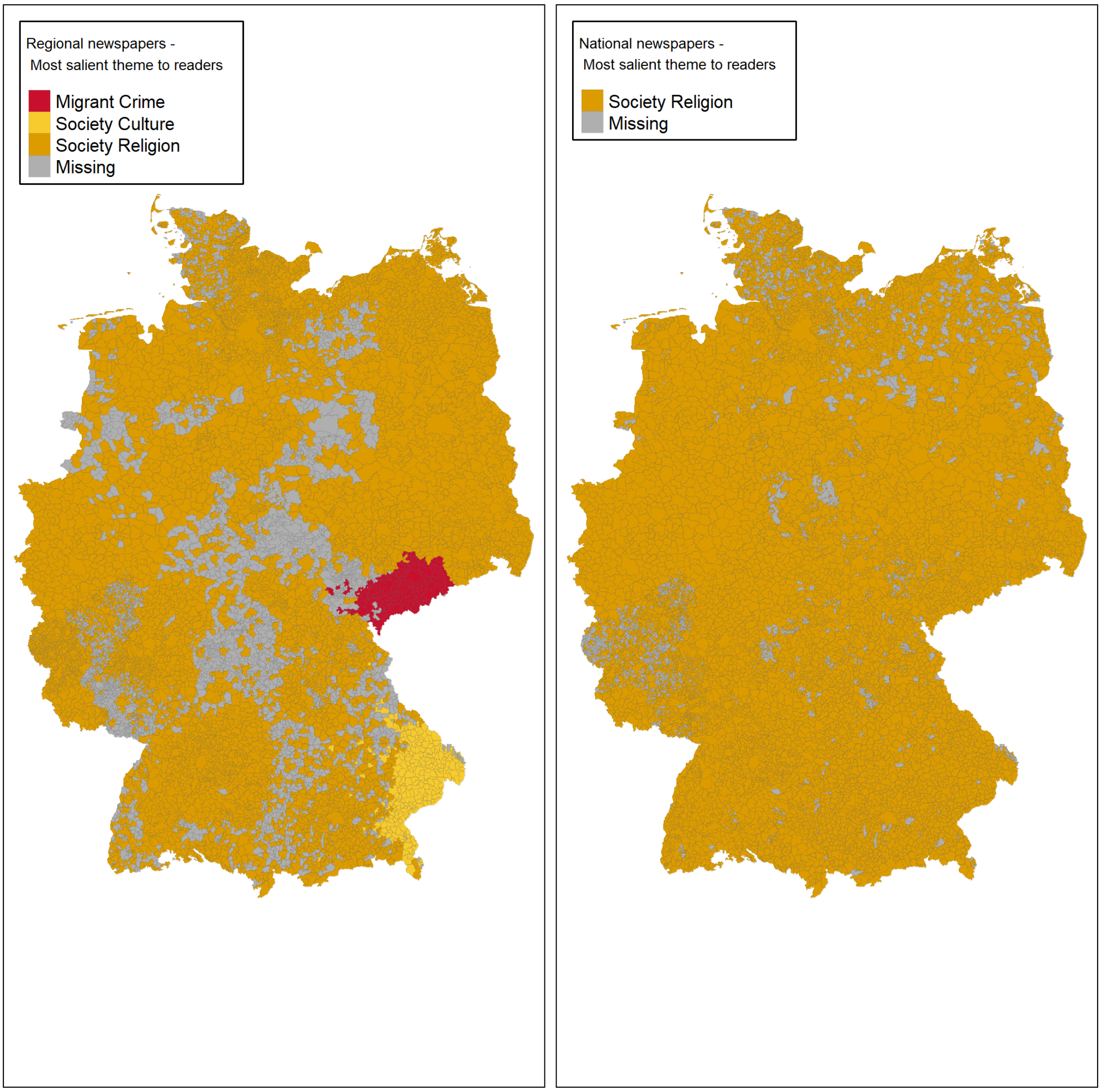


Figure C.7: Average sentiment of sentences about immigrants between 2013 and 2019 (monthly)

Appendix D Decomposition of sentiment

In the following, we decompose the change in sentiment shown in Panel d) of Figure **??** into three distinct contributions: a change in sentiment due to shift in salience between themes (which have distinct sentiments), a change in sentiment of each separate theme, 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 isolate the contribution of changes in theme shares (f^{th}):

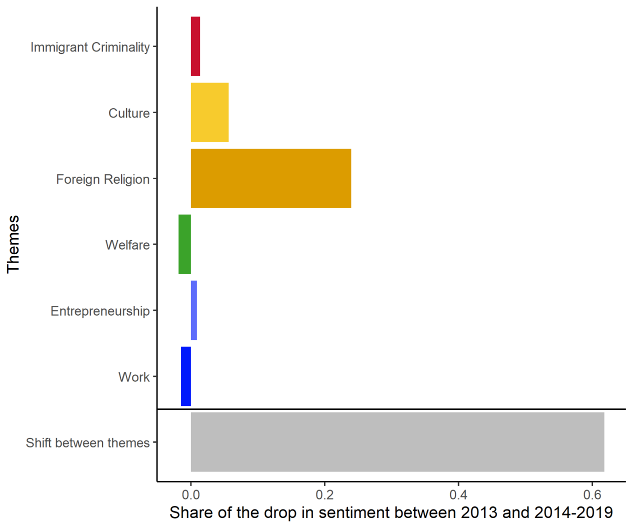
$$\Delta S^{between}_t = \sum_{th} S^{th}_{2013} \times \left(f^{th}_t - f^{th}_{2013} \right) \tag{2}$$

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

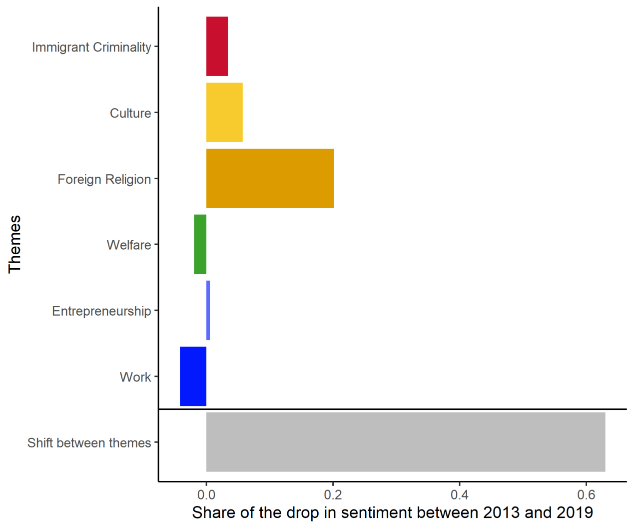
$$\Delta S^{within}_t = \sum_{th} f^{th}_{2013} \times \left(S^{th}_t - S^{th}_{2013} \right) \tag{3}$$

Following, $S_t - S^{between}_t - S^{within}_t$ is the residual component of the sentiment shift, which is driven by correlations between shifts between and within themes. However, in most time periods this contribution is small, as shown in Panel d) of Figure **??**. 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 whole 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 immigrants 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

Appendix E Determinants of human coding

In this section, we explore the determinants of human coding. First of all, we are interested in effects shared among human coders and secondly in effects based on background variables of the students. Note that the number of students is relatively low: we nevertheless cluster standard errors on the student level.

Table E.1 reports the coefficients of a linear probability model regression estimated by OLS of a binary variable whether a sentence is classified or not on a set of coding task-specific, but also coder-level covariates. We find that the order of the sentence is negatively related with classification, suggesting that due to fatigue, students pick up less sentences as they go deeper in the file. The average of this effect is relatively modest: as the average file has 5000 sentences, the last sentences in a file have a 1.5 pp lower classification rate than the first sentences. As students received 4 sub-batches of around 5000 sentences, we are also interested in learning effects over the sub-batches. We find that the second and third sub-batches have lower classification rates than the first, whereas the fourth and last is comparable to the first.

Moving to the third column, we see that the female coders classify considerably more sentences than the male coders. We also find that there is a strong positive association between age (or being a Masters students, we can't disentangle these). Having a migration background seems to be negatively related to the classification rate, although this effect is not robust to inclusion of additional coder-level covariates in the fourth column. We furthermore find that studying political science (rather than economics) is associated with higher classification rates, just as being a master student (rather than a bachelor student).

Table E.2 provides a similar table, but then for sentiment, where the outcome is the assigned

Table E.1: Determinants of human classification of sentences (classified=1, not classified=0)

	Base	Sentence FE	Basic person vars	Full person vars	Student FE
Order of sentence in article	-0.0001***				
	(0.0000)				
Order of sentence in file (in 1000s)	-0.0033	-0.0033***	-0.0021**	-0.0011	-0.0038**
	(0.0021)	(0.0008)	(0.0009)	(0.0013)	(0.0015)
batch_order::2	-0.0070**	-0.0070***	-0.0025	-0.0107	-0.0066
	(0.0028)	(0.0011)	(0.0080)	(0.0075)	(0.0063)
batch_order::3	-0.0123***	-0.0123***	-0.0086	-0.0127**	-0.0118**
	(0.0046)	(0.0014)	(0.0055)	(0.0056)	(0.0055)
batch_order::4	0.0031	0.0031**	0.0041	-0.0056	0.0032
	(0.0040)	(0.0013)	(0.0092)	(0.0084)	(0.0066)
Gender::Female			0.0611***	0.0611***	
			(0.0194)	(0.0137)	
Age			0.0093***	-0.0007	
			(0.0025)	(0.0039)	
migration_bg::.			0.0205		
			(0.0380)		
migration_bg::Yes			-0.0414**	-0.0187	
			(0.0161)	(0.0203)	
parents_uni::Yes				-0.0396	
				(0.0267)	
study_broad::Pol. Sci.				0.1218***	
				(0.0381)	
study_level::Master				0.0873**	
				(0.0327)	
N	374833	374833	299072	262291	374833
R^2	0.17	0.60	0.65	0.69	0.62
Article FE	X				
Sentence FE		X	X	X	X
Student FE					X

sentiment (-1,0,1). Note that this is *conditional* on being selected: e.g. a higher average per student can be caused by (1) finding more positive sentences or (2) being more positive given the sentence. In the second column we introduce sentence FEs, which partials out (1). We find no systematic relationship between coding task-specific variables and sentiment. We do find that Political Science and Master students report more negative sentiments.

Table E.2: Determinants of sentiment, conditional on selection

	Base	Sentence FE	Basic person vars	Full person vars	Student FE
Order of sentence in article	0.0006**				
	(0.0003)				
Order of sentence in file (in 1000s)	0.0015	-0.0021	0.0028	-0.0070	-0.0021
	(0.0103)	(0.0061)	(0.0055)	(0.0110)	(0.0048)
batch_order::2	-0.0187	-0.0088	-0.0211	-0.0141	-0.0079
	(0.0129)	(0.0069)	(0.0139)	(0.0146)	(0.0095)
batch_order::3	-0.0013	-0.0011	-0.0262	-0.0216	-0.0097
	(0.0161)	(0.0083)	(0.0197)	(0.0207)	(0.0157)
batch_order::4	0.0071	0.0051	-0.0212	-0.0168	0.0051
	(0.0134)	(0.0075)	(0.0165)	(0.0161)	(0.0150)
Gender::Female			-0.0576	-0.0075	
			(0.0547)	(0.0516)	
Age			-0.0083	0.0302	
			(0.0064)	(0.0227)	
migration_bg::.			0.1027		
			(0.0624)		
migration_bg::Yes			0.0341	-0.0612	
			(0.0336)	(0.0619)	
parents_uni::Yes				0.1729	
				(0.0997)	
study_broad::Pol. Sci.				-0.3907**	
				(0.1745)	
study_level::Master				-0.3100**	
				(0.1350)	
N	55427	55427	45356	38797	55427
R^2	0.37	0.84	0.86	0.89	0.85
Article FE	X				
Sentence FE		X	X	X	X
Student FE					X

Appendix F 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 jonathan.oeztunc@googlemail.com and charlotte.robert@awi.uni-heidelberg.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**.