ALMA MATER STUDIORUM-UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO DI SCIENZE POLITICHE E SOCIALI

CORSO DI LAUREA MAGISTRALE IN

Politica, Amministrazione, Organizzazione - Politics and Social Policy

Tesi di Laurea Magistrale in

Social Stratification and Economic Inequality

Trajectories of local development and public opinion

Environmental policy in Umbria

Candidato: Lorenzo Mattioli

Relatore: Prof. Marco Albertini

Co-relatrice: Prof. Arianna Tassinari

ANNO ACCADEMICO 2024/2025

Table of contents

CHAPTER 1

Introduction

The last decades have witnessed the rise of an interrelation in the public debate between industrial and environmental policies. In layman terms, the discussion has often revolved around two opposite arguments: a supposed trade-off between environmental policy and economic development, thus including job security (Vona 2019). This may seem unreasonable to an academic audience for two reasons. First and foremost, there are fairly strong theoretical grounds to the claim that an energy transition would lead to an increase in labour demand at the aggregate level (Smith and Carbone 2007). This fact alone would make framing the problem as a zero-sum game irrational in the eyes of an economist. Second, to the average natural scientist, balancing concerns on occupation with the perspective of mass extinction might seem silly at best. To the workers employed in those industries which face the threat of extinction, however, the choice between job security and mitigating climate change is all but obvious. The key to explaining support for environmental policy could be found in the fear of experiencing a feeling of *loss* and noxious deindustrialisation (Feltrin, Mah, and Brown 2022; Jackson and Grusky 2018). This project therefore insists on the importance of analysing labour markets to explain

public opinion formation around the issue of climate and environmental policy. In particular, the project aims to answer the following research question:

> Is public discourse on environmental policy correlated to a local economy's *exposure* to the energy transition?

In order to start explaining this phenomenon, however, one first has to *establish* it. The present study will hence delineate the theoretical premise such question is based on, but refrain from advancing any causal claims. Using the Umbria region (central Italy) as a case study, an empirical exploration of the phenomenon will be conducted. Such endeavour will be carried out through computational methods, by analysing a sample of online press releases and videos. The content of the subsequent chapters is summarised below:

In the second chapter, the importance of adopting a localised perspective will be stated and discussed. The concept of *loss* will be defined and adapted to the specific case of environmental policy. The theoretical lens through which the present analysis of public discourse is conducted will be spelled out, before briefly discussing the arguments behind case selection.

In the third chapter, the two selected cases will be described and compared from a historical point of view, in order to provide the necessary context for the subsequent analysis.

In the fourth chapter, the sample selection strategy will be presented. The two resulting corpora will be then described in detail before defining the analytical strategy adopted to produce the results enumerated in chapter five.

The results will be finally discussed in the sixth chapter, in which the limitations of

the present study will be also assessed, and the key findings summarised in a brief concluding section.

CHAPTER 2

Environmental policy and public discourse

2.1 The importance of a local perspective

Climate change is a complex, multifaceted phenomenon. Its consequences manifest in different ways across territories and regions, and are borne unequally by different groups (Calvin et al. 2023). In the most pragmatic sense, some areas may be prone to flooding, while others are subject to desertification. Through a breakdown of the impacts of climate change at regional level, Rodríguez-Pose and Bartalucci (2023) found that European cities are less vulnerable than rural areas. There is thus a consolidated consensus both in the natural sciences and in literature on the energy transition that climate change should be understood as a *localised* phenomenon, where each political community faces different problems and tailors its solutions to pre-existent socio-economic and environmental conditions (Pörtner et al. 2023; Smith and Carbone 2007). The latter should be considered path-dependent, as in informed and constructed by the history of a certain region. It follows that the social scientist's approach on this topic should not fundamentally differ from that of a natural scientist. Just like the biologist takes the ecological context into account when analysing the

health of a natural habitat, a sociologist or political scientist cannot forget to consider local idiosyncrasies when analysing the socio-economic conditions of an artificial one (Dunlap and Brulle 2015). This becomes especially apparent in areas whose economic activity is more clearly linked to its climate: Hamilton and colleagues, for instance, found that areas which used to thrive off of skiing suffer more acutely from the adverse economic effects of the climate crisis as compared to other, less vulnerable areas (Hamilton et al. 2003). A similar argument can be made for socio-economic vulnerability to environmental policy. A regional economic system which is based on inherently polluting activities will have to undergo more fundamental change than one whose economy is centred around *green-er* ones. In a context where goals and strategies are set at a higher-than-regional level, environmental policy might find some regions more *ready* than others. To differentiate it from *vulnerability* to the natural consequences of climate change, this concept will be hereinafter referred to as a local economy's *exposure* to environmental policy.

The political outlook on climate change also appears to be geographically informed. In the most down-to-earth fashion, the occurrence of climate anomalies and rising temperature trends in a particular area have been proven to be reliable predictors of individuals' outlook on climate change as an artificial phenomenon (Hamilton and Keim 2009; Hamilton and Stampone 2013). This makes the case for adopting a local perspective even more compelling in a study concerned with public discourse.

2.2 De-industrialisation and patterns of *loss*

As discussed above, the overwhelming presence of job-loss arguments in the public discussion on environmental policy is hardly explainable with the traditional tools of economic *rationality*, since its effects should benefit the whole population indiscriminately both in terms of occupation and health. In *exposed* regional economies, however, workers of *brown*, polluting jobs might have a harder time dismissing the notion of environmental policy as a *zero-sum* game, where an improvement in public health is balanced out by a decline in employment and job security.

In a post-industrial context, workers in exposed economies often experience *noxious* de-industrialisation. *Noxiousness* is a translation from the Italian *Nocività*, a word used by the Italian labour movement in the sixties to identify *production-induced damage against both human and non-human life* (Feltrin and Sacchetto 2021). By noxious de-industrialisation, Feltrin, Mah, and Brown (2022) identify a scenario in which job loss and toxic pollution coexist, where the old factories are still operative (and polluting), but employ a fraction of the original workforce. To the inhabitants of an area which used to thrive off of industrial activity, the post-industrial experience is one of loss and decline. As Jackson and Grusky (2018) poignantly point out, *concentrated* loss, as in loss attributable to a single group, can lead to its *politicisation*. What this means is that those active or employed in *exposed* economies might interpret environmental policy as a covertly distributional zero-sum game, where blue collar industrial workers lose out to urban, service-class elites, active in *green*, future-proof sectors of the economy. The truthfulness of the (re-) distributional claim is of little interest in this instance, although a fear of loss in exposed economies does not appear to be groundless: in Europe, the intensity of toxic pollution is still positively

associated with employment and wages at the industrial level (and specifically in the energy sector), while a reduction of pollutants at facility level is associated to lower employment and wages (Bez and Virgillito 2022). Following Jackson and Grusky (2018) and their *Post-liberal theory of social stratification*, resentment, resistance, and retrenchment against environmental policy should be then expected from those who feel that their jobs are endangered by it.

2.3 Theoretical framework and research questions

As discussed above, in exposed economies, we expect the dominant narrative around environmental policy to be one of a zero-sum game (Jackson and Grusky 2018). The aim of this study, however, is to *establish* rather than explain the phenomenon at hand. The approach adopted is therefore clearly a grounded, descriptive one.

The central research question is thus the following:

> *Can the characteristics of public discourse around environmental policy be correlated to a regional economy's* exposure *to environmental policy?*

2.3.1 Public discourse analysis

In order to detect shifts in public discourse, we turn to media. This is a somewhat unorthodox approach: in stark contrast with the more consolidated practice of considering media to have an active role in shaping narratives and political concepts, they will be here thought of as responsive to public opinion *on the ground*. This is a rather common concept in social movements research, where issue salience and awareness

are thought to be influenced by mobilisation (see, for instance: Carey, Branton, and Martinez-Ebers 2014). Classic concepts in agenda-setting theory, however, still represent this study's theoretical core. Environmental policy will be here analysed in terms of its *salience* in local media (McCombs 1976). Heavy media emphasis on environmental policy (salience) will be treated as a sign of it ranking higher in personal agendas in a certain area. Its *attributes*, as in its sub-issues and relationship with other policy areas, will be here aptly qualified.

2.3.2 Economic *exposure*

The concept of economic *exposure* (as in vulnerability to the energy transition) is here operationalised through the identification of two localised labour market conditions:

- Having previously experienced industrial decline and suffered the consequences of noxious deindustrialisation;
- Presenting a prevalence of brown jobs.

The hypothesised mechanism underpinning the relationship between economic exposure and zero-sum discourse can be summarised as follows: living in an area experiencing industrial decline means being exposed to a narrative in which members of previous generations remember times when they could sustain a comparatively higher standard of living. When combined with the lack of immediate economic alternatives compatible with the green transition, this feeling could develop into a structured belief in the fact that progress towards a greener future does not, in fact, benefit anyone else other than those who already have better life chances, thus pro-

ducing ubiquitous *loss*. Such fear of future downward mobility would then manifest itself in public discourse.

2.3.3 Methodological pursuit

This study's goal is to advance knowledge on the topic by integrating standard quantitative approaches with those of computational social science. Using the web as a data source, online news outlets and YouTube videos will be analysed through natural language processing methods borrowed from computational linguistics. Through the application and comparison of different analytical strategies, this study will offer a complete overview of the phenomenon.

2.4 Case selection

The strategy of choice is that of a comparative case study. For this investigation's purpose, the ideal cases would be two areas comparable in terms of their cultural and political background, but displaying variation in terms of the labour market conditions above. The Umbria region in central Italy thus represent a good candidate. Its two main cities (Perugia and Terni) have a similar stock of residents, respectively 162099 and 106436 in 2024, and share a common cultural and political background, their polities having been intertwined at least since 1970, when they were chosen as the two provincial capitals of the newborn Umbria region (ISTAT 2025). Their economic and social history, however, is deeply different.

CHAPTER 3

The case of Umbria

3.1 Umbria as a case study

Terni has a long industrial history, its identity revolving around its steel factory. After decades of decline and job loss, the factory now employs a fraction of the people it did in the 1980s, while still exerting a very negative impact on the city's air quality. Terni is a textbook case of noxious de-industrialisation, the extensive employment losses becoming even more painful as they are the proof of multiple labour failures over decades (Portelli 2023, 2017). Perugia, on the other hand, thrives off of its long academic tradition, its economic activity being mostly centred around the service economy, hospitality and confectionery. De-industrialisation is mostly unknown to the population of a city that was simply never truly industrialised. Its expansion over the last twenty years made transportation become the main topic of political debate.

Perugia and Terni clearly display very different characteristics in terms of the labour market conditions: the experience of industrial decline of noxious deindustrialisation so familiar to Terni's population is completely alien to Perugia's, where heavy indus-

try plays an almost quantitatively irrelevant role in the local economy. Nonetheless, their polities and politics have been closely intertwined for at least half a century, due to the fact that they are the only two provincial capitals of the Umbria region. These factors make them perfect to conduct a comparative case study on.

3.2 Terni's industrial history

Up to the 1880s, Terni was a medium-sized rural town of no particular interest other than the occasional visit from English painters and poets to the nearby waterfalls or the castles along the Nera valley. Although, as stressed by Vaquero Piñeiro (2020), the area was historically no stranger to manufacture, the establishment of the *Società degli altoforni, fonderie e Acciaierie di Terni* ("Society of blast furnaces, foundries and steelworks of Terni") in 1884 was a defining moment in the city's history, turning it into a full-fledged industrial town. The steel factory underwent numerous name changes as it expanded, and was finally nationalised in the 1930s. With the steel came weapons, chemicals, electrical plants and pipe mills. The production, which initially comprised solely of stainless steel, came to include cutting-edge magnetic steel, engines, turbines, all while pulling the city toward higher incomes and better education, at the expense of air quality and intense urbanisation. Since then, Terni's identity evolved around heavy industry. Its urban design was catered to the needs of its industry, before and after being heavily bombed during WWII due to its strategic relevance. Workers' dwellings and guest quarters still crowd every area of the town. A bright-green, 12-ton steel press greets the occasional visitor right in front of the train station. In the city centre, a loud siren resounds every day at noon signalling the end of the morning shift at the factory.

In recent years, however, the town's industrial heritage started clashing with its present. After the steel factory's privatisation in 1994, the new German-led corporate ownership immediately exported patents to Germany and started progressively cutting down on personnel. A first *traumatic* event for the town was to witness Thyssen-Krupp's decision to move the highly coveted production of magnetic steel to Germany in 2004, non-withstanding the massive protests which flooded the city (Portelli 2023). In 2012, Finnish group Outokumpu tried acquiring the whole of Inoxum (Thyssen-Krupp's inox steel division), but was forced by the European Commission to give up Inoxum's assets in Italy, including *Acciai Speciali Terni* ("Terni Special Steel", which was then the factory's new name) (Bellomo 2012; Meneghello 2012). After looking for a buyer for two years to no avail, Outokumpu gave AST back to Thyssen-Krupp, which then decided to downsize it in order to sell it more easily (TerniOggi 2013). To this end, Lucia Morselli (whose nickname was already *la tagliateste*, the axewoman) was appointed CEO, with the explicit mandate of dramatically cutting down on the workforce. This sparked an unprecedented struggle between the ownership and the workforce. The tension became unsustainable, with all-out strikes being called by every union, and Morselli being even detained in her own office a whole night by protesters (Ugolini 2014; UmbriaJournal 2014). The result of the 2014 struggle, however, was considered by many an utter and complete loss for unions and the workforce. Morselli offered the protesters "compensation" (equal to the unpaid salary for the time they spent on strike), which most workers accepted (Portelli 2023).

The personnel of a steel factory that used to employ over 10000 people up to a few decades ago, now amounts a mere 2300 units ("AST: ARVEDI AST" 2025). The in-

dustry's adverse effects on public health, however, did not falter one bit. We can consider Terni to be a textbook case of noxious deindustrialisation, as Feltrin, Mah, and Brown (2022) define it, since widespread job loss came with no reduction in toxic pollution. The historical presence of three waste incinerators (of which only one is currently operative) unsurprisingly contributes in fuelling a considerable amount of spontaneous public health and environmental activism (Toni 2024; TerniToday 2020). From the 2010s onwards, a wide range of local pressure groups emerged: one movement (*No Inceneritori*, "no incinerators") became especially popular and influential, while the high school students at the head of Terni's *Fridays for Future* branch assumed a prominent role in its national organisation. Although the COVID19 pandemic put a noticeable dent in the city's grassroots political activity, this moment of its recent history is still fresh in Terni's collective memory. The ubiquitous feeling of *loss* generated by the city's industrial decay in Terni's population, joined with its recent mobilisation around environmental policy, makes it a perfect case to be here analysed.

3.3 Perugia as a comparison case

After centuries as part of the papal state, Perugia's economic history is essentially one of stagnation up until the 20th century. In 1922, Hotel Brufani was the starting point of the fascist March on Rome (Albanese 2006). This made Perugia suddenly become a symbol to the newborn regime, which financed significant embellishments to the city. Its industry, however, remained largely underdeveloped. The city's marginal strategic importance meant that it was mostly spared by Allied bombers during WWII.

The city of Perugia is also home to one of the Europe's oldest universities (*Università di Perugia*, founded in 1308), as well as to the University for Foreigners (*Università per Stranieri di Perugia*, UNISTRA) (The Editors of Encyclopaedia Britannica 2020). Perugia's rich cultural heritage was a decisive factor in 1970, when it was chosen to become capital of the newborn Umbria region.

Its economy nowadays is mostly based on the service economy, driven by festivals and large cultural events such as *Umbria Jazz* and the Journalism Festival. Although heavy industry does play a (quantitatively marginal) role in non-urban areas in the province of Perugia, manufacture-related economic activity in the city's immediate vicinity pertains mostly to the confectionery and textile sectors. Notable examples of the like are *Perugina*'s headquarters and *Nestlè*'s main Italian branch, as well as *Luisa Spagnoli*'s textile plant. A considerable amount of smaller mechanical and artisanal firms bloomed in the outskirts of the city, which expanded almost incontrollably during the last 20 years.

CHAPTER 4

Data and methods

In this chapter, the data sources and the analytical methods used will be explored and discussed. First, a description of the data collection process will serve as an introduction to the two corpora themselves. The difficulties encountered and the resulting limitations in the quality of the data will be here documented and thoroughly examined. Second, the process of text retrieval, as well as the analysis and mining techniques employed in the following chapters will be identified and explained.

4.1 Sample selection

Having now established that media will be here considered a viable a proxy for public discourse, we turn to the task of constructing a sample of it to be used in this analysis. In the current age of digital communication, a *multi-modal* approach is in order. To capture a wider range of political communication, we turn here to both written text and video.

4.1.1 Transcribing YouTube videos

YouTube has long been home to a rich and diverse collection of communicative content. Containing both structured press reports by *traditional* news outlets and bystanders' contributions, it proves extremely valuable in obtaining an overview of public discourse that is not just limited to what professionals and *insiders* think or write. Since, for the purpose of the present study, we are interested in what are the most *visible* opinions or points of view, we could even take advantage of YouTube's own retrieval system. The videos included in the corpus were selected through automated querying. A Python script was set up to cycle through a set of keywords associated with each of the two cities. Every video was then converted to audio and transcribed using OpenAI's Whisper API (OpenAI 2022). This method, however, presented a few critical difficulties, connected to the nature of the medium itself. First of all, not all videos identified by the queries were relevant to the object of this study. Keywords connected to the biosphere and ecosystems were often linked to videos catered to tourists, as they overlapped with the image of Umbria as Italy's green heart, which has nothing to do with environment policy per se. In other cases, the match between title and keyword was completely coincidental. For instance, the keywords "spazi verdi" (green areas) matched with a video essay about renovating Terni's old theatre, Teatro Verdi. Another problem was that the quality of audio recordings was often very poor, and the dialogue delivered in the local dialect. This is equally true in the cases of demonstrations and press conferences. Professional politicians tend to drift from a linguistic register to the other, as well as between standard italian and Umbrian dialect, within even a few sentences of the same speech. Steel workers, on the other hand, tend to use very common idiomatic expressions such as *me ne volevo*

anna' via ("I wanted to leave"), or *ce stanno a pija' pe' lu culo* ("they're screwing with us"), which Whisper was not able to transcribe adequately. The simplest (although definitely time-consuming) solution was manually clean the data, preparing it to be *digested* through natural language processing.

4.1.2 Scraping local news articles

More *traditional* textual media was not overlooked. Three on-line media outlets were chosen to be part of a second corpus:

Corriere dell'Umbria is an established regional newspaper, available both on-line and in print. Due to its sheer size, it has become somewhat of a staple in local press, notwithstanding its relatively young age (it was founded in 2007). Its articles prove particularly useful to a comparative study of this sort, as two of its six newsrooms are based in Perugia and Terni.

Terninrete and *Perugia Today* are some of the oldest *purely* on-line newspapers in Umbria (founded respectively in 2009 and 2011). Embracing the fast pace of internet-based communication, they tend to publish short, rapid-fire articles. As will be discussed in a few paragraphs, this clearly impacts the quality of sentiment analysis results: *fast journalism* tends to be closer in tone to news agency flashes, as in more neutral in their word choices. Their responsiveness, however, is of great use in investigating a topic's salience.

4.2 The corpora

Two corpora were built: the first comprises of 530 transcriptions of YouTube videos selected through keyword querying, the second comprises over 150000 articles published by three major on-line news outlets. Although we concede that the tone and topic composition of public discourse might differ from what can be read in the press or found on YouTube, it would be reasonable to expect that news outlets would have commercial reasons to mirror what is talked about "in the streets", and that on-line, decentralised media platforms (like YouTube) could offer some, more or less direct, insight into public discourse.

4.2.1 YouTube

As Figure 4.1, the distribution of the videos' upload date is skewed towards more recent results. The number of usable videos in Perugia is also considerably higher than those in Terni. While the difference between the number of videos per city is easily explained through population and the vastly different amount of political activity (Perugia being the regional capital), the distribution of videos over time could offer some insight on the salience of each topic.

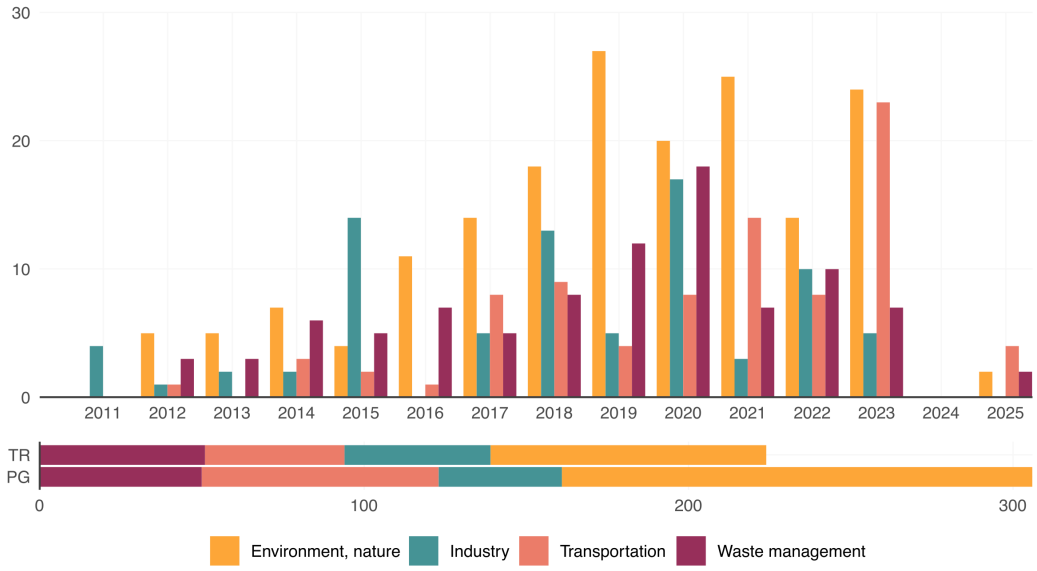

Figure 4.1: Transcription distribution over time, by query category and city

4.2.2 Umbria Press

The second corpus in this study comprises of articles from three on-line local newspapers: *Terninrete*, *PerugiaToday*, and *Corriere dell'Umbria*[1]. The first two are linked to the cities of Terni and Perugia, respectively, as their articles refer strictly to these two cities. *Corriere dell'Umbria*, on the other hand, is a regional news outlet, with separated newsrooms for the two cities. This was taken into account, and the articles divided accordingly.

The corpus' composition is represented in Figure 4.2. A few of its features are evident to the eye: first of all, *PerugiaToday* is by far the most prolific news outlet of the

[1]The technology used to retrieve all articles in machine-readable form was the R package `rvest`, which allows the user to identify different sections of an html page by specifying their CSS attributes and store the text associated with each one in a data frame (Wickham 2024). Parallel sessions were set up taking advantage of the `future` package, in order to reduce run times (Bengtsson 2021). The complete dataset is available on-line, complete with the source code for the scrapers (Mattioli 2025).

three, with more than 125000 articles published. In general, the two *fast-journalism* outlets (*PerugiaToday*, *Terninrete*) flood the corpus with information, making the couple hundreds of articles from the *Corriere* almost disappear.

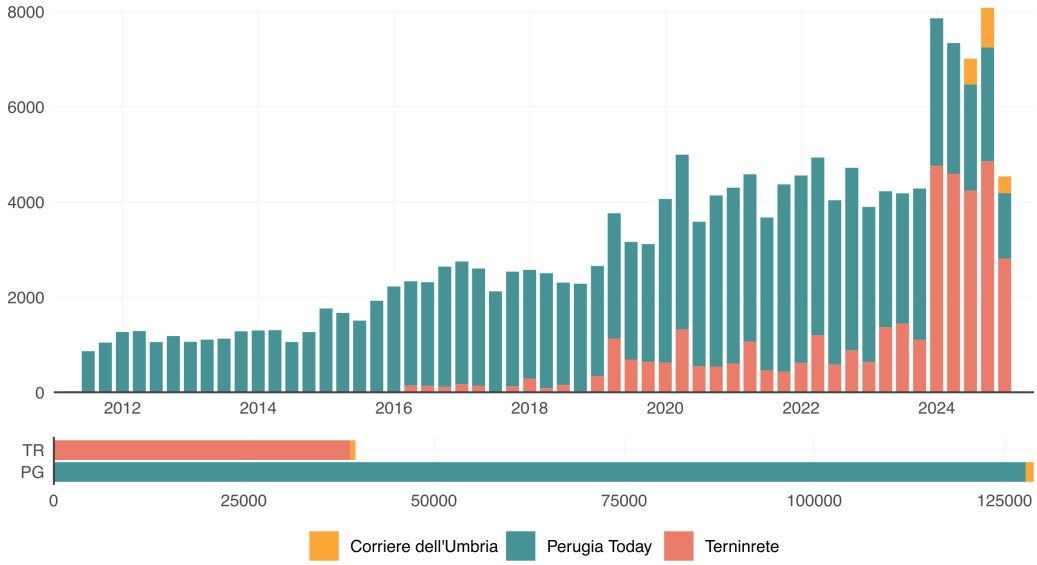

Figure 4.2: Article distribution over time, by newspaper and city

4.3 Analytical strategy

This study must be understood as one of exploratory nature. Its aim is not to *explain* a phenomenon, rather to establish its existence (Merton 1987). After a thorough description of the corpus, it will focus on two dimensions of public discourse: salience and polarisation. The tools employed to do so are here explained and compared.[2]

[2]The analysis was conducted entirely using the statistical coding language R and mostly adheres to the *tidy data* principles defined by Wickham (2014), thus relies heavily on the `tidyverse` family of packages (Wickham et al. 2019).

4.3.1 Putting ideas in context

Examining the context in which a word is placed is paramount in understanding the different ideas to which it is connected. KWIC (KeyWords In Context) is probably the most basic approach to this task: by simply querying a keyword, one can retrieve its immediate context and qualitatively assess it. The R package `quanteda` provides the researcher with a very simple interface to do so (Benoit et al. 2018).

4.3.1.1 Ranking words

Being able to rank words in terms of their relative importance to each document is a very useful tool to dissect a corpus. The standard method to do this is a simple statistical measure called tf-idf. The index is decomposable in two parts, term (relative) frequency and inverse document frequency. For a term "t" and a document "d":

$$\text{tf}_{t,d} = \frac{\text{frequency of term } t \text{ in document } d}{\text{total number of terms in document } d}$$

$$\text{idf}_t = log_{10} \frac{\text{total number of documents}}{\text{total documents with term } t}$$

Tf-idf is computed as $\text{tf-idf}_{t,d} = \text{tf}_{t,d} \cdot \text{idf}_t$. By finding the highest tf-idf in each document or topic, one can qualify the context in which a concept or construct lives.

4.3.1.2 Finding meaning *between* words

Occurrences within the corpus can be represented as relational data, allowing us to build networks of words. Using bigrams as a unit of analysis, the relationships between words will be analysed and dissected graphically (Pedersen 2024a, 2024b).

A relatively new approach used in text retrieval and mining is word embeddings. A very basic explanation of what word embedding means would be finding a numerical representation for each word in a corpus (Jurafsky and Martin 2025). What a word embedding model does is basically finding a point in a multidimensional space representing the *meaning* of each word relative to all others. The distance of each word in this *semantic* space can be then understood as a measure of their difference in meaning. Word embeddings are often used as a featurisation step for supervised classification models. In this study, however, they will be used mostly as a retrieval device, aiding us to find the words which are closest to pre-identified keywords relating to environmental policy.

4.3.2 Sentiment analysis

To acquire insight on the tone of public discourse around environmental policy, dictionary-based sentiment analysis was conducted. In very pragmatic terms, this means joining a dictionary of positive and negative words with the text contained in the corpora, then analyse the frequency of matches. Although this kind of sentiment analysis can be seen as a rather crude instrument, it can still prove useful in exploring the corpora (Silge and Robinson 2017). The biggest obstacle encountered while working on this section of the study was the difficulty in finding a suitable

dictionary. While dictionaries for the English language are more than abundant, acquiring one for Italian is no simple task. Although some extremely valuable work was conducted by Basile and Nissim (2013) more than ten years ago, their dictionary (SENTIX) requires PoS tagging, which was not possible to perform on the corpora in this study. The solution was to use the multilingual lexicon published by Chen and Skiena (2014). As is often the case with dictionary-based techniques, some minor tweaking was necessary. The word *rifiuto* ("rubbish"), for instance, was listed as negative, since it can be used as an insult when referring to a person. Of course, being our topic of interest connected to waste management, the word had to be excluded from the dictionary altogether.

CHAPTER 5

Results

In this chapter, the analyses conducted on both corpora will be presented and discussed, in light of the historical and political context already explored in chapter 3.

5.1 Analysing the YouTube corpus

As was already discussed in the previous chapter, the present corpus contains relatively low-quality data. The unstructured, colloquial communicative style of YouTube videos, together with wavering sound quality and intelligibility, limit greatly the range of viable analytical tools. Nevertheless, one must never underestimate "how much data science you can do with just counts and a little basic arithmetic" (Wickham, Çetinkaya-Rundel, and Grolemund 2023). As will be visible in the next few paragraphs, there is more than enough meaning to extract from this dataset.

5.1.1 Salience

The first question to be answered is: how much do people talk about environmental policy in Umbria? Is it salient at all? And then again: *when* was it more salient? To answer these questions, simply counting the amount of documents in which terms connected to environmental policy occur is the simplest, but probably most effective strategy.

The amount of videos retrieved per city varies drastically from a query category to another. As one would expect, numerous videos related to heavy industry can be found about the city of Terni. Their amount peaks in 2015, when the protests climaxed, then drop immediately the next year only to peak again in 2022, when the Arvedi group bought the steel factory from Thyssen-Krupp (Terni in Rete 2022). Shifting from foreign to Italian property was a very conflictual topic.

Similarly, transportation is a very hot topic in Perugia. The videos in this category tend to be very positive short documentaries about good practices in sustainable transportation, especially around 2023. A considerable peak can be detected in 2021, year in which the *Bus Rapid Transfer* (BRT)[1] service was established, requiring considerable investment and urban rearrangement. Sustainable mobility became more salient in Terni recently, due to new cycling infrastructure being under study. A peak in 2017 was caused by the news of a considerable investment by the municipality in bus transportation.

Waste management tends to be generally more relevant in Terni, home of three incin-

[1] The BRT was defined by the municipality as an "innovative electric transport system, based on an advanced road transport concept, with particularly high standards, characterised by low emissions and high transport capacity" (Comune di Perugia, n.d.). In plain English, electric buses.

erators. The discourse around them became particularly heated in 2020, when one of them was finally shut down (Tuttoggi 2020).

Finally, the salience of environment- and nature-related posts in Terni was apparently influenced by the relatively crowded Fridays for Future protests in 2019. The movement's popularity prompted local research centres to publish *Sentieri* and *Mal'aria* (Zona et al. 2019; Legambiente 2019), two very influential studies on air quality in the area, and local politicians to take on the problem. The same thing cannot be said for Perugia, where the results of environment-related queries comprise mainly of videos catered to tourists.

At a first glance, environmental policy seems to be more *politicised* in Terni than in Perugia, and its salience linked to social movements' success.

5.1.2 Tone and polarisation

At a superficial level, sentiment values in each category seem to be fairly balanced between Perugia and Terni. Looking at individual keywords, however, reveals more accurate insight. The videos retrieved with the "Industry" query show an interesting discrepancy: the ones connected to more *general* terms, such as *industria* (industry), or even *acciaieria* (steel factory) are markedly positive. It is only when one specifically looks for the exact name of the steel works in Terni (Acciai Speciali Terni - AST) that one gets polarised results. This is due to the fact that the more general keywords tend to match to political press releases, which are usually characterised by an either artificially neutral or markedly positive tone. The only exceptions are the few videos taken during protests, displaying very negative sentiment values. The transportation category in Perugia offers some evidence in the opposite direction:

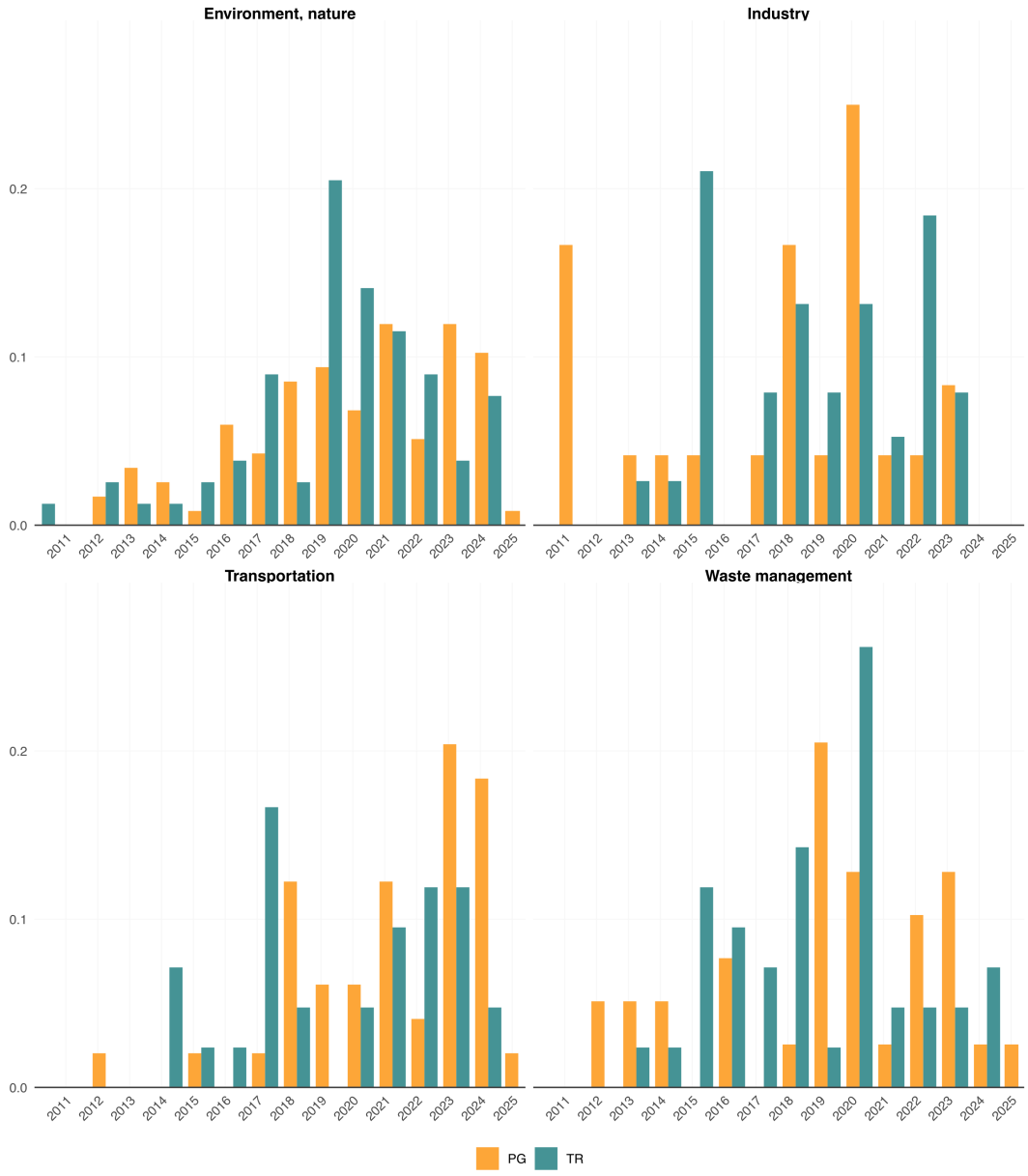

Figure 5.1: Relative salience per month by query category, computed as monthly percentage of videos over total within category

more general queries about buses retrieve mostly negative videos about the current state of public transport, whereas querying a specific project (BRT) retrieves enthusiastic political speeches about new public investment. Waste management in Terni is especially interesting, the number of videos about the incinerator with positive and negative sentiment being almost exactly the same.

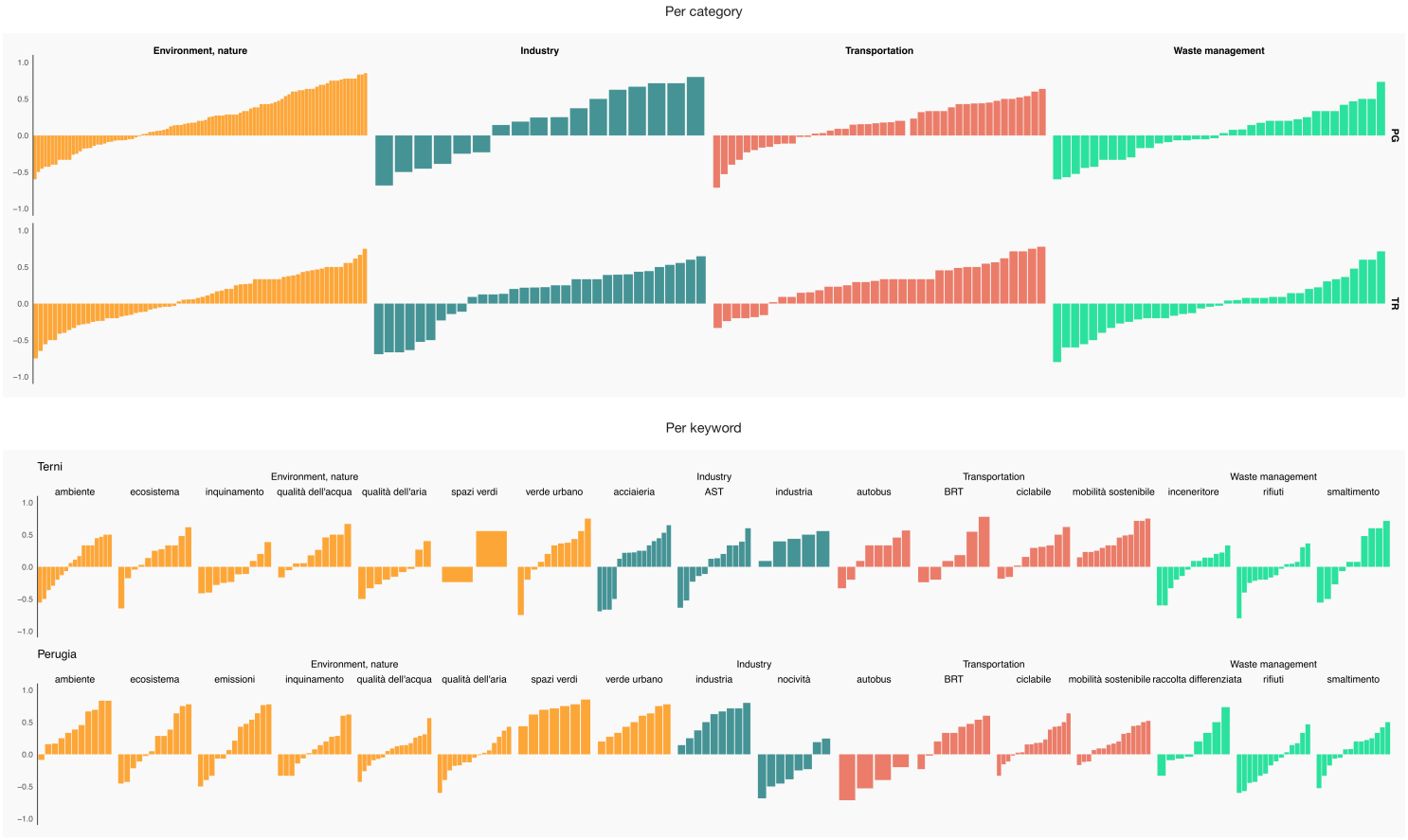

Figure 5.2: Dictionary-based sentiment analysis on YouTube videos, computed as ratio of positive over total matches per video

5.1.3 Co-occurring topics

In general, we can confirm that environmental policy is more polarised and *politicised* in Terni, as even the sentiment computed on the basic query "environment" is extremely polarised, as opposed to Perugia, where videos collected with the same query are clearly positive. The problem of air quality seems to drive negative sentiment on environment-related topics in Terni.

The issue of insufficient air quality in Terni is clearly connected to the *noxiousness* of its industrial activity. However, when videos talk about industry, they almost never mention environmental policy. If one searches for videos about industrial activity in Terni, they will hear absolutely no mention of its effects on the environment. Nevertheless, a significant portion of the videos talking about the environmental conditions of the same town mention industrial activity at least once. This fact speaks volumes about the population's priorities: that the relationship between heavy industry and air pollution is definitely recognised, but it is simply deemed irrelevant in a discussion on industrial planning.

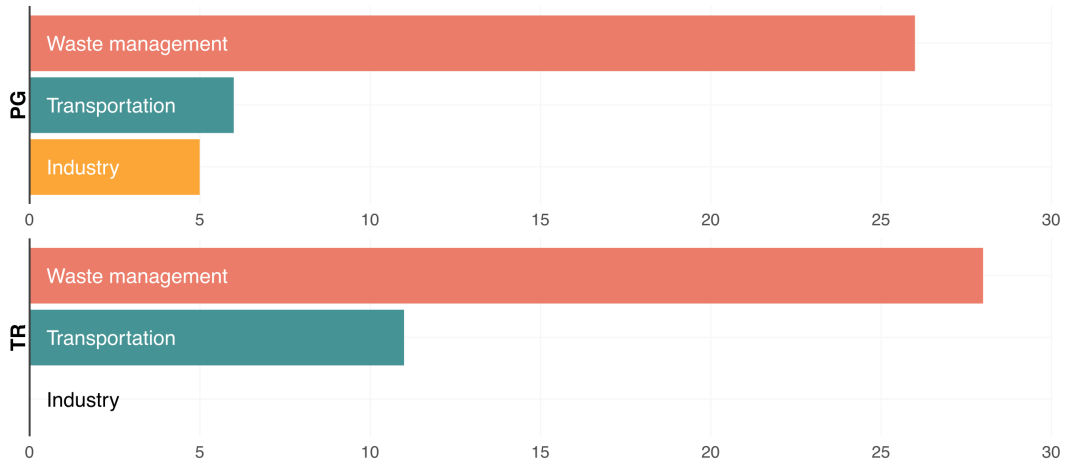

Figure 5.3: Number of videos not retrieved using environment-related queries in which environment-related keywords appear, by city.

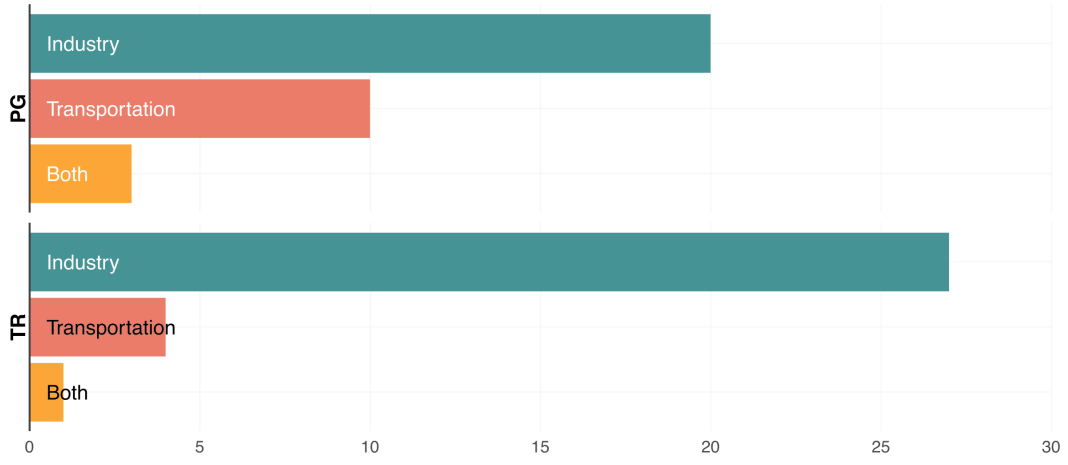

Figure 5.4: Occurrences of keywords pertaining to industry and transportation in videos retrieved using environment-related queries, by city.

5.2 Analysing the *UmbriaPress* corpus

Due to the nature of the data source, the quality of the text contained in UmbriaPress is considerably higher than that of the YouTube corpus. Together with its much bigger size, this means that more sophisticated modelling can be a viable option. The analysis still begins with keyword-based retrieval and counting, to investigate salience. An exploration of how environmental policy is defined in the corpus was made possible by treating co-occurring tokens as nodes of a relational database and training a word-embedding model. The tone of public discourse around environmental policy was analysed through dictionary-based sentiment analysis.

5.2.1 Salience

Consistently with what was found in previous analyses, the salience of transportation and industry follow very different paths in Terni and Perugia. A keyword-based retrieval returns pretty stable results in Perugia, where environmental policy comes under the spotlight in 2019, following the success of Fridays for Future, only to then yield its spot back to transportation, the main topic of discussion in the city. The situation in Terni is more erratic, each topic falling in and out of *fashion* way more easily.

Interestingly enough, the co-occurrence pattern detected in the YouTube corpus is partially confirmed by this second chunk of the analysis. In Terni, 22% of the articles containing terms connected to Environmental policy also include Industry-related terms. This relationship is not completely symmetrical: the articles in which Industry and Environment-related terms co-occur represent merely 15% of those in

which Industry-related tokens appear. This connection is way less significant in Perugia, where the percentages never reach 10%. Transportation and Environment are also less often connected in Perugia than in Terni. This might offer a hint about how environmental policy is discussed in the two cities: Perugia shows a more compartmentalised situation, where each topic is treated individually; whereas in Terni, environmental policy is often connected to other topics.

Figure 5.5: Topic salience by city

5.2.2 From salience to meaning

The relationship between words may help us in discovering both the definitions and connotations of certain words in our corpus. Representing the co-occurrences of significant terms and their immediate neighbours as relational data is a very valuable technique in this sense.

5.2.2.1 Bigram networks

The terms with the most connections, both in Terni and Perugia, are those relating to waste management. Interestingly enough, the word "waste" (*rifiuti*) is connected to "pollution" (*inquinamento*) in Perugia, but not in Terni.

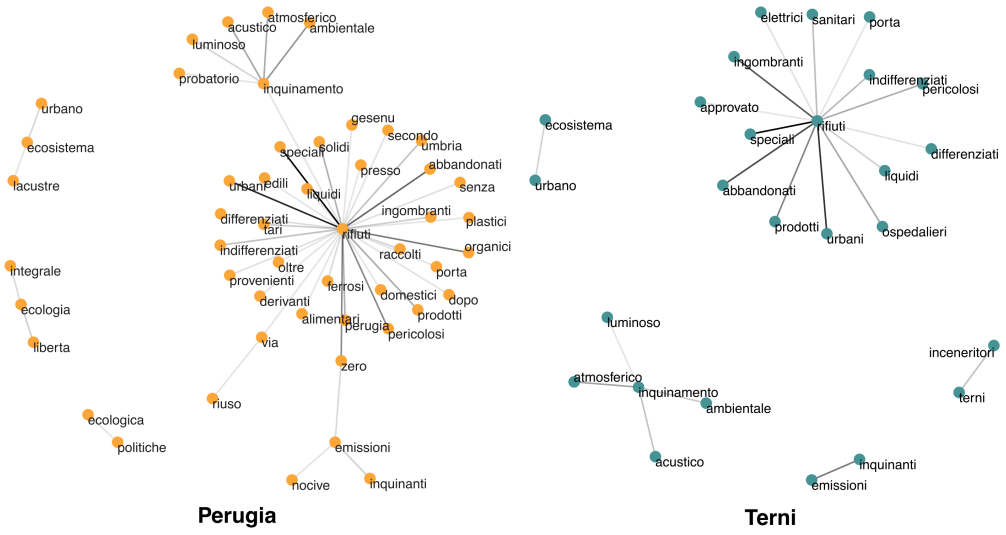

Figure 5.6: Bigram relationship graph

5.2.2.2 Word embeddings

After splitting the data in two sub-corpora, one for each city, a word-embeddings model was able to represent words as points in two multi-dimensional spaces, one for each sub-corpus. Performing Principal Component Analysis (PCA) on the values associated to relevant terms in each one lets us read them in a *topographic* representation, the nearest words being more closely related semantically.

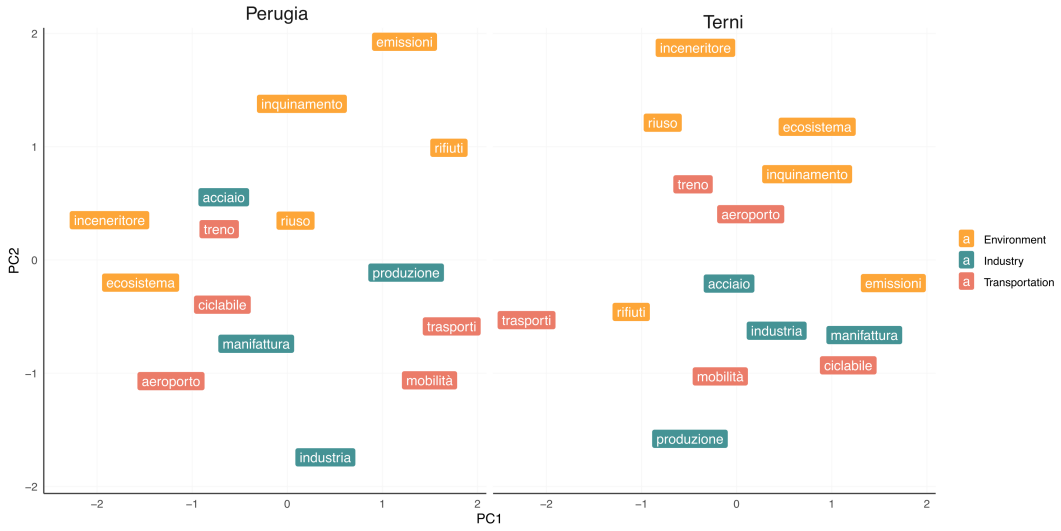

Figure 5.7: Principal Component Analysis (PCA) of word embeddings in the sub-corpora containing respectively all articles published in Terni and in Perugia

A first cluster could be identified in the Terni subcorpus around industry. (Heavy) industry in Terni is connected to polluting emissions and toxic waste. "Waste incinerator" (*inceneritore*) is instead found at the top of the representation, together with the words "ecosystem", "reuse", and "pollution". This might explain the asymmetry in co-occurrences between the two queries "Environment" and "Industry": while some areas of environmental policy (namely air pollution) are connected to industrial

activity, waste management is treated separately.

The same cannot be said for Perugia, where "pollution", "emissions", "reuse", and "waste" all coexist in the otherwise lonely top-right corner, while waste management is closely related to transportation, and industry-related terms seem to float without an apparent order.

These results contribute to explain the differences in salience and *politicisation* of environmental policy between the two cities. While the *perugini* see the environment as a coherent, unified topic of discussion, the *ternani* clearly mark a difference between *free-standing* policy areas, such as *civilian* waste management, and those connected to the *noxiousness* of economic activity, such as air quality and toxic waste.

5.2.3 Sentiment analysis

Through dictionary-based sentiment analysis, an exploration of the tone in each of the three main topics analysed is possible. At a first glance, no real difference can be identified between the two cities, despite a certain variability between topics being present.

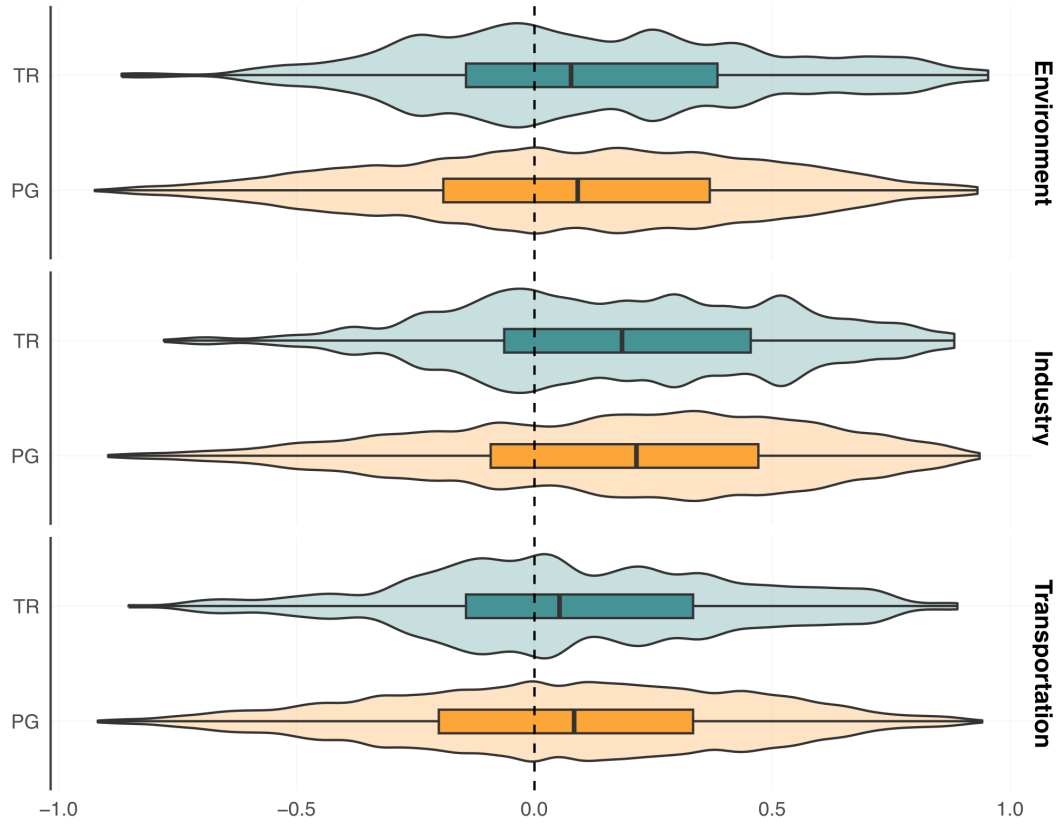

Figure 5.8: Distribution of the difference between positive and negative word count divided by the total, visualised by topic and city.

5.2.3.1 Restricting the corpus

A first element to be taken into account is the differences in tone across newspapers: the *fast-journalism* outlets Terninrete and PerugiaToday tend to have a very neutral tone, due to the adoption of a very short format and the subsequent shallowness of their content.

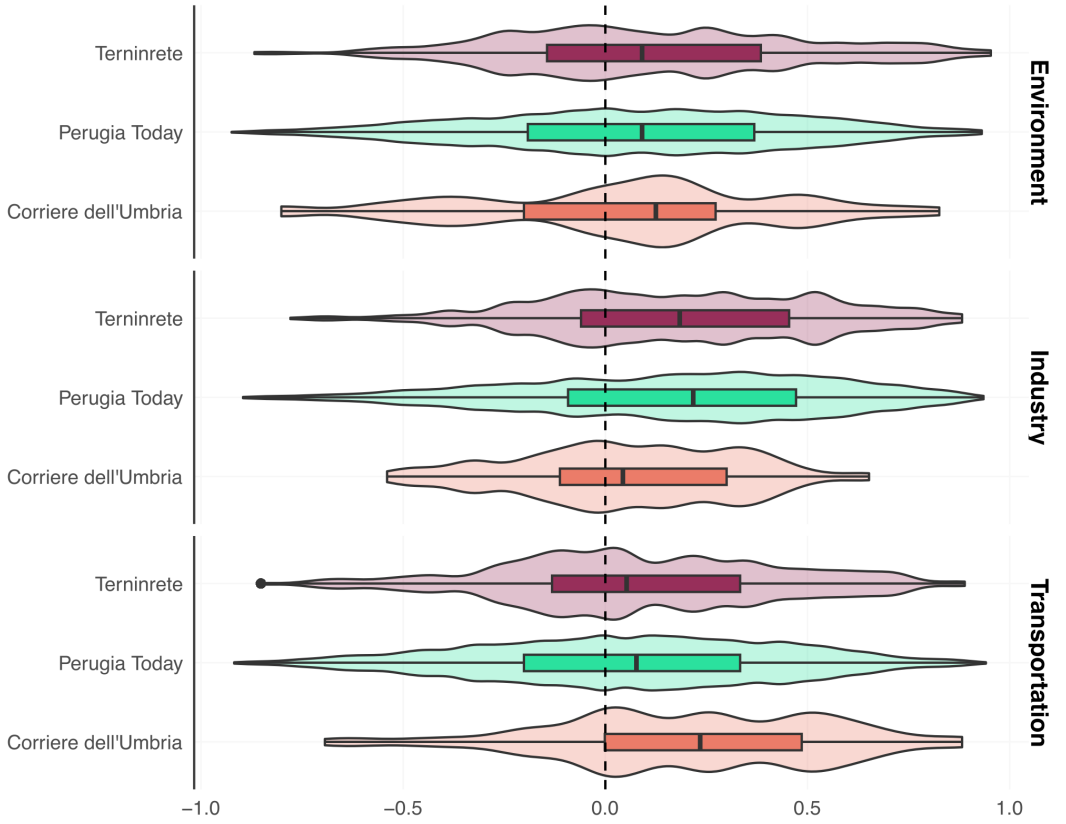

Figure 5.9: Distribution of the difference between positive and negative word count divided by the total, visualised by topic and newspaper

By restricting the corpus only to the articles published on *Corriere dell'Umbria*, one can easily see how sentiment values become more erratic: articles being way more negative in Terni than in Perugia in general, but especially when talking about environmental policy. However, the articles driving the negative *tail* of the distribution are, for the most part, related to single episodes or individuals' behaviour (e.g. illegal dumps in the countryside).

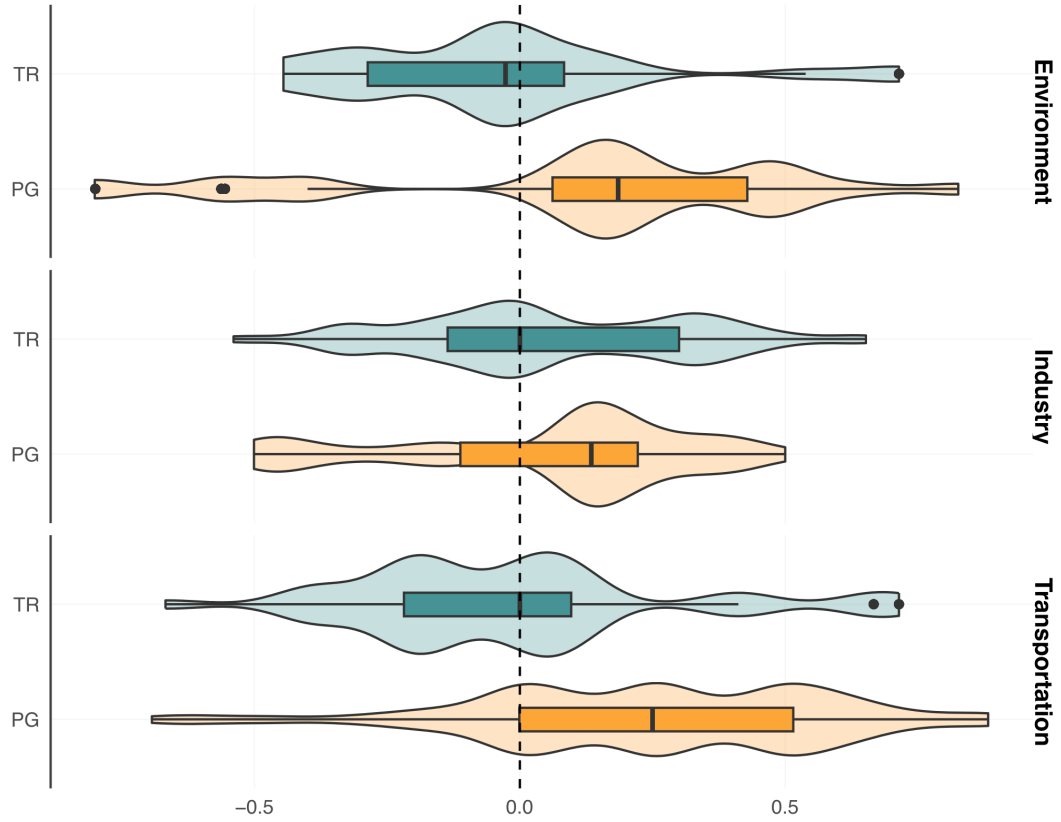

Figure 5.10: Distribution of the difference between positive and negative word count divided by the total, visualised by topic and city (sample restricted to articles published on Corriere dell'Umbria)

Due to the extremely small size of the resulting sample, however, it is necessary take these last results without a grain of salt. The only takeaway from this last chunk of the analysis is that future research should take the *type* of journalism into account at the moment of compiling a corpus.

5.2.3.2 Contextual sentiment

By retrieving the tokens found immediately before or after terms connected to environmental policy, one can get a feel for *where* these keywords appear, as in their

context in the overall text. By performing sentiment analysis on these *contextual* tokens, we can infer how environmental policy is conceived by the public in each city.

The sentiment appears generally more negative in Terni, which is hardly surprising considering that it has the regional record for the worst air quality, and overly complicated waste management conditions. Terms connected to ecology and nature, however, are markedly positive in Terni. A possible explanation could be an idealisation of *nature* as opposed to the noxious urban environment.

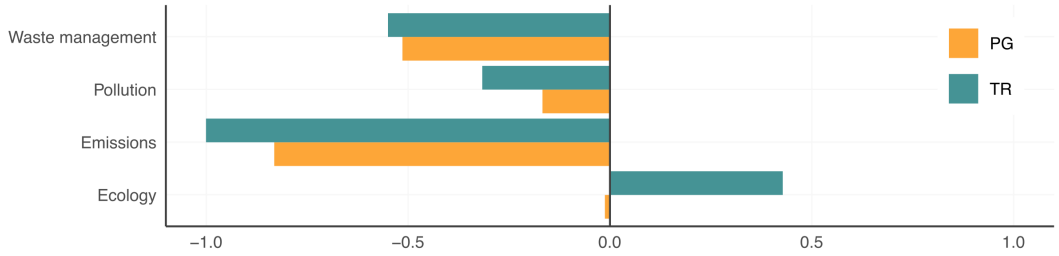

Figure 5.11: Sentiment values computed on KWIC - trigrams centred around relevant keywords

CHAPTER 6

Conclusions and discussion

6.1 Discussion

The salience of environmental policy displays rather different trajectories in Terni and Perugia. Its distribution over time follows social movements' and interest groups' success rather closely in Terni, while it seems rather independent from other policy areas in Perugia. This emerges from the analysis of both the YouTube and UmbriaPress corpora: Environmental policy in Perugia holds a stable place in public discourse, while its salience seems more issue-oriented in Terni.

Topic co-occurrence adds a significant piece to the *puzzle*: a larger portion of public discourse around environmental policy in Terni is related to the *noxiousness* of industrial activity, both on YouTube and in online news articles. The opposite, however, is not true. While the proportions of what we could call topic *overlap* is somewhat stable in Perugia, they are almost completely one directional in Terni: though articles and videos about the environment trace a connection with industrial activity, articles and videos about heavy industry do not mention environmental policy whatsoever. This asymmetrical relationship is further substantiated by a semantic analysis of the

corpus. The *overlap* is not only identified in terms of term occurrence in the corpora, rather it shows even as a coexistence of terms connected to air quality and industry in the same semantic space.

The emerging picture is that of environmental policy as a highly *politicised* topic in Terni. This makes its salience *fluctuate* together with the popularity and success of social movements and interest groups who centre their political activity around climate change or environmental policy. This was especially the case between 2017 - the year in which the relevance of anti-incinerator mobilisations peaked - and 2019, when *Fridays for Future* came under the spotlight.

These actors' tendency to connect it to industrial *noxiousness* is reflected in the media, but industrial activity seems to be discussed without ever taking environmental concerns into account. This feeds back to Vona's claim that job security is of utmost priority even for those who suffer the most from the adverse effects of toxic pollution (Vona 2019).

Differences in tone between between Terni and Perugia are subtle and nuanced. They appear mostly negligible when analysing news articles as a whole, but become rather interesting if one restricts the analysis to the keywords' immediate context. This kind of analysis unearths surprisingly high sentiment in Terni around ecology-related keywords, and generally low values in every other one. This might hint at an idealisation of nature as opposed to a noxious urban environment.

6.2 Limitations

This study has a series of limitations. First and foremost, the retrieval techniques employed here employed are rudimentary. While quick and computationally *cheap*, keyword-based techniques are inherently prone to error. Future developments of the present research might benefit from a more "sophisticated" approach to text retrieval, i.e. a classification model trained to classify news articles and press releases based on their content. Second, the data collection process directly influences the results of text analysis. Each different language variety included in a sample of local political communication has a series of idiosyncrasies, especially in terms of tone. What this means is that in order to accurately assign a sentiment value to a chunk of text, it would be necessary to take into account the political alignment and stylistic habits of at least every news outlet. Finally, and most importantly, this study does not advance any explanatory claims. Its role is paving the road for future contributions to define the causal chain linking patterns of loss to the support for environmental policy.

6.3 Conclusion

Empirical results highlight an interesting connection between salience and *exposure*. When faced with a trade-off between job security and environmental concerns, the public's priorities seem to lie with job security. This is certainly consistent with an interpretation of public discourse around environmental policy centred off patterns of *loss*, although the way in which it manifests in media is different than expected. Salience (or rather, its absence) appears to be more telling than tone. The most distinctive trait of Terni's public discourse is its utter disregard of environmental

concerns when discussing industrial planning.YouTube videos concerning industrial policies were, in fact, completely void of information on environmental policy. In the same corpus, however, a connection between toxicity and industrial activity can be identified. Additionally, "nature" as an idealised concept has an extremely positive connotation in a (formerly) industrial town like Terni in comparison to the "baseline" case of Perugia. In the context of this study's proposed framework, this could be interpreted as a way to avoid what is perceived as a zero-sum game. What is left for future research to ascertain is whether it is, in fact, a fear of loss that causes these differences in the connotation of public discourse, and to investigate the subsequent connection between public discourse and electoral results, as well as public opinion formation.

References

Albanese, Giulia. 2006. *La marcia su Roma*. 1. ed. Roma [etc.]: GLF editori Laterza.

"AST: ARVEDI AST." 2025. https://www.acciaiterni.it/chi-siamo/.

Basile, Valerio, and Malvina Nissim. 2013. "Sentiment Analysis on Italian Tweets." In, 100107. Atlanta.

Bellomo, Sissi. 2012. "L'inox di Thyssen a Outokumpu." http://www.ilsole24ore.com/art/finanza-e-mercati/2012-02-01/linox-thyssen-outokumpu-064544.shtml?uuid=AaMmGGlE.

Bengtsson, Henrik. 2021. "A Unifying Framework for Parallel and Distributed Processing in r Using Futures" 13. https://doi.org/10.32614/RJ-2021-048.

Benoit, Kenneth, David Muhr, and Kohei Watanabe. 2021. "Stopwords: Multilingual Stopword Lists." https://CRAN.R-project.org/package=stopwords.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An r Package for the Quantitative Analysis of Textual Data" 3: 774. https://doi.org/10.21105/joss.00774.

Bez, Charlotte, and Maria Enrica Virgillito. 2022. "Toxic Pollution and Labour Markets: Uncovering Europe's Left-Behind Places." *LEM Working Paper Series*.

Calvin, Katherine, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter W. Thorne, Christopher Trisos, José Romero, et al. 2023. "IPCC, 2023:

Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups i, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, h. Lee and j. Romero (Eds.)]. IPCC, Geneva, Switzerland." Edited by Paola Arias, Mercedes Bustamante, Ismail Elgizouli, Gregory Flato, Mark Howden, Carlos Méndez-Vallejo, Joy Jacqueline Pereira, et al., 1–34. https://doi.org/10.59327/ipcc/ar6-9789291691647.001.

Carey, Tony E., Regina P. Branton, and Valerie Martinez-Ebers. 2014. "The Influence of Social Protests on Issue Salience Among Latinos." *Political Research Quarterly* 67 (3): 615–27. https://doi.org/10.1177/1065912914534074.

Chen, Yanqing, and Steven Skiena. 2014. "ACL 2014." In, edited by Kristina Toutanova and Hua Wu, 383389. Baltimore, Maryland: Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-2063.

Comune di Perugia. n.d. "REALIZZAZIONE DELLA LINEA BUS RAPID TRANSIT (BRT) - INFRASTRUTTURAZIONE CORRIDOIO PIEVAIOLA - SETTEVALLI (LINEA CASTEL DEL PIANO - FONTIVEGGE)." https://pnrr.comune.perugia.it/progetto/realizzazione-della-linea-bus-rapid-transit-brt-infrastrutturazione-corridoio-pievaiola-settevalli-linea-castel-del-piano-fontivegge/.

Dunlap, Riley E., and Robert J. Brulle. 2015. *Climate Change and Society: Sociological Perspectives*. Oxford University Press, USA.

Feltrin, Lorenzo, Alice Mah, and David Brown. 2022. "Noxious Deindustrialization: Experiences of Precarity and Pollution in Scotland's Petrochemical Capital." *Environment and Planning C: Politics and Space* 40 (4): 950–69. https://doi.org/10.1177/23996544211056328.

Feltrin, Lorenzo, and Devi Sacchetto. 2021. "The Work-Technology Nexus and Working-Class Environmentalism: Workerism Versus Capitalist Noxiousness in Italy's Long 1968." *Theory and Society* 50 (5): 815–35. https://doi.org/10.1007/s11186-021-09441-5.

Hamilton, Lawrence C., and Barry D. Keim. 2009. "Regional Variation in Perceptions about Climate Change." *International Journal of Climatology* 29 (15): 2348–52. https://doi.org/10.1002/joc.1930.

Hamilton, Lawrence C., David E. Rohall, Benjamin C. Brown, Gregg F. Hayward, and Barry D. Keim. 2003. "Warming Winters and New Hampshire's Lost Ski Areas: An Integrated Case Study." *International Journal of Sociology and Social Policy* 23 (10): 52–73. https://doi.org/10.1108/01443330310790309.

Hamilton, Lawrence C., and Mary D. Stampone. 2013. "Blowin' in the Wind: Short-Term Weather and Belief in Anthropogenic Climate Change." *Weather, Climate, and Society* 5 (2): 112–19. https://doi.org/10.1175/WCAS-D-12-00048.1.

ISTAT. 2025. "Popolazione Residente Al 1° Gennaio : Tutti i Comuni." http://dati.istat.it/Index.aspx?QueryId=19101.

Jackson, Michelle, and David B. Grusky. 2018. "A Post-Liberal Theory of Stratification." *The British Journal of Sociology* 69 (4): 1096–1133. https://doi.org/10.1111/1468-4446.12505.

Jurafsky, Daniel, and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rd ed. https://web.stanford.edu/~jurafsky/slp3/.

Legambiente. 2019. "Mal'aria 2019, il rapporto annuale annuale sull'inquinamento atmosferico nelle città italiane • Legambiente." https://www.legambiente.it/

comunicati-stampa/malaria-2019-il-rapporto-annuale-annuale-sullinquinamento-atmosferico-nelle-citta-italiane/.

Mattioli, Lorenzo. 2025. "UmbriaPress Corpus – Lorenzo Mattioli." https://lormatt.github.io/posts/UmbriaPress%20corpus/.

McCombs, Maxwell E. 1976. "Agenda-Setting Research; A Bibliographic Essay." https://eric.ed.gov/?id=ED132575.

Meneghello, Matteo. 2012. "L'acciaieria di Terni va sul mercato." http://www.ilsole24ore.com/art/impresa-e-territori/2012-10-10/lacciaieria-terni-mercato-064320.shtml?uuid=AbASpArG.

Merton, Robert K. 1987. "Three Fragments From a Sociologist's Notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials." *Annual Review of Sociology* 13 (Volume 13, 1987): 1–29. https://doi.org/10.1146/annurev.so.13.080187.000245.

OpenAI. 2022. "Introducing Whisper." https://openai.com/index/whisper/.

Pedersen, Thomas Lin. 2024a. "Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks." https://CRAN.R-project.org/package=ggraph.

———. 2024b. "Tidygraph: A Tidy API for Graph Manipulation." https://CRAN.R-project.org/package=tidygraph.

Portelli, Alessandro. 2017. "La città dell'acciaio. Due secoli di storia operaia." https://www.donzelli.it/libro/9788868436025.

———. 2023. *Dal rosso al nero*. Donzelli editore.

Pörtner, H.-O., A. Alegría, V. Möller, E. S. Poloczanska, K. Mintenbeck, S. Götze, and (eds.). 2023. "Global to Regional Atlas." In, 2811–96. Cambridge University Press.

Rodríguez-Pose, Andrés, and Federico Bartalucci. 2023. "The Green Transition and

Its Potential Territorial Discontents." *Cambridge Journal of Regions, Economy and Society*, 5. https://doi.org/10.1093/cjres/rsad039.

Silge, Julia, and David Robinson. 2016. "Tidytext: Text Mining and Analysis Using Tidy Data Principles in r" 1. https://doi.org/10.21105/joss.00037.

———. 2017. *Welcome to Text Mining with R | Text Mining with R*. https://www.tidytextmining.com/.

Smith, V. Kerry, and Jared C. Carbone. 2007. "Should Benefit–Cost Analyses Take Account of General Equilibrium Effects?" In, 23:247–72. Bingley: Emerald (MCB UP). https://doi.org/10.1016/S0193-5895(07)23011-8.

Terni in Rete. 2022. "Terni: Arvedi ha acquisito l'Ast, nasce così un Gruppo da 7,5 miliardi di euro." https://terninrete.it/notizie-di-terni-terni-arvedi-ha-acquisito-l%E2%80%99ast-nasce-cosi-un-gruppo-da-75-miliardi-di-euro/.

TerniOggi. 2013. "Cessione Ast, vertice al ministero, Mise: "Valutare piano industriale per scegliere acquirente"." https://ternioggi.it/cessione-ast-vertice-al-ministero-mise-valutare-piano-industriale-per-scegliere-acquirente.

TerniToday. 2020. "Sii Con l'acqua Alla Gola, No Inceneritori: Sciogliere La Società e Costituirne Una Nuova e Totalmente Pubblica." https://www.ternitoday.it/politica/crisi-sii-terni-proposta-comitato-no-inceneritori-sciogliere-societa.html.

The Editors of Encyclopaedia Britannica. 2020. "University of Perugia." https://www.britannica.com/topic/University-of-Perugia.

Toni, Fabio. 2024. "Prisciano e Terni Est tornano alla carica su polveri e ambiente: «Il problema è di tutti»." https://www.umbriaon.it/prisciano-e-terni-est-tornano-alla-carica-su-polveri-e-ambiente-il-problema-e-di-tutti/.

Tuttoggi. 2020. "Terni, Chiude l'inceneritore Di Maratta | 'Esulta' La Città." https:

//tuttoggi.info/terni-chiude-linceneritore-di-maratta-esulta-la-citta/559558/.

Ugolini, Vanna. 2014. "Lucia Morselli, Una Tagliatrice Di Teste Consulente Alle Acciaierie Di Terni." https://www.ilmessaggero.it/umbria/terni_ast_morselli_acciaierie_thyssenkrupp-353296.html.

UmbriaJournal. 2014. "Ast, l'ad Morselli bloccata in fabbrica fino all'alba." https://www.umbriajournal.com/economia/ast-lad-morselli-bloccata-in-fabbrica-allalba-120835/.

Vaquero Piñeiro, Manuel. 2020. "Prima Della Grande Fabbrica: Manifatture e Opifici a Terni (Secoli XVI-XIX)." *STORIA URBANA*, no. 165 (October): 47–62. https://doi.org/10.3280/SU2020-165004.

Vona, Francesco. 2019. "Job Losses and Political Acceptability of Climate Policies: Why the 'Job-Killing' Argument Is so Persistent and How to Overturn It." *Climate Policy* 19 (4): 524–32. https://doi.org/10.1080/14693062.2018.1532871.

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (September): 1–23. https://doi.org/10.18637/jss.v059.i10.

———. 2024. "Rvest: Easily Harvest (Scrape) Web Pages." https://CRAN.R-project.org/package=rvest.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the {Tidyverse}" 4: 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Second edition. Beijing ; Sebastopol, CA: O'Reilly.

Zona, Amerigo, Ivano Iavarone, Carlotta Buzzoni, Susanna Conti, Michele Santoro, Lucia Fazzo, Roberto Pasetto, et al. 2019. *SENTIERI - Studio*

epidemiologico nazionale dei territori e degli insediamenti esposti a rischio da inquinamento. Quinto Rapporto. https://epiprev.it/pubblicazioni/sentieri-studio-epidemiologico-nazionale-dei-territori-e-degli-insediamenti-esposti-a-rischio-da-inquinamento-quinto-rapporto.

Appendix

A. Pre-processing

Tokenisation

Depending on the task at hand, different preprocessing methods were used. In all use cases, the corpora needed to be split into simpler linguistic units, or *tokens*. This process, known as *tokenisation*, can be carried through with a plethora of different techniques. The predominant approach in this study is *rule-based*. This method entails identifying the set of grammatical rules defining the ways to signal the end of a word and the beginning of another. Once this first stage of desk research is done, the only problem left to tackle is that of translating the rules in machine-readable form, so that a computer can divide long-form text into words (or sets of words). In many western languages, for instance, blank space divides one word from the next, and sentences are divided by punctuation (e.g. full stops, question marks, and so forth). In this specific case, a tokeniser designed for the English language was manually adjusted to Italian punctuation, using the `tidytext` package (Silge and Robinson 2016). More advanced, machine-learning techniques were also tested. Taking advantage of `spacyr`, a wrapper around the Python package `spacy`, a pre-trained tagging and tokenising model was also implemented. At the price of immensely higher

computational demands, this method promises much richer results, complete with PoS (Part of Speech) tagging and even entity recognition. Even at a first glance, however, the results appeared unreliable in more complicated tasks, and indistinguishable from simpler approaches in simpler ones. This approach was therefore abandoned, in favour of `tidytext`'s less demanding methods.

The last aspect of tokenisation to be taken into account is defining what a token is. In this study, tokens will be defined as either individual words or *bigrams* (two consecutive words). The reason for including bigrams in the analysis is that a relational database of co-occurring words might be more informative, especially in the case of dictionary-based analyses.

Deleting unnecessary content

After tokenisation, the corpora were stripped of all unnecessary content. This includes the more or less obvious punctuation and numbers, but also all words which have no real value to the analysis. There are many words which we use in passing, but add no semantic *value* to the text. All words which do not contribute in defining a document's meaning and topic are called *stopwords*. The corpora were all matched with a dictionary of Italian stopwords, using the `stopwords` package (Benoit, Muhr, and Watanabe 2021).

Stemming and lemmatisation

Especially when using dictionary-based text mining methods, stemming is a common pre-processing step. It entails turning words into *stems*, as in only keeping their roots

instead of their conjugated or declined form. An obvious case in which this would be useful is verbs: the Italian translation for "to eat", *mangiare*, can be found in the conjugated forms *mangio, mangi, mangiarono, mangerà,* and so on, which can all be reduced to the stem *mang-*. There are, however, very common cases in which a verb's conjugated form disregards the infinitive form's root. This is the case, for instance, of the verb *andare* ("to go"), whose conjugated forms include *vado, vai, vanno*. A solution to this is adopting a more sophisticated approach called *lemmatisation*. It consists of automatically recognising a word's *dictionary form*, and assigning it to each token. Once again, this task is hardly a simple one for a computer to take on, so it tends to be rather demanding in computational terms. An attempt was made at the early stages of this study to adopt lemmatisation, with very unsatisfactory results. The method of choice was thus that of taking on this problem at a later stage, using tokens as they are and then adjusting analytical dictionaries to accommodate this issue.

B. Text retrieval

YouTube querying

The Python script used to retrieve and transcribe YouTube videos took advantage of the native querying system, cycling through lists of keywords. Below is a table of all queries associated with each city.

Query category	PG	TR
Waste management	rifiuti, smaltimento, raccolta differenziata	inceneritore, rifiuti, smaltimento
Environment, nature	ambiente, ecosistema, qualità dell'aria, qualità dell'acqua, inquinamento, spazi verdi, verde urbano, emissioni	ambiente, ecosistema, qualità dell'aria, qualità dell'acqua, inquinamento, spazi verdi, verde urbano
Industry	industria, nocività	acciaieria, AST, industria
Transportation	ciclabile, mobilità sostenibile, BRT, autobus	ciclabile, mobilità sostenibile, BRT, autobus

Keyword-based retrieval in Umbria Press

News articles were queried through keyword-based matching. The keyword are listed below.

Query category	Keywords
Industry	acciaieria, industria, acciaio, Arvedi, Thyssen, Thyssen-Krupp
Transportation	treno, aeroporto, Trenitalia, ciclabile, mobilità, BRT, trasporti
Environment	emissioni, PM10, inquinamento, ecolog, riuso, ecosistem, rifiuti, inceneritor

Contextual sentiment querying

The “environment” category was further decomposed and expanded in order to study the contextual sentiment of significant terms. All terms are listed below.

Query category	Keywords
Emissions	emissioni, PM10
Pollution	inquinamento,
Ecology	ecolog, ecosistem
Waste management	riuso, rifiuti, inceneritor

Acknowledgements

My heartfelt thanks go first and foremost to my family. Thanks for listening, for caring, and for making sure I didn't completely lose it during these five years at university.

Many thanks to my supervisors, too, for their patience while teaching a newbie how to actually write something.

To all my friends in Terni—thanks for keeping me grounded and cheering me on, even from far away. Growing up together has been such a privilege and a gift.

To my friends in Bologna, thank you for sharing your experiences and thoughts with me. I would not be half the person I am today if it wasn't for the conversations we had between classes.

To my friends in Ippogogo, thanks for conjuring a theatre out of thin air and allowing me to take part in the lovely task of telling stories.

To my friends in Ferrara, thanks for welcoming a stranger so late in the game. Our chats made me rethink the very way in which I understand the space around me, and almost made the weather bearable.

Finally, to all my friends around the world, whether you're at home or off on some big adventure, thanks for sharing a bit of yourselves with me.