International Journal of Heritage Studies

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjhs20

Anthracite memories: semantic tagging and coal mining oral histories

Madeline Brown & Paul Shackel

To cite this article: Madeline Brown & Paul Shackel (2023): Anthracite memories: semantic tagging and coal mining oral histories, International Journal of Heritage Studies, DOI: 10.1080/13527258.2023.2243461

To link to this article: https://doi.org/10.1080/13527258.2023.2243461

Published online: 31 Aug 2023.

INTERNATIONAL JOURNAL OF HERITAGE STUDIES
https://doi.org/10.1080/13527258.2023.2243461

Anthracite memories: semantic tagging and coal mining oral histories

Madeline Brown and Paul Shackel

Department of Anthropology, University of Maryland, College Park, Maryland, USA

ABSTRACT

Oral histories are a critical source of information about lived experiences of past events. They have been analysed both for their form – linguistically as texts, performances, and expressive accounts – and their content for understanding historic events and personal experiences. Here we focus on sentiment analysis approaches frequently applied to big data research questions, but less often utilised by anthropologists working with oral histories. Oral histories collected half a century ago in the anthracite mining communities of northeastern Pennsylvania are examined by considering methodological and historical questions. This project explores how oral history and data science might be productively combined to understand these now historic communities' everyday lives and working conditions. Bakhtin's (1981) concept of chronotope helps us understand the memory of these anthracite coal mining communities' daily life and working conditions.

ARTICLE HISTORY
Received 15 March 2023
Accepted 29 July 2023

KEYWORDS
Text mining; oral history; chronotope; mining communities; labour history

Introduction

Oral histories are important anthropological sources for understanding particular historical contexts or chronotopes. These recorded narratives offer insights into the direct experiences of a particular place and time and how the narrator perceives and describes these experiences. Though oral histories are often analysed qualitatively, incorporating text mining and natural language processing methodologies into oral history analysis offers the potential for new understandings. Treating 'text as data' enables efficient processing of lengthy texts to identify patterns and themes (Grimmer, Roberts, and Stewart 2022). For example, sentiment analysis involves tagging terms based on their association with broader emotions, categories, or sentiments (Silge and Robinson 2017). However, McGillivray (2021) notes that because words can mean multiple different things across contexts and this meaning can change over time (lexical polysemy and semantic change) computational text analysis tools do not always accurately apply to historical texts; they further outline future directions to address this issue.

Here, we analyse oral histories using Bakhtin's concept of the chronotope (Bakhtin 1981). We combine close and distant reading approaches (McGillivray 2021) to these texts to highlight how community members remember these places. In particular, we compare sentiment tagging results from both general purpose and domain-specific lexicons. Finally, we discuss the potential for further integration of sentiment analysis for oral histories and other anthropological texts.

Our analysis focuses on the oral histories from the anthracite coal mining communities in northeastern Pennsylvania. While anthracite coal was discovered in the region during the late

CONTACT Madeline Brown mtbrown@umd.edu Department of Anthropology, University of Maryland, College Park, Maryland, USA

1760s, large-scale extraction of this carbon fossil fuel was only realised by the 1840s. The development of railroads and canal systems allowed the shipment of large quantities of coal to the east coast for use in industries and domestic heating. Anthracite coal ignited the burgeoning industrial revolution, which enabled U.S. industries to become an international leader in manufacturing by the middle of the nineteenth century (Palladino 2006; Wallace 1987).

The growing number of coal patch towns, which developed with the increase in coal extraction, attracted a new immigrant workforce and their families. The first coal miners to the anthracite region came from England, Wales, and Germany. By the 1840s and 1850s, many new workers migrated from Ireland, escaping famine and political oppression. By the 1880s, many coal operators began to recruit workers from eastern and southern Europe. These new immigrants were easily enticed to the coalfields because of their home countries' deteriorating economic and political conditions. These new immigrants to northeastern Pennsylvania faced poor living and working conditions, and underemployment (Dublin 1998; Greene 1968; Miller and Sharpless 1998, 170–173; Roberts 1970, [1904]; Roller 2015, 2018; Shackel 2023).

The coal baron's poor treatment of their new workforce was legitimised by the new immigrant's racialisation. The workers were placed in jobs based on their ethnic background, so the new immigrants were underemployed, were provided lower wages than their peers, and lived in substandard housing. They tended to be in perpetual debt, a form of debt peonage, and had a limited supply and limited diversity of foodstuffs. Communities also faced environmental stresses, with the air filled with coal dust and the drinking of toxic water drained from the coal mines, leading to the early onset of several chronic diseases. The community faced additional stress with the decline of the coal industry, the region's major economic driver. This loss caused emotional trauma as well as nutritional stress. Racial hierarchies that were naturalised in the late nineteenth century legitimised the foundation of what became a long-term history of structural violence and the legacy of poverty and health disparities (Dublin and Licht 2005; Greene 1968; Shackel 2019, 2023; Turner 1977, 1984).

The anthracite coal industry peaked around WWI, employing about 180,000 workers. While anthracite mining had one the highest mortality rates of any profession in the United States, the industry continually recruited labour from overseas, offsetting the high death rate and ensuring low wages and high profits (Smith 1886; 1887; Wallace 1987). Daily, coal workers endured the mental stress associated with underground mining, thinking they might not survive until the end of the day. For instance, after the 1922 strike, noted economist H.S. Raushenbush (1924, 37) observed the coal miners' general mood in the early twentieth century;

> They said men accustomed to facing death in routine darkness did not whine when there was no butter for their bread and when the coffee grounds were boiled to their fourth weakness. They said that the women they married, and the women who were the mothers of such men, had that quality it takes to carry on. They had handled life without gloves and without perfume. They had seen limp bodies carried from the shaft to their own street, to be taken into some quiet little house. They never knew when it would be their own house. They knew their life was like that, was always to be like that. They did not ask much. They knew all these things. But they were not sure that others knew. (Raushenbush 1924, 37)

The coal industry began to decline after WWI as other fossil fuels became readily available to households and industries. The Johnson Act of 1921 and later the Johnson-Reed Act of 1924 restricted the number of new immigrants. The new law based the number of immigrants on a proportion of the country of origin of new immigrants found in the 1890 US federal census. In 1890 the numbers were much smaller for those from southern and eastern European countries compared to the early twentieth century. Without the influx of new immigrants, those from eastern and southern Europe remained the primary workforce in this industry. Coal extraction decreased precipitously after World War II. Today, only a few hundred people work in the northeastern Pennsylvania anthracite coal industry (Dublin and Licht 2005; Shackel 2023; Turner 1977; 1984 .

Historian; John Bodner (1983, 11) writes, 'No other American industry inflicted more heedless destruction on men and the environment than anthracite mining'.What remains now in the anthracite region is torn earth with mine waste, culm banks, and deteriorating machinery and

buildings – all reminders of the former prowess of the region's industrial past. The mined areas lay in ruins, and acid mine drainage continues to impact the region. Hundreds of miles of streams and rivers are considered dead because of the high acid content of the waters. Vegetation only slowly colonises the acid rich deposits on the landscape with lichens, wiry clumps of grass, goldenrod, and briar bushes. Several decades later birch and locust trees will root in these soils (Marsh 1987, 347; Zawacki 2017).

With the decline and forecasted disappearance of the anthracite coal industry, historians and folklorists began collecting oral histories of the men and women who worked in and around the vanishing anthracite coal industry in the 1970s. The oral histories for this analysis came from the northern section of the anthracite coal region collected by the Pennsylvania Historical and Museum Commission in 1973 from the Scranton, Pennsylvania area and called the Scranton Oral History Project. This collection contains 26 interviews documenting work and daily life from about the 1910s through the next half-century. These oral histories were transcribed in the 1980s and are on file at the Pennsylvania State Archives in Harrisburg, Pennsylvania. Here, we use text mining to examine the sentiments embedded in these oral histories. Additionally, we extend beyond existing sentiment tagging lexicons to consider how anthropologists can incorporate text mining approaches to historical oral histories and other anthropological texts.

Chronotopes of the anthracite

While examining these 'historic' oral histories, the concept of chronotope is a useful framework for understanding how these anthracite coal mining communities remembered their everyday lives and working conditions. The chronotope is how configurations of time and space are represented in language and discourse. The term, translated as 'time-space', was first used by Mikhail Bakhtin (1981) in 1937, when he described it as a central element in his theory of meaning in language and literature. He explains how literary genres operate with different configurations of time and space, giving each genre its particular narrative character. 'Discourses of place, as chronotopic representations, provide important reference points that communities use to make sense of their lives and the spaces in which they live' (Britt 2018, 255). Keith Basso (1984) uses the concept of chronotopes to discuss Western [Apache] stories linked with places. Specific landscapes and geographic features reminded tribal members of the events that occurred there in important moral narratives. They are places on the landscape where time and space intersect and fuse. "Time takes on flesh and becomes visible for human contemplation; likewise, space becomes charged and responsive to the movements of time and history and the enduring character of a people. . . . Chronotopes thus stand as monuments to the community itself, as symbols of it, as forces operating to shape its members' images of themselves" (Bakhtin quoted in Basso 1984, 44–45).

As Peeren (2006, 69) points out, 'subjects do not stand above a chronotope as its masters, but are within it or, indeed, of it', indicating that chronotopes provide a frame that may enact limits on the kinds of identities that individuals may occupy. This suggests that individuals may wish to challenge dominant representations of place and time to open up new possibilities for their identities. For instance, Britt (2018) describes how residents of Flint, Michigan responded to media and outsider's perceptions of their community using oral histories to counteract externally imposed 'chronotopes' or the 'official picture, or collective understanding of space and time' that they were supposed to be a part of (Britt 2018, 254). By serving as the collective memory of space and time, chronotopes, by implication, directly impact the identities of subjects that are subsumed under them. Britt (2018) explains that since the 1980s, the media has depicted Flint as a dangerous and miserable place. The perceptions of Flint serve as chronotopic representations that locate the city and its people in space and time. The same can be true for the anthracite region in northeastern Pennsylvania. Today, the anthracite region is a place of poverty, sickness, and addiction, and these devastating conditions are prime for a critique of unchecked capitalism. For some, the region is now a destination for dark

heritage tourism. For others, the place is a destination to remember and commemorate the struggles of a familiar past (Shackel 2018, Shackel 2023).

When coal companies abandoned the region, they left behind a ruined landscape. About a quarter of the anthracite region's 484 square miles have been directly affected by almost two centuries of mining, and much more have been indirectly affected. The coal mining industry's collapse has led to high unemployment rates and a significant out-migration of its traditional population. Scattered rubble and waste litter patches of the landscape, and people are abandoning the region at an alarming rate. There are black mounds and hills of coal waste, some reaching 125 feet high. Water running through abandoned mines interacts with the coal and other surrounding materials and discharges into streams and rivers, polluting numerous waterways inside and outside the region (Shackel 2017).

In 1960, during this abandonment, Penn State geographers Deasy and Griess (1961, 3) came to the area and bleakly assessed the landscape as a 'man-made Bad Lands'. People have been making pilgrimages to gaze at and roam around the abandoned industrial landscape to reinforce their memories of the industrial era. Material culture may have some ascribed meanings – past and present – associated with it. These meanings vary among individuals and interest groups. 'This is a landscape that has gone through two periods of asymmetry as a place – first an early industrialising stage, when the landscape provided a means (i.e. jobs) but relatively little specific meaning to the new inhabitants, and now a declining stage, when the landscape is rich with meaning from that past but deficient in means' (Marsh 1987, 339).

In 1969, Paramount Pictures filmed *The Molly Maguires* in one of the coal patch towns, an event that was influential in creating the region's chronotope. The Mollies became active in the anthracite region around 1870, confronting coal operators for their unjust labour practices through subversive activities. The Pinkerton National Detective Agency infiltrated the Molly Maguires, which led to the conviction of 20 men for murder, arson, kidnapping, and other crimes. They were executed by hanging in 1877 and 1878. Memorializing the struggles of the working class and this 'heroic defeat' in Mahanoy City, a memorial was erected in 2010 to commemorate the efforts of the Molly Maguires. The memorial is a blindfolded man, hands tied behind his back, about to be hanged. It is located on the town's main street, although it sits behind a block wall, making it difficult to view from the road (Hand 2015; Shackel 2020). (Figure 1)

Other memorials scattered throughout the region create and reinforce the region's chronotope. For example, many larger towns have memorials for coal workers near the town's centre. Some are a sole bronze coal miner with a pick axe in hand and standing on a pedestal. Memorial markers throughout the region commemorate mining disasters and the loss of life, like the 1869 Avondale mine disaster and the 1959 Knox mine disaster. There is also a monument recognising the 1897 labour massacre at Lattimer, where a sheriff and his deputies killed 25 unarmed striking miners protesting for fair wages and better working conditions (Shackel 2018, 2019).

The state-sanctioned museums at Eckley Miners' Village, near Weatherly, Pennsylvania, and the Anthracite Heritage Museum in Scranton, Pennsylvania, interpret the life and working conditions people once endured under the rule of the coal barons (Shackel 2019). (Figure 2) Adjacent to the Anthracite Heritage Museum is the Lackawanna Coal Mine Tour. It is one of three anthracite mine tours. (The other two are in Lansford and Ashland). These mine tours take you several hundred feet below the surface, and guides describe the harsh manual labour and dangerous conditions in which the miners worked. As James Abrams (1994, 27) explains, these are places and landscapes that have been abandoned by capital. They are spaces for remembering in which industrial artefacts and places become distinctive places to otherwise displaced social environments.

Driving through the anthracite region, it is stunning to notice the decay and destruction in many coal mining towns. For instance, Centralia, Pennsylvania, is one town impacted by a dangerous underground mine fire. Underground mine fires have a long history in the anthracite region (Conlogue 2013, 57). Now mostly abandoned, Centralia draws tens of thousands of people yearly to observe a town's destruction and slow decay. The landscape has

Figure 1. (Molly Maguire statue). The Molly Maguire memorial located in the center of Mahanoy City. The "Mollies" were an underground organization of coal workers who fought to undermine the exploitative treatment of the coal operators. Many of them were caught and hanged in 1877 and 1878. (Photograph by Author)

abandoned roads and empty house lots. A few stop signs still stand at several street intersections, although a bit awry. Steven and Lewis (2007) write that capitalism makes destruction seem inevitable, even progressive. Capitalism controls the landscape and moves workers and their communities to the margins. A ruined landscape is about job loss, health disparities, and environmental degradation – in the past and present (High 2013a, 2013b; Mah 2009, 2010, 2010, 2012). Centralia, Pennsylvania, is an example of a once-thriving town that lost its economic engine and is now abandoned. Its narrative in time and space is forever grounded in anthracite history.

Underground mining subsidence has impacted the anthracite region for over a century and has significantly impacted the landscape with tragic outcomes. The most noted and well-commemorated subsidence is the Stockton Mine disaster. The Stockton Mine subsidence occurred on 18 December 1869, claiming 10 lives. The coal company mined beneath a residential area in Stockton when two houses collapsed into the ground, falling 40 feet and killing everyone in the houses. A third home went into the subsidence without any fatalities (*New York Times*, 19 December 1869a, 20 December 1869b). Subsidence events still occur throughout the region. There are newspaper accounts of sinkholes appearing on roads or in a household's backyard. Sometimes there are stories about a house slowly moving and tilting because of a foundation giving way to subsidence (Gallo 2017).

Several published (in print and on the web) surveys regarding the region's general health and well-being indicate that some consider this area as the unhappiest place in the USA, a product of the region's declining employment and economic outlook, as well as its poor

Figure 2. (Eckley Miners' Village). Eckley Miners' Village is the location of Paramount Pictures The Molly Maguires. The Pennsylvania Historical and Museum Commission now owns the town and is used to interpret anthracite mining life. (Photograph by Author)

general health, among other factors. Economists compared 367 MSAs (Metropolitan Statistical Areas) in the USA in a study that considered overall happiness. Northeastern Pennsylvania (Scranton, Wilkes-Barre, Hazleton MSA) is ranked last (Glaeser, Gottlieb, and Ziv 2014). The region continues to rank relatively low in recent health and well-being studies.

The Center for Disease Control and Prevention documents a high rate of deaths related to heart disease in the American South among the white population (Steckel and Senney 2015). This high incidence rate reaches into the anthracite coal region. People living near open-air mining (strip mining and mountaintop removal) who rely on well systems can be drinking water drained through mines (Acid Mine Drainage – AMD) and consume an abnormally high rate of metals being washed out of the coal beds. These communities have cancer rates of 14.4% compared to 9.4% for people elsewhere in the coal region. In addition, the rate of children born with congenital disabilities is 42% higher in mountaintop removal mining areas (Ahern et al. 2011; Hendryx et al. 2012; Hitt Nathaniel 2010). While we do not have comparable historical data for the general health conditions of coal communities a century ago, it is easy to assume that many chronic diseases would have existed and were probably worse.

For Bakhtin (1981, 84), chronotope refers to how time and space are represented in literature. For many, the anthracite chronotope has developed in an environment of a ruined landscape, and the working class, hard luck narrative is reinforced through memorials, state-sanctioned museums, and the media. While significantly diminished, the effects of the mining industry continue to impact the community's general poor health and well-being. While coal is no longer the centre of the region's economy, there are many reminders of the heedless destruction of the landscape and the environment. While the area is now attracting new immigrants to work in low-skilled occupations in fulfilment centres, the region continues to embrace its coal mining history. It is the backbone of the region's heritage and what the outside world recognises as the region's identity. The following

oral histories show how people remembered time and space in the anthracite coal region in the 1970s. This remembering helped to create a consensus narrative, a collective memory of the region's past that emphasises strategies of success and the importance of collective actions.

Sentiment analysis and oral history

Natural language processing and text mining approaches have transformed how social scientists and humanities scholars approach the analysis of texts (Grimmer, Roberts, and Stewart 2022; McGillivray 2021). As Grimmer and colleagues argue, these computational methods make it possible to rapidly analyse large volumes of text in a way that ' . . . *augments* our reading ability', which thereby 'help[s] us read differently, not avoid reading at all' (Grimmer, Roberts, and Stewart 2022, 24).

For our study we recognise High's (2017, 110–129) challenge to us to think differently as we re-use evidence in oral history projects. In his case study he explains that oral history research can go beyond the 'juicy quotes syndrome' and supply contextual detail in order to develop an understanding of contextual circumstances in lived experiences. He explores how particular voices are recorded, shared, accessed, and reused, particularly using new technologies.

Key to our analysis is Moore's (2006, 21–32) call for understanding the importance of context and reflexivity because they are intrinsic to the process of qualitative research. The context of the original research has been privileged and is often reduced to the reflexive production of the data by the researcher. We can shift attention from context as something as fixed to the processes of the identification and construction of context. In this way we can understand reusing qualitative data as being about the process of re-contextualising data. This process can then open up more possibilities of meaning-making from reusing data.

For anthropologists, qualitative text analysis and semantic coding for themes in ethnographic field notes and interviews are widely practiced methodologies (Bernard 2011). As anthropologists turn to examining high volume data sets, including social media, news archives, oral histories, and folklore, structured tools for text analysis become critical. Here, we first describe the state of text mining community oral histories. Next, we compare tools for general purpose and domain-specific sentiment taggers for analysing coal mining oral histories from the 20th century. Finally, we discuss challenges and potential future pathways for bringing text mining tools into analyses of oral histories and historical texts.

Texts are frequently annotated or coded as part of the qualitative data analysis workflow (Bernard 2011). Text annotation can also help standardise forms of speech and link historical terms across sources, time periods, and languages. One example of this approach is exemplified by the Classics Department at the University of Cambridge. They describe their xml publishing process for developing the Cambridge Greek Lexicon, through which words are tagged according to themes, language roots, or other custom tagged characteristics for easy sorting, analysis, and comparisons (Classics Department n.d.). Another approach is exemplified in the Oral History Metadata Synchronizer (OHMS), focused on tagging and indexing metadata for oral histories (Boyd, Fernheimer, and Dixon 2015), while Rieping (2022) combined manual topic tagging with natural language processing methodologies.

When working with oral histories and historical datasets, researchers may seek to parse and tag existing texts. Manual text tagging can often be time-consuming and may not be easily standardised (McGillivray 2021). Instead, using natural language processing (NLP) tools enables more efficient general analysis of sentiments, parts-of-speech (POS), word roots (e.g. lemmatisation and stemmers [Penn Treebank-2 corpus]), among other aspects of text data (Coden et al. 2005; Silge and Robinson 2017). For example, the Natural Language Toolkit (NLTK) compiles multiple open source tools for conducting text analyses that may be of interest to anthropologists (NLTK Project 2022), while the Classical Language Toolkit (CLTK) similarly compiles tools for analysing classical Eurasian texts (Johnson 2021).

McGillivray (2021) notes the important distinction between *text* searches and *semantic* searches in documents. The former can return specific strings of characters, while the latter enables researchers to find all words associated with a particular concept, which might vary based on the historical and social context (McGillivray 2021). McGillivray illustrates this with the example of the difference between searching a text for the word 'happy' as opposed to searching for a set of terms associated with the emotion 'happiness' (McGillivray 2021, 261). Both approaches will return useful insights, but the former is focused on specific words, while the latter allows for a more nuanced thematic analysis. Semantic tagging is often conducted as part of part-of-speech (POS) or sentiment analysis, wherein terms are tagged according to the type of words (e.g. nouns, verbs, etc.) or the emotions with which they are associated (Silge and Robinson 2017).

Potential exists for further development of disciplinary or issue-specific tagging libraries. Coden et al. (2005) note that general purpose English language natural language processing part-of-speech tagging lexicons do not accurately interpret medical texts, suggesting the need for domain-specific corpora. Moreover, they find that incorporating a small-sized domain-specific lexicon facilitates more accurate results than the general English-language tagging lexicon alone (Coden et al. 2005). These benefits may be even more significant when domain-specific lexicons are made open-source and shared among researchers and practitioners working in the same fields.

Drawing inspiration from the concept of chronotopes, constructing specific tagging dictionaries for texts from particular times and places may add an additional layer of interpretation to these histories. Here we examine oral histories from the anthracite region using general purpose sentiment libraries and domain-specific tagging libraries. The domain-specific tagging library presented here focuses on three sentiments: family, work, and community. Both individual terms and bigrams (pairs of words) are tagged according to these categories.

Methods

Data description

This study analyzes the text of 26 oral histories from Scranton, Pennsylvania, recorded in the 1970s, that detail memories about coal mining, labour, and community in this region. These are members of working class mining communities, who are mainly elderly and lived in the anthracite region their entire lives. We interpret these texts using sentiment analysis conducted with R, including packages for data wrangling, visualisation, and text mining (Hvitfeldt 2022; R Core Team 2022; Rinker 2018; Silge and Robinson 2016; Wickham 2016, 2022; Wickham et al. 2019). The aim of the sentiment analysis used here is to assess the applicability of general-purpose emotion-focused sentiment lexicons to oral histories as well as to develop a domain-specific lexicon focused on key topics appearing in the Scranton Oral Histories.

Data analysis

In order to analyse the text of the Scranton Oral Histories, the documents were first converted into a plain text format and imported into R. Once each oral history was imported, we further cleaned the data with the following steps. First, lines of text associated with metadata were removed from the top of each history. Second, we removed apostrophes and numbers and converted the text to lowercase using the textclean package (Rinker 2018). Next, tokens were unnested into single terms and bigrams, from which generic stopwords were removed using tidytext (Silge and Robinson 2016). Stopwords include common terms such as *am*, *and*, or *of*, which often appear in texts but add little semantic content (Burns 2018; Hvitfeldt 2022). Stopword removal is common in natural language processing and text mining, though it is also noted that any time words are removed from a corpus, there is a risk of also removing important information for interpreting meaning in the text

(Hvitfeldt 2022). Both generic and custom or domain-specific stopword lists continue to be used and developed in text analysis (Sarica and Luo 2021; Burns 2018; Hvitfeldt 2022). After generic stopwords were removed, we further removed terms from a custom stopword list since additional words and abbreviations still added messiness to the data. This follows the trend identified by Sarica and Luo (2021, 2), who note that general stopword lists may 'leav[e] many domain-related uninformative and repetitive terms in the data'. Our custom stopword list included terms such as 'interviewed', 'they'd', 'yeah', and the names of the interviewers and interviewees. The full stopword list is included in the Appendix and article GitHub repository (Brown and Shackel 2023). Finally, we further standardised the words and bigrams in terms of tense, parts of speech, and plurality. For example, 'brother' and 'brothers' are combined into the singular form of the term, while 'mentioned' and 'mention' are combined into the latter term. This data cleaning step decreases the occurrence of treating each version of a word as a distinct term.

Cleaned datasets were analysed using sentiment analysis. We used general purpose sentiment tagging lexicons and developed a custom, domain-specific tagging lexicon. For general sentiment tagging, we used the Bing sentiment lexicon (Minqing and Liu 2004) and the AFINN lexicon (Nielsen 2011). The Bing lexicon tags words as either 'positive' or 'negative', while AFINN assigns a value from −5 to + 5, denoting the degree of positivity or negativity for each term. In addition, we created a custom sentiment lexicon for the three themes of *community*, *family*, and *work*. The most frequently occurring individual terms and bigrams from the oral histories were tagged based on whether or not they were associated with each category. Annotating these terms based on expertise in the field and then using these categories as a framework for further analyses combines the qualitative and quantitative aspects of interpreting texts. In support of open and reproducible science, these tagging libraries and stopword lists are available in the Appendix and online in the project GitHub repository (Brown and Shackel 2023).

Results

How is sentiment encoded across these oral histories?

Oral histories often include detailed information about the interviewee's perception and memory of events, including emotion and sentiment-based values. Sentiment analysis is a common technique in natural language processing, yet it has not often been applied to oral history analysis. Several different libraries exist for analysing sentiments. For example, the NRC library allows text to be analysed by different types of emotions, including 'anger', 'trust', and 'joy', while other libraries, such as Bing, focus on classifying terms into positive or negative sentiments. A challenge in using these libraries – perhaps particularly the NRC – to analyse oral histories is that the meaning or sentiment of a word might vary widely across social and historical contexts. Here, we apply the Bing and AFINN sentiment analysis libraries to the 1970s oral histories to examine the most common terms associated with both positive or negative sentiments.

Given that the general sentiment libraries are not specifically designed for particular historic or cultural contexts, the utility of tagging terms from oral histories with them remains somewhat limited. However, these types of tagging libraries can still yield some insights into the general mood or tone of a particular interview. For this reason, our analysis focuses on binary (positive/negative) taggers, rather than one assigning particular emotions or sentiments to words. Figure 3 conveys the most frequently occurring terms associated with negative and positive sentiments in the Scranton interviews using the Bing sentiment lexicon (Minqing and Liu 2004).

Using a general tagging lexicon means that not all the words will be tagged in ways that make sense for a particular context. This is particularly true for historical texts, as the use and meanings of words may change over time (McGillivray 2021). In this case, some terms may retain positive or negative sentiments, while the sentiment of other terms may differ between the tagging lexicons and oral history texts. For example, the term 'strike' is the most frequently occurring term from the

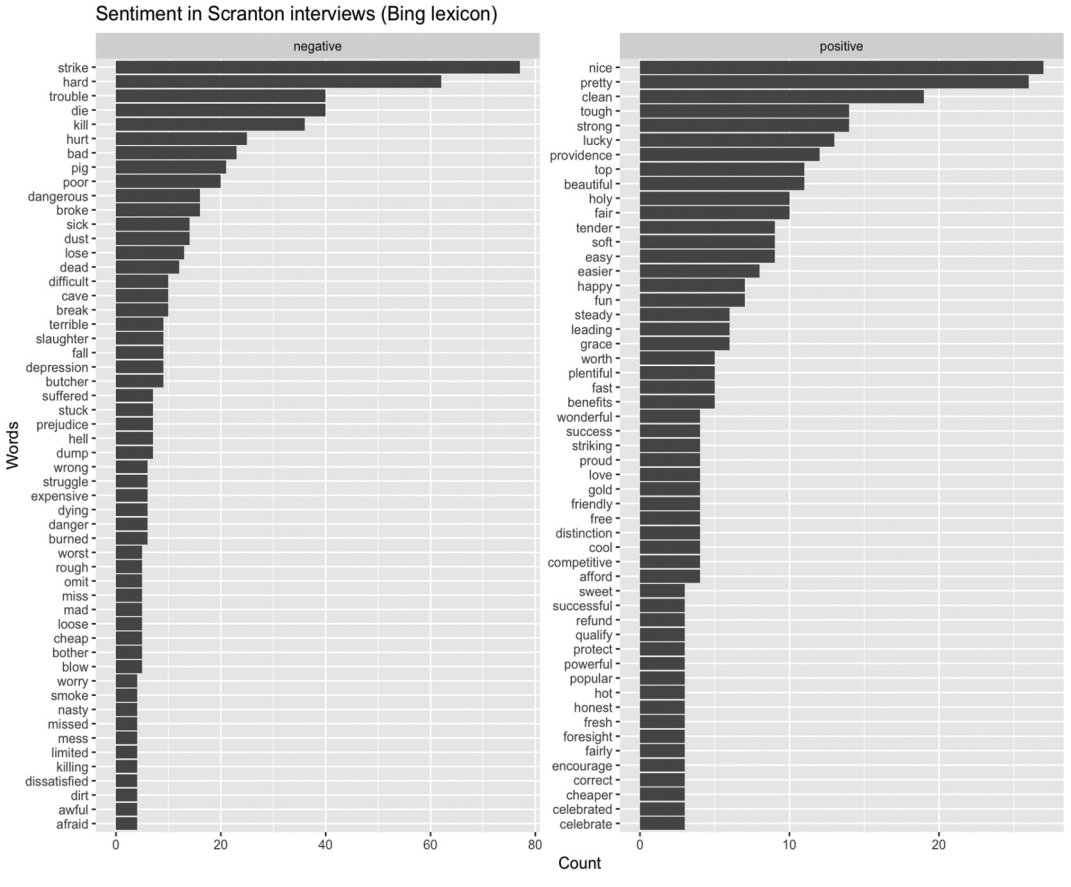

Figure 3. Most frequently occuring words associated with positive and negative sentiments (Bing sentiment dictionary) in Scranton interviews. Note: some words that would also be tagged positively, such as "well" or "work" are removed from the analysis as stopwords.

Scranton Oral Histories that is tagged as 'negative' by the Bing lexicon (Figure 3) This negative tagging may be the result of multiple meanings for the term, or due to the process by which this tagger was constructed. However, in the context of the Scranton interviews, the term 'strike' is perhaps better tagged as positive, rather than negative. The term describes workers' collective action protesting for better working and living conditions. Similarly, three terms related to butchery (pig, butcher, and slaughter) are tagged as negative, when they may not be negative in this context. Rather, these words help to describe the process and consumption of often lacking protein in the diets of members of this community.

The AFINN sentiment tagging assigns terms a value of positivity/negativity, rather than using binary categories. Figure 4 shows that the terms with the highest positivity include 'nice', 'help', 'clean', 'care', and 'united'. In this case study, the word *united* is associated with the national union – the United Miners Workers of America. More of the common words were considered negative than positive under the AFINN tagger.

Identifying themes: family, community, and work

In addition to analysing texts with general purpose sentiment taggers, we also tagged terms and bigrams based on domain-specific categories to examine how these concepts are embedded in the

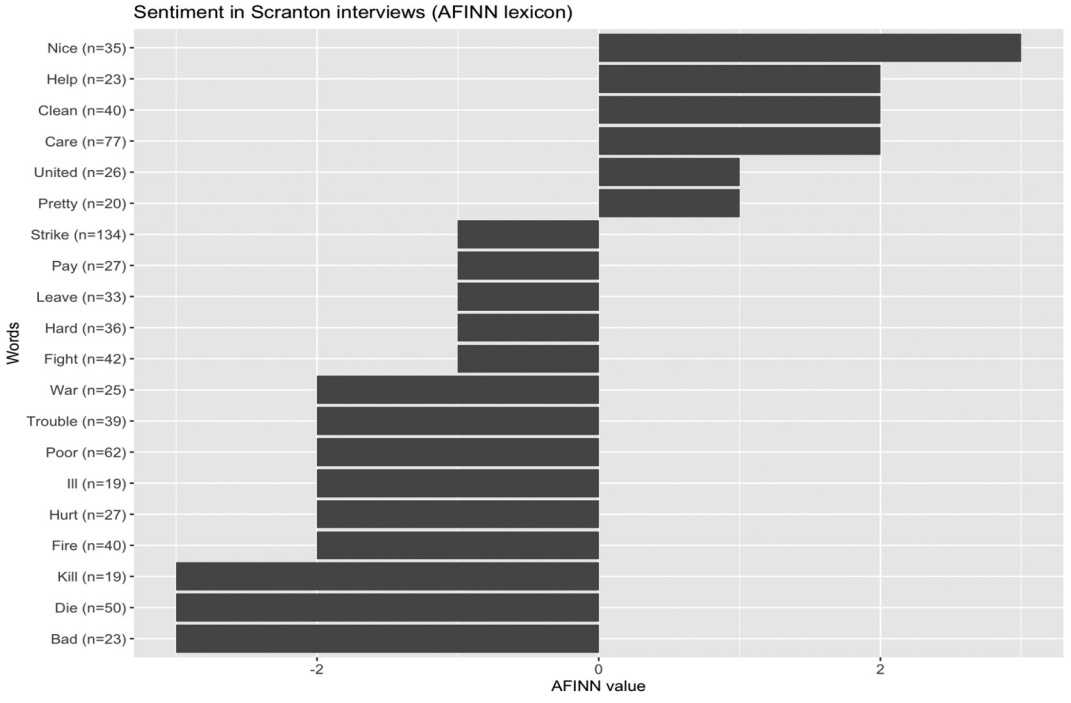

Figure 4. AFINN Sentiment values among top words in the Scranton interviews. The n values refer to how frequently a term appears in the oral histories.

oral histories (Figure 5 and 6). This approach follows methods from sentiment analysis, where words are tagged with particular sentiments or themes, which can then be used to analyse the prevalence of these themes across different texts. Here, we created a unique custom tagging lexicon based on the themes of *work, community*, and *family*, three concepts that are central to understanding the context of historic mining towns. This approach combines qualitative assessments from the researcher with quantitative natural language processing tools.

Under the theme of community, the term 'people' was often used to describe the other. For instance, Luigi Ligl described his community: 'Yes, there was a lot of Polish and there were some Irish but not too many. There was only seven families of Italian *people*. But mostly Polish'.

Under the theme of family, 'mother' is often used to describe the roles and duties of the person who often became the head of the household and the family caretaker after the father's untimely death. For instance, Walter Dutchak (Interview 1973) explained, 'My *mother* not only had to take care of the needs of the family, but she had to take care of the needs of these men (borders). She did a lot of hard work. She got up early, and she had a garden to take care of. . . . My *mother* made most of the clothing. From old flour sacks or sugar sacks. And I know some of my under garments were made from flour sacks and sugar sacks. And my *mother* had a second-hand sewing machine, and she made quite a bit of clothing on that machine'. The term 'family' was used to describe relationships with the closest relations. For example, Stanley Gawrych (Interview 1973) went to New York to earn money for the family. He stayed in New York, and his family remained in the Scranton area. He stayed in a boarding house, 'And I sent the money home to my *family*'.

Under the theme of work, the words 'mines' and 'coal' are common. For instance, Chester Brazina (Interview 1981) noted, 'Well, my mother was a housewife all her life. My father worked in the *mines*, my father was a miner . . . . I was actually the only one that really worked in the *mines*'. When describing life in Italy that he left, Brazina explained, 'there was nothing there, so they all came and migrated here. They worked in the coal *mines*, and that was it'. Edward Lutchko

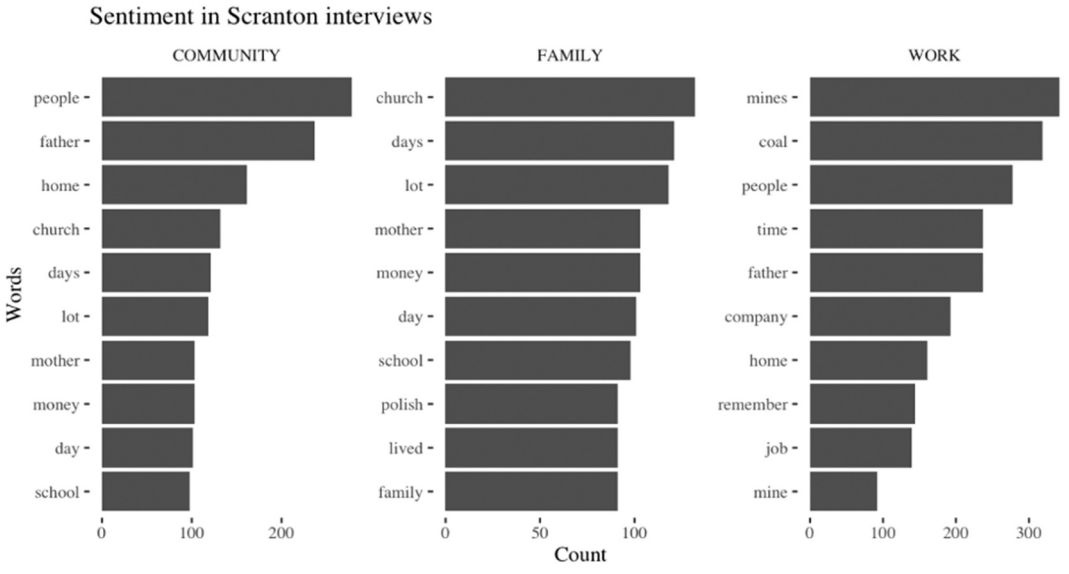

Figure 5. Top 10 words tagged by domain-specific themes in Scranton interviews.

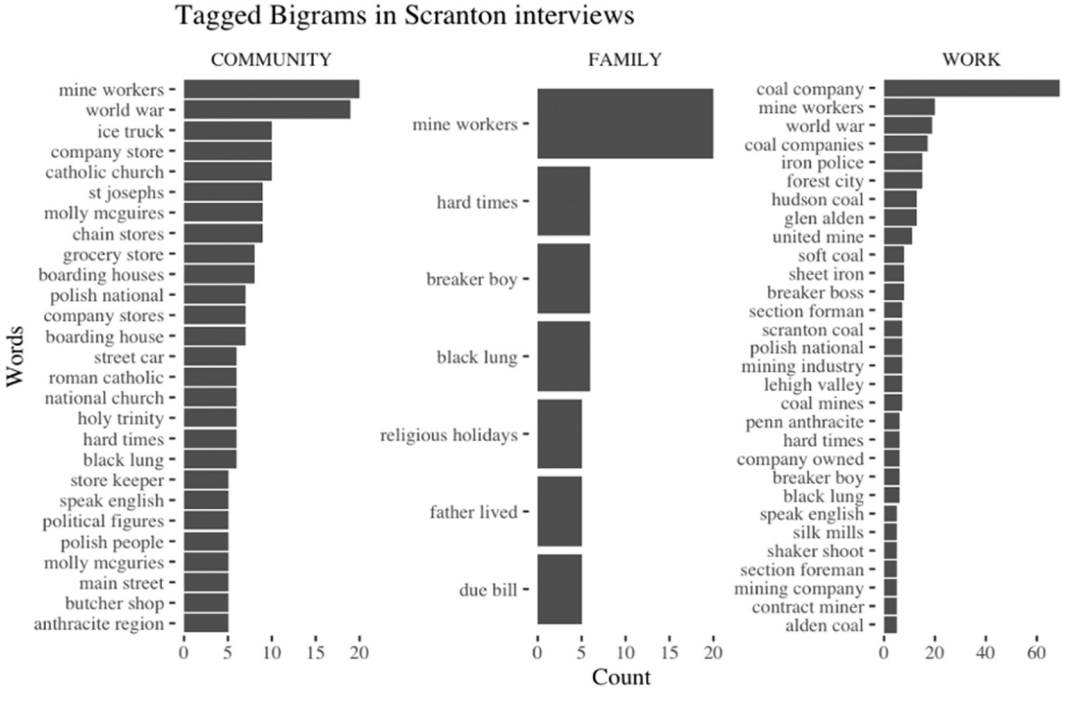

Figure 6. Tagged bigrams in Scranton interviews using domain-specific themes. Bigrams are ordered by the frequency with which they appear in the interviews.

explained, 'And in the *mines* you was your own boss. You load your *coal* when you go in and load your *coal*. They used to call me "here he comes and there he goes" . . . I had so much work to do and then you were out'.

In addition to tagging singular words, we also tagged pairs of words (bigrams) using the same themes (community, family, and work). Pairs of words are analytically interesting as they can reveal multi-word concepts and individual words that tend to co-occur. The high frequency of two-part concepts (e.g. theme of community – 'catholic church', theme of family – 'hard times', theme of work – "coal company ") in these oral histories demonstrates the strength of this approach for identifying relevant themes in these types of sources (Figure 6).

Figure 6 details the bigrams from the Scranton interviews tagged based on the three semantic themes. For instance, under the theme of family, the term 'Catholic Church' is frequently noted when describing daily life in this coal mining town. The term was sometimes used to describe the various forms of Catholicism. Edward Lutchko (Interview 1973) was born in the late nineteenth century and explained, 'Yes, I went to the Greek *Catholic Church* on 7th Street'. Walter Dutchak (Interview 1973), who is a second-generation Ukrainian-American explained how his father helped build a church. 'It became known as a Ruthenian-Greek *Catholic Church*. And one digs into the name of Ruthenian; it was still Latin for Ukrainian. And that was the church that my father was married in'. Earl Lamb (Interview 1973) explained the enclaves that developed around churches. He explained that in Centralia, 'there was a distinct division between the Protestant and Roman Catholic areas. The Roman Catholic areas were very built around the Roman *Catholic Church*. They were mostly the Irish descent. The Protestant community was a mixture of Welsh, English, Scotch'.

Under the theme of family, the term 'hard times' is frequently used to describe family struggles. Louis Glowaki (1973) described his wife's mother, who lost her husband and had to provide for her family. 'Her mother had *hard times* in those days'. His wife, Margie Glowaki, responded, 'We all had *hard times*. But thank goodness it all worked out. We worked hard for what we got'. In another case, John Parroccini (Interview 1973), an Italian immigrant, reflected on life in Italy. He was thankful that he did not have to experience the World Wars, although he recognised the difficulties his family faced in Italy. 'Over there, there was not much work in those days. . . . Of course, they went through *hard times*, during those wars. For that I thank God we were here'.

Under the theme of work, the term 'coal company' is often referred to as the place of employment. For instance, Stanley Guntack (Interview 1973) explained that he worked for the 'The Hudson *Coal Company*'. Anthony Naugin (Interview 1973) noted, 'when I hit the Hudson *Coal Company*, I stayed there for a while, for about 40 years'. Walter Dutchak (Interview 1973) described that when his father first arrived in Scranton, he was looking for a job and slept on a back porch. 'After three days, my father obtained work in the Sandy Banks colliery of the Scranton *Coal Company*'.

Discussion

What was life like in anthracite mining communities during the early 20th century? One snapshot of this chronotope is documented in the Scranton oral histories. Sentiment analysis with the Bing and AFINN lexicons highlights how particular terms used in the oral histories may be considered positive or negative in sentiment. Additionally, we piloted a custom, domain-specific lexicon to examine the themes of *family*, *community*, and *work*, as they appear in this context. Words explicitly drawn from the text analysis listed below clearly link our interpretations to the words used by narrators of the oral histories. The first approach (binary positive/negative sentiment tagging) suggests that memories of the coal mining era include the negative experiences associated with 'dangerous' and 'hard' labour, as well as the 'prejudice', and 'struggle' experienced by community members. On the positive side of experiences, we see qualifiers about 'clean', 'nice', 'lucky', and aesthetics appearing in the histories.

The second approach (tagging based on three themes) highlights the specific categories of places, people, experiences, and events that would have shaped daily life in these communities. For example, terms such as 'church' and 'school' highlight community spaces that might also have

been important for family life. Meanwhile, the names of various coal companies, actors, or materials in the coal mining industry describe the tangible and perceptual experiences of *work* in community members' memories.

Semantic tagging is a powerful tool for oral history analysis. When texts can be effectively searched for key terms related to topics of interest, it is possible to quickly parse lengthy or complex documents. There are several benefits this method can offer. First, it enables researchers to pinpoint sections of text to include for more detailed reading and analysis. Second, quantitative analysis may identify patterns across lengthy texts that could be overlooked through qualitative analysis alone. Combining a close reading of texts with systematic text analysis allows for a comprehensive approach to complex social texts such as oral histories and other texts of interest to social scientists and humanities scholars (Grimmer, Roberts, and Stewart 2022). A mixed-methods approach to both locating oral history sources and analysing their texts may also avoid the pitfall noted by Frisch: 'In the search-engine world it is simply too seductively easy to snag something on the run rather than to learn the skill of angling for it with just the right fly or lure suited to the setting, the conditions, and the object sought' (Frisch 2008, 224). Critically using general purpose and domain-specific lexicons can be one approach to supporting more focused searches through oral history data.

We piloted a tagging lexicon for words associated with *family, work*, and *community*. This domain-specific lexicon was compared with general purpose sentiment taggers to assess their relative strengths and weaknesses for analysing coal mining oral histories. Coding ethnographic and interview transcripts are not new approaches in anthropology, making them well-suited for potential integration into more computational natural language processing and text mining methodologies for anthropology. We echo calls made in other fields about the benefits of domain-specific sentiment and part-of-speech lexicons, stopword lists, and other text analysis tools for improving the interpretation of diverse text sources from particular thematic, spatial and temporal contexts (Burns 2018; Coden et al. 2005).

The pervasive physical reminders of this once powerful industry, still visible in the scarred landscape and culm banks scattered throughout the region, invoke a heritage of hard work, tough economic times, and the struggle of new immigrants to survive in an unyielding, unsympathetic, largely unregulated capitalist economy (Tarone 2004; Turner 1977, 1984).

Many people survived in the anthracite region in a climate of unchecked capitalism while confronting different forms of structural violence. Historically, residents have faced food insecurity, underemployment, a toxic environment, and unsafe work and living conditions. While coal is no longer the area's economic engine, many physical and ecological reminders point to the heedless destruction of unchecked capitalism. A chronotope is rooted in how one perceives the flow of events and its representation of particular worldviews and ideologies. These chronotopes of loss and decay predominate many Rust Belt narratives, often created by outside observers, often limiting the story of the communities being described. Our analysis allows us to see how this now historic community viewed and created its chronotope a half-century ago. Text mining these oral histories makes it possible to ascertain how community members themselves describe their memories of a particular place/time. While a system of unchecked capitalism oppressed the coal mining communities, they often remembered their past as powerful, surviving against the crushing odds of the industry. They spoke about collective action (United Mine Workers of America) and bettering the lives of their families. They spoke about the importance of family and how family members found different strategies to stave off starvation. Community in the form of church and company served as an anchor and support system. If we carefully delve into these oral histories, new, complimentary narratives can be created about these Rust Belt communities.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Madeline Brown is an Assistant Professor of Anthropology at the University of Maryland, College Park. Her work focuses on human-environment interactions with an interest in digital and mixed-methods research.

Paul A. Shackel is a Professor of Anthropology at the University of Maryland, College Park. His research projects focus on the role of archaeology in civic engagement and heritage activities related to race and labor.

References

Abrams, J. 1994. "Lost Frames of Reference: Sightings of History and Memory in Pennsylvania's Documentary Landscape." In *Conserving Culture: A New Discourse on Heritage*, edited by M. Hufford, 24–38. Urban/Champaign, Illinois, USA: University of Illinois Press.

Ahern, M. M., M. Hendryx, J. Conley, E. Fedorko, A. Ducatman, and K. J. Zullig. 2011. "The Association Between Mountaintop Mining and Birth Defects Among Live Births in Central Appalachia, 1996–2003." *Environmental Research* 11 (6): 838–846. https://doi.org/10.1016/j.envres.2011.05.019.

Bakhtin, M. M. 1981. *The Dialogic Imagination: Four Essays by M.M. Bakhtin*, Translated by Caryl Emerson & Michael Holquist. Austin, Texas, USA: University of Texas Press.

Basso, K. 1984. "Stalking with Stories: Names, Places, and Moral Narratives Among the Western Apache." In *Text, Play, and Story: The Construction and Reconstruction of Self and Society*, edited by E. Bruner, 19–55. Washington: American Ethnological Society.

Bernard, R. 2011. *R*esearch Methods in Anthropology: Qualitative and Quantitative Approaches. 5th ed ed. Lanham, MD, USA: AltaMira Press.

Bodnar, J. 1983. *Anthracite People: Families, Unions and Work, 1900-1940*. Harrisburg, PA: Pennsylvania Historical and Museum Commission.

Boyd, D. A., J. W. Fernheimer, and R. Dixon. 2015. "Indexing as Engaging Oral History Research: Using OHMS to "Compose History" in the Writing Classroom." *The Oral History Review* 42 (2): 352–367. https://doi.org/10.1093/ohr/ohv053.

Britt, E. 2018. "Oral History and the Discursive Construction of Identity in Flint, Michigan." *Journal of Linguistic Anthropology* 28 (3): 252–272. https://doi.org/10.1111/jola.12196.

Brown, and Shackel. 2023. Anthracite Oral Histories Analysis Repository. https://github.com/maddiebrown/OralHistories.

Burns, P. 2018. "Constructing Stoplists for Historical Languages." *Digital Classics Online* 4 (2): 4–20.

Classics Department, University of Cambridge. n.d. "Tagging the Lexicon." Accessed October 21, 2022. https://www.classics.cam.ac.uk/research/projects/glp/tagging.

Coden, A. R., S. V. Pakhomov, R. K. Ando, P. H. Duffy, and C. G. Chute. 2005. "Domain-Specific Language Models and Lexicons for Tagging." *Journal of Biomedical Informatics* 38 (6): 422–430. https://doi.org/10.1016/j.jbi.2005.02.009.

Conlogue, B. 2013. *Here and There: Reading Pennsylvania's Working Landscape*. University Park, PA, USA: The Pennsylvania State University Press.

Deasy, G. F., and P. R. Griess. 1961. "Tourism for the Anthracite Region: An Alternative for Unemployment." *Mineral Industries* 30 (7): 1–8.

Dublin, T. 1998. *When the Mines Closed: Stories of Struggles in Hard Times*. Ithaca, NY: Cornell University Press.

Dublin, T., and W. Licht. 2005. *The Face of Decline: The Pennsylvania Anthracite Region in the Twentieth Century*. Ithaca, NY: Cornell University Press.

Frisch, M. 2008. "Three Dimensions and More: Oral History Beyond the Paradoxes of Method." In *Handbook of Emergent Methods*, edited by S. N. Hesse-Biber and P. Leavy, 221–238. New York: The Guilford Press.

Gallo, A. 2017. "More Holes Open on Property of Family Who Nearly Lost Home." *WNEP com*. http://wnep.com/2017/01/30/more-holes-open-on-property-of-family-who-nearly-lost-home/.

Glaeser, E. L., J. D. Gottlieb, and O. Ziv. 2014."'Unhappy Cities.' Working Paper 20291." Cambridge, MA: National Bureau of Economic Research, Accessed December 19, 2022. http://www.nber.org/papers/w20291.

Greene, V. R. 1968. *The Slavic Community on Strike: Immigrant Labor in Pennsylvania Anthracite*. South Bend, IN: University of Notre Dame Press.

Grimmer, J., M. E. Roberts, and B. M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton, NJ, USA: Princeton University Press.

Hand, M. 2015 September 29. "West Virginia Miners Play Second Fiddle to the Molly Maguires." *CounterPunch*. Accessed December 19, 2022. http://www.counterpunch.org/2015/09/29/west-virginia-miners-play-second-fiddle-to-the-molly-maguires/.

Hendryx, M., L. Wolfe, J. Luo, and B. Webb. 2012. "Self-Reported Cancer Rates in Two Rural Areas of West Virginia with and without Mountaintop Coal Mining." *Journal of Community Health* 37 (2): 320–327. https://doi.org/10.1007/s10900-011-9448-5.

High, S. 2013a. "Beyond Aesthetics: Visibility and Invisibility in the Aftermath of Deindustrialization." *International Labor and Working-Class History* 84:140–153. https://doi.org/10.1017/S0147547913000276.

High, S. 2013b. "'The Wounds of Class:' A Historiographical Reflection on the Study of Deindustrialization, 1973–2013." *History Compass* 11 (11): 994–1007. https://doi.org/10.1111/hic3.12099.

High, S. 2017. "Beyond the Juicy Quotes Syndrome." In *Going Public: The Art of Participatory Practice*, edited by E. Miller, E. Little, and S. High, 110–129. Vancouver: UBC Press. https://doi.org/10.59962/9780774836647-005.

Hitt Nathaniel, P. A. M. H. 2010. "Ecological Integrity of Streams Related to Human Cancer." *Ecohealth* 7 (1): 91–104. https://doi.org/10.1007/s10393-010-0297-y.

Hvitfeldt, E. 2022. "textdata: Download and Load Various Text Datasets. R Package Version 0.4.4. https://CRAN.R-project.org/package=textdata.

Johnson, K. 2021. "The Classical Language Toolkit (CLTK)." Accessed November 16. http://cltk.org/ .

Mah, A. 2009. "Devastation but Also Home: Place Attachment in Areas of Industrial Decline." *Home Cultures* 6 (3): 287–310. https://doi.org/10.2752/174063109X12462745321462.

Mah, A. 2010. "Memory, Uncertainty and Industrial Ruination: Walker Riverside, Newcastle upon Tyne." *International Journals of Urban and Regional Research* 34 (2): 398–413. https://doi.org/10.1111/j.1468-2427.2010.00898.x.

Mah, A. 2012. *Industrial Ruination, Community and Place: Landscapes and Legacies of Urban Decline*. Toronto, Ontario, Canada: University of Toronto Press.

Marsh, B. 1987. "Continuity and Decline in the Anthracite Towns of Pennsylvania." *Annals of the Association of American Geographers* 77 (3): 337–352. https://doi.org/10.1111/j.1467-8306.1987.tb00163.x.

McGillivray. 2021. "Computational methods for semantic analysis of historical texts." In *Routledge International Handbook of Research Methods in Digital Humanities*, edited by K. Schuster and S. Dunn, 261–274. London: Routledge. https://doi.org/10.4324/9780429777028-20.

Miller, D. L., and R. E. Sharpless. 1998. *The Kingdom of Coal: Work, Enterprise, and Ethnic Communities in the Mine Fields*. Philadelphia: University of Pennsylvania Press.

Minqing, H., and B. Liu. 2004. "Mining and Summarizing Customer Reviews." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004), Seattle, Washington, USA, Aug 22-25.

Moore, N. 2006. "The Contexts of Context: Broadening Perspectives in the (Re)use of Qualitative Data." *Methodological Innovations Online* 1 (2): 21–32. https://doi.org/10.4256/mio.2006.0009.

New York Times. 1869a. "The Sunken Coal Mine: The Latest Pennsylvania Mining Horror—A Block of Houses Sink into a Mine—Ten Persons Engulfed—Apprehension of Further Disaster." December 20.

New York Times. 1869b. "Ten or More Persons Buried Alive. Disaster in the Coal Region. Caving is of a Pennsylvania Coal Mine. The Shaft Choked up. Fall of Two Dwellings with the Bank." December 19.

Nielsen, F. A. 2011. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages (718) in {CEUR} Workshop Proceedings, 93–98. May. http://arxiv.org/abs/1103.2903.

NLTK Project. 2022. "Natural Language Toolkit." Accessed November 16. https://www.nltk.org/.

Palladino, G. 2006. *Another Civil War: Labor, Capital, and the State in the Anthracite Regions of Pennsylvania, 1840-1868*. New York: Fordham University Press.

Peeren, E. 2006. "Through the Lens of the Chronotope: Suggestions for a Spatio-Temporal Perspective on Diaspora." In *Diaspora and Memory: Figures of Displacement in Contemporary Literature, Arts and Politics*, edited by M.-A. Baronian, S. Besser, and Y. Jansen, 67–78. New York, NY: Rodopi.

Raushenbush, H. S. 1924. *The Anthracite Question*. New York: H.W. Wilson Co.

R Core Team. 2022. *"R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rieping, H. A., 2022. *Audio Segmenting and Natural Language Processing in Oral History Archiving* (Masters dissertation, Massachusetts Institute of Technology).

Rinker, T. W. 2018. "textclean: Text Cleaning Tools." https://github.com/trinker/textclean.

Roberts, P. (1904) 1970. *Anthracite Coal Communities*. New York: Arno Press and the New York Times.

Roller, M. P. 2015. ""Diabolical Consumerism: Mass Psychology a Between The Gilded and the Golden Ages." In *Historical Archaeologies of Capitalism*, edited by M. Leone and J. Knauf, 25–34. New York: Springer International. https://doi.org/10.1007/978-3-319-12760-6_2.

Shackel, P. A. 2017. "Anthracite Heritage: Landscape, Memory and the Environment." *Open Rivers: Rethinking Water, Place and Community*. Summer. http://editions.lib.umn.edu/openrivers/article/anthracite-heritage-landscape-memory-and-the-environment/.

Shackel, P. A. 2019. "Structural Violence and the Industrial Landscape." *International Journal of Heritage Studies* 25 (7): 750–762. https://doi.org/10.1080/13527258.2018.1517374.

Shackel, P. A. 2020. *An Archaeology of Unchecked Capitalism: From the American Rust Belt to the Developing World*. NY, USA: Berghahn Books. https://doi.org/10.2307/j.ctv1dwq0wr.
Shackel, P. A. 2023. *The Ruined Anthracite: Historical Trauma in Coal Mining Labor Communities*. Champaign, IL: University of Illinois Press.
Silge, J., and D. Robinson. 2016. "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *JOSS* 1 (3): 37. https://doi.org/10.21105/joss.00037.
Silge, J., and D. Robinson. 2017. *Text Mining with R: A Tidy Approach*. O'Reilly Media, Inc. https://www.tidytextmining.com/.
Smith, R. M. 1886. "American Labor Statistics." *Political Science Quarterly* 1 (1): 45–83. https://doi.org/10.2307/2139302.
Smith, R. M. 1887. "Review: Emigration and Immigration: Reports of the Consular Offices of the United States." *Political Science Quarterly* 2 (3): 520–522. https://doi.org/10.2307/2139189.
Steckel, R. H., and G. Senney. 2015. "Historical Origins of a Major Killer: Cardiovascular Disease in the American South." *Working Paper 21809*. National Bureau of Economic Research. December. Accessed December 19, 2022. http://www.nber.org/papers/w21809.
Steven, H., and D. W. Lewis. 2007. *Corporate Wasteland: The Landscape and Memory of Deindustrialization*. Ithaca, NY: Cornell University Press.
Tarone, L. A. 2004. *We Were Here Once: Successes, Mistakes, & Calamities in Hazleton Area History*. Hazleton, PA: Citizen Publishing, Company, Inc.
Turner, G. A. 1977. "The Lattimer Massacre and Its Sources." *Slovakia* 27 (50): 9–43.
Turner, G. A. 1984. "Ethnic Responses to the Lattimer Massacre." In *Hard Coal, Hard Times: Ethnicity and Labor in the Anthracite Region*, edited by D. L. Salay, The Scranton Museum Association, in cooperation with the Commonwealth of Pennsylvania, Pennsylvania Historical and Museum Commission, 26–152. Scranton, PA: The Anthracite Museum Press.
Wallace, A. F. C. 1987. *St. Clair: A Nineteenth-Century Coal Town's Experience with a Disaster Prone Industry*. New York, NY, USA: Knopf.
Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://doi.org/10.1007/978-3-319-24277-4.
Wickham, H. 2022. "stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.1." https://CRAN.R-project.org/package=stringr.
Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.
Zawacki, G. 2017. "Abandoned Mine Drainage: Past Causes and Present Impacts. Anthracite Environmental." https://anthraciteenvironmental.org/2015/06/12/abandoned-mine-drainage/.