Energy Research & Social Science 130 (2025) 104398

Contents lists available at ScienceDirect

Energy Research & Social Science

journal homepage: www.elsevier.com/locate/erss

Original research article

Dirty talk: Media discourse and the struggle over South Africa's coal transition☆

Charlotte S. Bez [a,b,*], Giacomo Raederscheidt [c], Jan C. Steckel [a,d]

[a] Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany
[b] The New School For Social Research, NYC, USA
[c] University of Potsdam, Potsdam, Germany
[d] Technical University of Munich, Munich, Germany

ARTICLE INFO

Keywords:
Energy transitions
Coal
South Africa
Discursive power
Topic modeling
Text as data

ABSTRACT

This paper investigates media discourses on coal in South Africa to understand the state of progress of its energy transition. We approach discourse as a key element in policy sequencing, shaping the conditions under which regulatory change becomes possible. Using an integrated text-as-data pipeline, including topic modeling, named entity recognition, and sentiment analysis, we analyze approximately 8,000 national newspaper articles from 2010 to 2024. Topic modeling reveals five discourse clusters: Mining industry, Transition politics, Energy crisis, Mining affected communities, and Politics in mining and energy. While new technologies receive attention – particularly around the announcement of the Just Energy Transition Partnership with South Africa in 2021 – explicit discourse on coal phase-out remains marginal. At the same time, we highlight that these technologies are frequently framed as substitutes for structural change. Positive sentiment toward the mining sector and transition politics suggests that the transition is not framed as a coal exit. In this sense, the observed framing reveals not only media priorities but also the discursive power of dominant actors in structuring transition pathways. Our findings underscore the value of combining computational and qualitative approaches innovatively to examine socio-political transition discourses, with implications for broader applications of text-as-data in policy research. Our results inform debates on the political economy of energy transitions at the interplay of international climate finance.

1. Introduction

Energy transitions and phasing down coal have been debated prominently, both in the political sphere and in academia. Yet, the majority of coal-dependent low- and middle-income countries are still expanding their capacity of coal-fired electricity generation [1]. South Africa constitutes a representative example of a national economy heavily reliant on coal production and export. Not only is it one of the major exporters of coal [2], it also generates almost all of its electricity from it [3]. While South Africa has committed to achieving net-zero emissions by 2050 [4], coal remains nationally significant. It provides jobs and revenues in a country with high unemployment rates, deep socio-economic inequalities, and persistent electricity shortages. Strong interest groups advocate continued coal use [5,6].

These circumstances make the notion of a *just* energy transition particularly relevant for South Africa, as it conceptualizes the shift from fossil-based energy production to renewable energy sources while ensuring fairness and inclusiveness in the process. Importantly, South Africa was the first country to engage in a Just Energy Transition Partnership (JETP) with a group of high-income countries, aimed at accelerating its coal phase-out with international support. By placing South Africa at the center of this new model of international climate cooperation, the JETP not only reflects the urgency of addressing coal dependence at a national level, but also establishes South Africa as the test case for global coal phase-out strategies. Its success or failure will strongly influence how other coal-dependent economies are engaged in the future, and it offers unique insights into the interplay of international climate finance, energy security, and the politics of socio-economic transformation.

This paper investigates the state of South Africa's energy transition away from coal since 2010, focusing on how public discourse reflects

☆ Charlotte S. Bez and Jan C. Steckel acknowledge financial support of the Werner Siemens Foundation, Switzerland and funding through the TRANSFORM project by the German Federal Ministry for Economic Cooperation and Development, Germany.

* Corresponding author at: Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany.
E-mail addresses: charlotte.sophia.bez@pik-potsdam.de (C.S. Bez), raederscheidt@outlook.com (G. Raederscheidt), Jan.Steckel@pik-potsdam.de (J.C. Steckel).

https://doi.org/10.1016/j.erss.2025.104398
Received 18 April 2025; Received in revised form 23 September 2025; Accepted 9 October 2025
Available online 5 November 2025

and potentially shapes this ongoing process. Given that the transition is still underway and policy outcomes are not yet fully observable ex post, we adopt a discursive lens to examine how the transition is framed, contested, and legitimated in the public arena. Specifically, we analyze shifts in newspaper coverage to trace evolving discourses, actor positions, and the salience of coal and coal phase-out in national media discourses.

Our theoretical approach builds on two interrelated strands of literature. First, we draw on the policy sequencing literature [7,8], which highlights the importance of the temporal order in which policy instruments are introduced to enable deep transitions such as coal phase-out. We extend this perspective by emphasizing the role of political and societal actors in shaping the conditions under which such sequences can be initiated and sustained. In line with [9], we argue that policy change is not only a matter of technocratic policy design but also of changing positions by relevant actors in politics and society.

Second, we engage with theories of discursive power and governmentality, following Foucault's insight that power operates through the production of knowledge and the shaping of norms and narratives. In the context of energy transitions, dominant discourses often reflect the interests of incumbent actors who seek to maintain the legitimacy of carbon-intensive systems [10,11]. Conversely, shifts in discourse can indicate the emergence of new coalitions, ideas, and policy windows.

Media discourse analysis generally speaking has gained traction in energy and sustainability studies as a way to understand the political dynamics underpinning transitions [e.g., 12–15]. We build on this literature by treating newspapers as decisive sites where political arguments are made, contested, and circulated. As agenda-setting platforms, newspapers shape public perceptions and influence the framing of energy issues within the policymaking sphere [16,17]. Thus, they provide a valuable empirical window into how actor groups attempt to influence policy debates and shape the trajectory of the coal transition. Against this theoretical background, our study maps discursive topics, actor groups, and attitudinal framings in South African newspaper coverage of coal. We aim to uncover how discourses around coal and the broader energy transition evolve over time, particularly in relation to international interventions such as the Just Energy Transition Partnership (JETP). By analyzing shifts in discourse, we assess the extent to which the coal phase-out is becoming politically and socially viable.

So far, existing research on South Africa's energy transition discourse has relied on qualitative methods such as expert and stakeholder interviews [see 5,18,19]. However, it remains unclear how the energy transition debate sits within the general public coal discourse in South Africa. It is crucial to comprehend the trajectory and importance of discourse associated with coal in order to contextualize its role in influencing public perceptions and policy.

Profiting from new computational techniques and the growing availability of textual data, social sciences increasingly use text data as valuable input for innovative ways to uncover patterns in discourse [20]. We apply state-of-the-art text-as-data methods to answer the above-mentioned research goals. In that way, we propose an exploratory combination of text-as-data approaches that can be applied to a wide range of research questions within the field of energy transitions and beyond. Specifically, we analyze a dataset of approximately 8,000 South African newspaper articles containing the word *coal*, published between 2010 and 2024, using dynamic topic modeling. We then integrate the results with a keyword search of a subset of technologies, with named entity recognition to identify key actors and finally apply automated sentiment analysis to detect prevailing sentiments by actors and topics.

The remainder of this paper is structured as follows. In Section 2, we discuss literature on policy analysis to evaluate discourses and the necessary background on the role of coal and other forms of energy in South Africa. Section 3 describes the newspaper data and the data preprocessing. In Section 4, we motivate the choice of each method, explaining the methodologies, and then sequentially presenting the results of the analysis. Section 5 discusses the findings and points to the limitations. Finally, Section 6 concludes.

2. Literature

Our study relates to several strands of the literature. In the following subsections, we outline how public policy studies analyze discourse, operationalized through text-as-data. We then highlight the role of coal in South Africa's political economy and assess the growing integration of carbon management technologies into this landscape.

2.1. Policy analysis to evaluate discourses

Public policy studies apply various traditions of discourse analysis to understand how policy agendas are constructed and contested, as outlined in [21]. In the tradition of Foucault it can uncover the ways in which language, knowledge, and power shape policy narratives and institutional dynamics. Discursive methods have become increasingly prominent in the study of energy *transitions*, particularly through discourse analysis, which examines how language, narratives, and framing influence energy policy and transition processes [see the review by15]. This body of scholarship has been especially insightful in analyzing institutional change and perceptions of technologies. Public discourse analysis typically seeks to identify dominant energy narratives and points of contention [e.g., 17,22,23]. Furthermore, [24] shows how discourse analysis can uncover the construction and contestation of policy narratives. Methodologically, the author advances the robustness of discourse analysis aimed at revealing how alternative interpretations can challenge dominant narratives. More recent contributions focus on how discourse changes over time within transitions, by identifying distinct temporal phases and detecting significant shifts [e.g., 25].

Yet, existing research focusing on news media discourse tends to concentrate on the European context.[1] For instance, [14] analyze how coal is portrayed across various European coal-dependent countries, highlighting strong overlaps between media narratives and stakeholder perspectives. Similarly, [13] underscore media neglect of coal mining's environmental impacts, while [16] reveal recurring discourse dynamics in coal phase-out debates in Northern Europe. Despite the particular relevance of public perceptions in coal-dependent countries of the Global South, such as South Africa [26], there remains a significant gap in empirical research addressing this context.

2.2. Text-as-data methods in the realm of discourse analysis

Recognizing the limitations of traditional discourse analysis tools for dealing with large text corpora [12,27], recent literature increasingly employs statistical techniques that treat text as data. Automated algorithms can analyze large amounts of text data quickly, saving considerable time and effort compared to manual categorization [27]. A cornerstone of quantitative text analysis methods is the (semi-)automated analysis and identification of latent topics in extensive text collections, known as topic modeling. Through detecting latent topics and their evolution across a large corpus of documents, topic modeling offers researchers a means to uncover hidden patterns and attain more profound insights into the underlying clusters and trends within discourses [28]. Topic modeling has been used in both newspaper media analyses [e.g., 29] and in the analysis of textual data related to energy policy [e.g., 12,30,31]. However, its application to energy-related newspaper discourses remains limited to a small number of studies [see 13,32], despite [33] having called for empirical novelty in energy social science for many years. Although a significant body of research employs text-as-data methods, most studies tend to rely on a

[1] Notably, also the review by [15] does not include empirical evidence from country studies in the Global South.

single isolated approach, such as topic modeling or sentiment analysis.[2] We argue that combining multiple text-as-data techniques can enhance the power of automated content analysis, particularly in the study of energy transition discourses.

2.3. Coal in South Africa

Historically, the importance of coal in South Africa's energy economy is largely attributable to the country's abundant coal resources. This has been supported by low-cost labor for extraction and government subsidies for affordable electricity generation, which has fueled energy-intensive mining and associated industrial activities [34]. Today, the coal mining sector and coal-fired power stations collectively employ over 130,000 workers [19], with the majority of them being located in the Mpumalanga province, which accounts for 80% of the national coal production [35].

South Africa's energy sector faces major challenges, particularly with Eskom, the largest consumer of domestically produced coal and responsible for nearly half of the nation's carbon emissions [36]. Mismanagement, corruption, and delays in coal plant projects, coupled with tariffs not reflecting underlying generation costs, have left the utility financially unstable [34]. This has resulted in severe national electricity shortages [37]. While Eskom is building new coal-fired power plants to meet demand [34], renewable energy capacity is rising, pushed by decreasing costs and growing involvement of Independent Power Producers (IPPs). To ensure energy stability, decision-makers are increasingly focusing on diversifying the sector [5]. At the same time, high levels of unemployment, inequality, and poverty complicate the country's efforts to meet its national climate targets and transition away from coal [5,19].

2.4. The political economy of coal revisited

Transition scholars have conceptualized the status quo in the country's energy politics as heavily influenced by the so-called *minerals-energy complex*, describing a powerful nexus of political and economic interests involving state-owned enterprises, major mining companies, and government policies that have historically prioritized coal as a cornerstone of the South African economy [38,39]. Even today, these incumbent actors, and in particular Eskom, cling to the preservation of coal mining and combustion, partly resisting the shift toward renewable energy systems [36]. Generally, economic beneficiaries of coal industries advance discourses of regional development risks and green growth, while highlighting employment loss, hence framing coal as essential for economic prosperity [40]. Indeed, research has shown that coal-related vested interests represent large barriers to South Africa's coal phase-out [5,19,41]. Additionally, social actors, in particular some of the trade unions advocating for mining workers, also play a role as they actively oppose coal phase-out [41]. In fact, the discussion regarding coal phase-out and transition to renewable energy is described as heavily focusing on jobs and associated risks for coal-dependent regions like Mpumalanga [5,42].

A connected discourse has emerged around speculative carbon management technologies, increasingly positioned as central to future climate strategies despite their uncertainty in terms of scientific viability and economic scalability [43,44]. Technological optimism underpins what some describe as "fossil fuel solutionism" by sustaining imaginaries of technological fixes, thereby delaying fossil fuel phase-out and aggravating carbon lock-ins [44,45]. Recently, a set of these technologies has been critiqued as "false solutions" [see 46]. Prior studies have examined these technologies in isolation: the nexus between carbon capture and storage (CCS) and hydrogen [44], CCS specifically [43], the interplay of renewables and natural gas [47], and direct air capture [48]. We notice a lack in *empirical* grounding in how carbon management technologies are discursively framed and deployed within specific fossil-fuel dependent economies. We address this gap by analyzing how carbon management technologies are represented and contested within the coal and energy transition discourse in terms of investment decisions, policy support, academic collaborations, and narrative promotion. By unpacking these discourse categories, we show how potentially contested technologies are legitimated and mobilized, thus ultimately sustaining fossil fuel dependence. Hereby, we make an empirical contribution to the concept of transition discourse co-optation by fossil interests [see 49,50].

3. Data and methods

3.1. Data

To analyze the national coal discourse comprehensively, we leverage the LexisNexis data archive's full-text access. We conduct a search query for English-language news articles published in South African news media within a time frame of ten years, between January 2010 and January 2024, containing the word *coal*. As part of data preprocessing, we exclude regional and international newspaper outlets to focus our analysis exclusively on the national level. Simultaneously, we make sure that our dataset contains the most relevant outlets in terms of readership. Moreover, we remove duplicate articles with a stringent 80 percent similarity threshold, determined empirically through multiple rounds of testing and validation, and supported by the distribution of similarity scores. Finally, we iteratively screen and filter out false positives[3] and drop documents incorrectly identified as full-text articles, i.e., articles containing less than 70 words. The final dataset consists of 8,071 articles coming from newspaper articles from 18 different outlets (refer to Figure A1 in Appendix A for further information).

Fig. 1 below illustrates the monthly distribution of coal-related articles within the timeframe of our analysis. Visibly, the number of published articles was relatively high throughout the earlier years and during the most recent years. The average number of articles per year is 577. Most articles were published in 2011, the fewest in 2020. Lastly, coal is mentioned on average three times per article.

3.2. Methods

This paper utilizes text-as-data techniques to explore the discourse on South African coal as presented in newspapers. We analyze discourse in a broad sense to refer to the systematic investigation of public discourses, actor framings, and attitudinal valences in media representation. Fig. 2 provides a comprehensive step-by-step overview of the research process. The methods (in purple) are interdependent: Dynamic Topic Model provides thematic context that informs both Named Entite Recognition (to focus on entity extraction within coherent discursive frames) and Sentiment Analysis (to ensure valence is measured in a topic-aware manner). We depart from the assumption that meaning depends on both *what* is being discussed and *who* is speaking in *which* tone. In other words, our methodological design allows us to capture not only discursive content but also its framing and sentiment over time, which is essential to understanding the public narrative around coal. All analytical steps are manually validated to achieve accuracy and reliability of the results (in orange). We begin with the pre-analysis steps, which involve data extraction, followed by the filtering and preprocessing of the articles (indicated in green).

[2] We are unaware of relevant scholarship in the sphere of climate- and energy studies that applies several methods to the same text corpus and interpret the results in association with each other. While recent advances in NLP often center around model development, our contribution lies in adapting these tools for substantive application within environmental social science.

[3] Examples include articles containing *coal* only as part of a composite word like *charcoal*, *coaling*, or *coalstove/coal stove*, or within idiomatic expressions like *to hoal someone over the coals*.

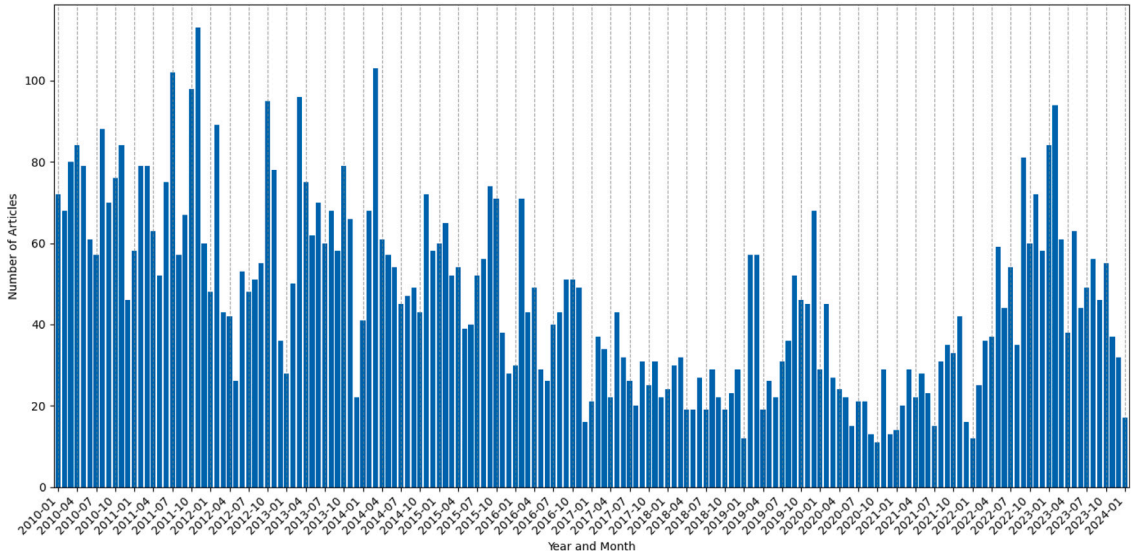


Fig. 1. Distribution of the number of articles per month between 2010 and 2024, labeled in 3-months intervals.

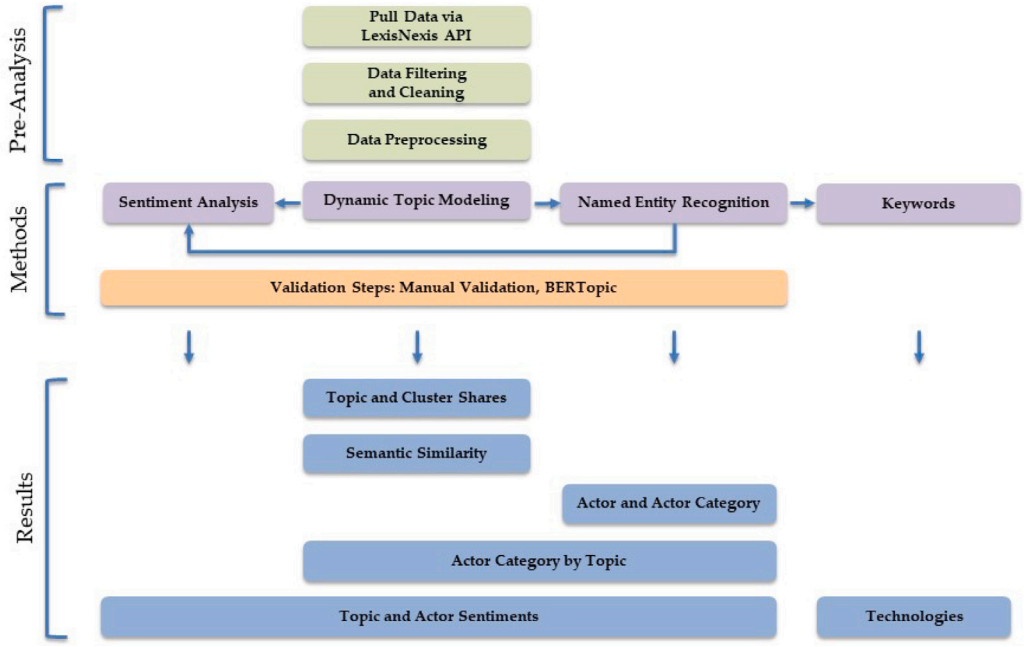


Fig. 2. Illustration of the research pipeline. Green panels refer to pre-analysis steps, purple panels refer to methods, where arrows indicate sequential dynamics of the methods' implementation. The orange panel refers to the validation steps as part of the method. Last, the blue panels indicate the results, together showing what is being discussed and who is speaking in which tone. The horizontal position of blue panels indicates which method has been leveraged for which result, e.g., "Topic and Actor Sentiments" as a joint result by leveraging Sentiment Analysis, Dynamic Topic Modeling, and Named Entity Recognition (in purple).

To exploratively uncover the discursive topics and their evolution over time in our large corpus of South African newspaper articles, we make use of automated topic modeling. We adopt the time-based DTM introduced by [51] as the preferred approach (compare "Dynamic Topic Modeling" in Fig. 2). Our time frame of 15 years encompasses significant developments that we expect to be reflected in discursive dynamics. [12] provide a robust implementation of this approach, whose guidance we follow. DTM enables modeling the evolution of topics relative to each other and capturing changes in language used within the same topic over time (the weight of words within each topic can vary over time). This is done by slicing the documents in the corpus into consecutive time periods (here, monthly) and creating model dependencies between topics in neighboring time bins. In line with [51], this approach links topic distributions across adjacent intervals so that topics evolve smoothly over time rather than being learned independently in each slice.

Prior to the topic model analysis, we have to pre-process the input data to improve model performance, reduce noise, and improve interpretability (compare "Data Preprocessing"), consciously acknowledging that these choices can influence the topic model results [27]. After tokenizing the text, we apply the standard approach of converting all tokens to lowercase and removing non-alphabetic characters. Moreover, we implement stemming to strip suffixes from words and

eliminate common stopwords.[4] In a sequential process, we use the topic model outcomes to improve pre-processing and ensure insightful and robust topic model results, e.g., by detecting and removing false positives. Finally, following the example of [52] we group the pre-processed articles by month to create time slices and convert each document into a bag-of-words format to produce the ready-to-use corpus, by creating a dictionary (vectorized) representation of articles.

One of the key parameters of unsupervised probabilistic topic models is the appropriate number of topics in the corpus, which has to be manually determined. As is common in the literature, we balance semantic coherence and exclusivity to maximize semantic interpretability and achieve interpretable topics [12]. To find the optimal number of topics k, we run models with k between five and thirty, in five-step intervals. A careful interpretation of Figure A2 in Appendix A yields an optimal model with twenty topics.

We implement the DTM by compiling the original software provided by [51].[5] The model produces two different matrices as output. One for the topic distribution over articles and another for the word (or token) distribution over topics. To make sense of the topics identified by the model, we need to interpret the content of each topic and find suitable topic labels. For this, following [12], we examined the 30 top-scoring words for each identified topic (see Table B1 in the Appendix B) and the ten most representative documents for each topic, i.e., documents with high topical content.

We use topic modeling to identify topics, analyze their prevalence, and their evolution over time. Additionally, we compute co-occurrence matrices to examine the semantic relationships between the identified topics (see "Topic and Cluster Share" and "Semantic Similarity"). It indicates how much a topic co-occurs with another topic, evaluated over its share of newspaper articles. We visualize weaker and stronger similarities by the use of a heat map.

As an intermediate analytical step, we search the data corpus for keyword mentions of so-called carbon management technologies (see "Keywords"), that have been labeled as "false solutions" in recent literature: Carbon capture and storage, clean coal, offset markets, and hydrogen [46]. These reflect a broad range of technologies, namely emission controls, alternative fuels, market-based mechanisms, and fuel efficiency, respectively. We identify approximately 300 mentions of these four technologies. We then leverage the software MAXQDA to qualitatively investigate the role of carbon management technologies.[6] For each narrative, we lay out examples of discourse representation and indicate its source (date of publication and news outlet), shown in Table B3 in Appendix B. This keyword search serves as an exploratory tool to gauge which carbon management technologies the actors mention and promote.

Topic modeling is an effective method for uncovering overarching clusters and patterns within large textual datasets. However, it falls short when it comes to precisely identifying the relevant actors. To systematically uncover the key actors in the South African coal discourse, we employ automated named entity recognition (NER) (see "Named Entity Recognition"). NER identifies and classifies entities in text based on pre-defined categories like persons, organizations, or locations. We operationalize NER using the SpaCy library,[7] known for its speed and efficiency. For its implementation, we extract the sentences containing the term *coal* and the two bordering sentences as we want to narrow down the actor analysis to the coal-related parts of the articles. We use a custom entity ruler – a tool that applies predefined patterns to detect and classify specific entities – to improve the performance of the model in recognizing actors specific to the coal discourse in South Africa and to assign them to the correct group. Finally, we predict entities within the extracted sentences and manually validate the results to ensure accuracy. We further group the identified entities into different actor categories (see "Actor and Actor Category"). Furthermore, inspired by [53], we link the actors identified on a sentence level with the most dominant topic of each article identified through DTM. By combining NER with topic modeling results, we can link entities to specific topics, providing a clearer picture of actor involvement in the discourse (see "Actor Category by Topic"). For example, if *Eskom* is mentioned in a snippet that is majorly attributed to the "Energy crisis" topic cluster, this will show up in Fig. 7 in *Eskom*.

As a further step in the analysis, we apply automated sentiment analysis (see "Sentiment Analysis"). It is a text-mining technique that evaluates language to determine underlying attitudes or opinions, positioning text along a positivity–negativity continuum and generating a compound sentiment score. This method reveals how topics are framed in the media by quantifying the sentiment expressed in the text. For the sentiment analysis, we use the Valence Aware Dictionary and sEntiment Reasoner (VADER) algorithm [54]. VADER is a pre-built, unsupervised (i.e., rule-based) sentiment analysis tool with an internal dictionary that labels text documents as positive, negative, and neutral. This algorithm approach relies on opinion words indicating sentiment polarity. Positive opinion words denote desired states, while negative opinion words denote undesired states. It has been widely used in the literature, see for instance [55,56], and has shown very good predictive sentiment performance.

In our analysis, we apply sentiment analysis in an advanced way by integrating it with topic modeling and NER (see "Topic and Actor Sentiments"). First, topic modeling extracts the main topics from the corpus, and then sentiment analysis is conducted at the sentence level to assess the tone associated with each topic cluster. Additionally, by identifying actors through NER, we can pinpoint sentiment not only about general topics but also directed toward specific entities. This novel approach, which combines several NLP techniques, builds on the few recent studies in this area [32,53].

Specifically, after computing the word-specific sentiment scores, we calculate the weighted average sentiment score for each sentence, where the weight is the frequency of the word in the sentence. First, we apply the sentiment analysis on the same dataset used in 4.4 on named entity recognition, hence calculating sentiments by actor. Second, we follow the same approach but calculate the sentiments by topic cluster, based on the dataset used in Section 4.1.

3.3. Validation and robustness

We manually validate all methods (in orange in Fig. 2). First of all, validating the DTM involves inspecting the coherence of topics by examining top-ranked words and representative documents within each cluster. We assess whether these words form semantically meaningful groupings and whether assigned documents reflect the intended theme.

In addition to manual validation, we benchmark DTM against machine learning-based BERTopic [57] as a robustness check to visualize

[4] Stemming is performed by using the Porter Stemmer contained in the nlkt library. Further, we use the nlkt library's standard list of English stopwords. Additionally, we manually check the articles and topic model outcomes to find more domain-specific words to add to the list of stopwords.

[5] In DTM, articles are treated as mixtures of topics rather than being allocated to a single most pervasive topic. For a detailed description of the DTM method, please refer to https://github.com/blei-lab/dtm/tree/master (last retrieved on 01/08/2024).

[6] MAXQDA supports systematic coding of the dataset by enabling the identification and organization of discourse elements related to different technologies. First, we performed a keyword-based search to retrieve relevant articles mentioning the carbon management technologies as introduced in Section 4.3. Next, we applied thematic coding in MAXQDA to categorize text segments by technology type, actor references, and framing (e.g., technological optimism, policy support). Each coded segment was linked to metadata, including the publication date and news outlet, as shown in Table B3 in Appendix B.

[7] For further documentation on SpaCy, see https://spacy.io/api/entityrecognizer/.

inter-topic relationships (see "Steps"). BERTopic leverages BERT embeddings to capture the semantic context in which words appear, producing more coherent and interpretable topic clusters than traditional models like LDA or DTM. Its ability to map semantically related articles and clusters in a low-dimensional space, combined with strong visualization capabilities, makes it particularly well-suited to address our second research question on how coal transition topics are embedded within the broader discourse. While BERT embeddings are widely used in text-as-data research [e.g., 58], BERTopic remains relatively underexplored [59,60], and its application in energy transition research is novel. For implementation, we follow [57,58], embedding documents, reducing dimensionality with UMAP, and clustering them into topics using HDBSCAN. Hyperparameters were tuned experimentally to balance topic coherence and granularity, leading to a refined set of 21 topics. As shown later in Section 4.2, despite methodological differences, BERTopic supports the DTM findings. It identifies a similar topic mix while providing additional insight into the semantic relationships between topics.

Second, validation of NER results requires reviewing extracted entities in context to confirm that they are correctly recognized and classified. We manually checked for false positives and false negatives (missed entities). Third, for sentiment analysis, validation included sampling texts with assigned polarity scores to verify that the sentiment aligns with human interpretation. We have also compared sentiment distributions across topics and entities to check for systematic biases. We have abstained from re-calibrating thresholds or adapting lexicons to ensure replicability and transparency.

Lastly, we conduct manual validation using MAXQDA as an exploratory tool for qualitative content analysis of media discourses on carbon management technologies. The initial keyword list, developed from existing literature (Section 2.2), was used to retrieve relevant articles. We then applied thematic coding to systematically categorize text segments by technology type, actor references, and framing dimensions. Each coded segment was linked to metadata such as publication date and news outlet (see Table B3 in Appendix B), allowing us to cross-check keyword retrieval with broader discursive patterns in the corpus.

4. Results

The findings are presented in a chronological way that reflects the sequential research pipeline. In 4.1, we explore the discursive topics and temporal developments within the South African coal debate. Section 4.2 dives deeper into the content of the transition topic cluster, a result of the topic model in 4.1, and its interconnectedness with the other topics. 4.3 delves into how technology is portrayed as a key enabler of the transition. We then identify and analyze the key actors engaged in the coal discourse in 4.4, and combine these insights with the topic modeling results. Finally, in 4.5, we examine the tones in which coal-related themes are discussed, disaggregated by actor and topic cluster.

4.1. Topics and time trends

Our analysis of the DTM output reveals twenty distinct topics.[8] To enhance the interpretability of our results, we classify the labeled topics into five different topic clusters: *Mining industry*, *Transition politics*, *Energy crisis*, *mining affected communities*, and *Politics in mining & energy*. To do so, we apply an inductive process, starting by associating each individual topic with a tentative category based on its most representative keywords and documents. We then refine and consolidate these initial associations until stable and coherent higher-level clusters emerge. We refer to them as topic clusters, as visualized in Fig. 3 in its temporal dimension and using a rolling 3-month average to smooth out short-term fluctuations in topic shares.[9]

The topic clusters *Politics in mining & energy* and *Mining industry* dominate the discourse, maintaining the largest cluster share up until around 2016–2017. After this period, the prominence of *Mining industry* begins to decline, with its share being gradually overtaken by other topic clusters. Particularly, *Transition politics* increases in relative importance between 2019 and 2022. To understand the underlying dynamics explaining these general trends, we need to take a more granular look at the individual topics within each topic cluster. Fig. 4 disaggregates the topic clusters and depicts the temporal evolution of individual topics.

Fig. 4 provides a detailed overview of how individual topics have evolved over time, within and across topics of the same topic cluster. We thereby identify considerable differences in the individual development topic importance over time, with some topics displaying relatively constant topic shares and others significant variations. Topics such as *Daily political business* and *Trade relations* exhibit relatively stable importance over time, suggesting their continuous relevance in the coal discourse. These topics cover reporting on routine governmental and political activities, debates, and foreign trade politics, which seem to remain important regardless of transient events or changes in related discourse areas. Similarly, the topic *Eskom crisis* exhibits relatively constant relative shares over time. Eskom finds itself in a constant state of crisis since the 2000s [41]. In particular, *Eskom crisis* reflects the ongoing discussion about the company's financing problems and long-planned structural reforms that have never been enacted, i.e., separating generation from transmission and distribution [see p.168 5, in].

In contrast, other topics show significant variability in their relative importance. By examining relevant articles and conducting background research on these peaks in salience, we identify key events that might have shaped the media discourse on South African coal.[10] The discussion around *State capture & corruption* peaks during notable scandals involving key figures in the mining sector and political leadership. A prime example is the arrest of former President Jacob Zuma in 2021 for his refusal to testify before the Zondo Commission.[11] The topic *Electricity crisis* often resurfaces alongside frequent power outages and the operational challenges faced by Eskom. The electricity shortages have had far-reaching implications for the country's economy and daily life [5,6]. Despite load-shedding being a persistent issue for almost 20 years, the situation has worsened in recent years, keeping it in the spotlight.

The topics *Mining investments* and *Mining market dynamics* are highly salient, in particular in the earlier years of the study period. They contain mostly financial statements, earnings reports, and other similar content relevant to financial markets. The topic *Safety & illegal activities* predominantly covers crimes and incidents such as illegal mining and associated risks in abandoned coal mines or accidents involving coal-carrying trucks. The topic *Air pollution* focuses mostly on the issue of high levels of local air pollution caused by coal plants and mines in the Mpumalanga region (see Section 2 for context). *Anti-mining activism* is

[8] Three of these topics were excluded from the analysis as they were not directly related to the coal discourse. These three topics are ID7, ID11, and ID17, see Figure A2. They discuss sports and education (ID7), the history of Indian South Africans (ID11), and cooking and fishing (ID17). For example, there are articles about fishing close to the coal wharf, or a person being born in a coal village. We could have followed a stricter exclusion rule ex-ante, but instead opted for removing whole topics after running the topic model.

[9] Using a rolling 3-month average is necessary, as due to the low number of articles published in some periods, the relative topic shares would soar too much, complicating a comprehensive analysis of discourse trends.

[10] Again, consult Table B2 in Appendix B for representative quotes.

[11] This Commission was set up to investigate a range of corruption cases, many of which were linked to the coal mining sector and Eskom. It uncovered extensive corruption networks and cases of state capture involving high-ranking officials and business leaders.

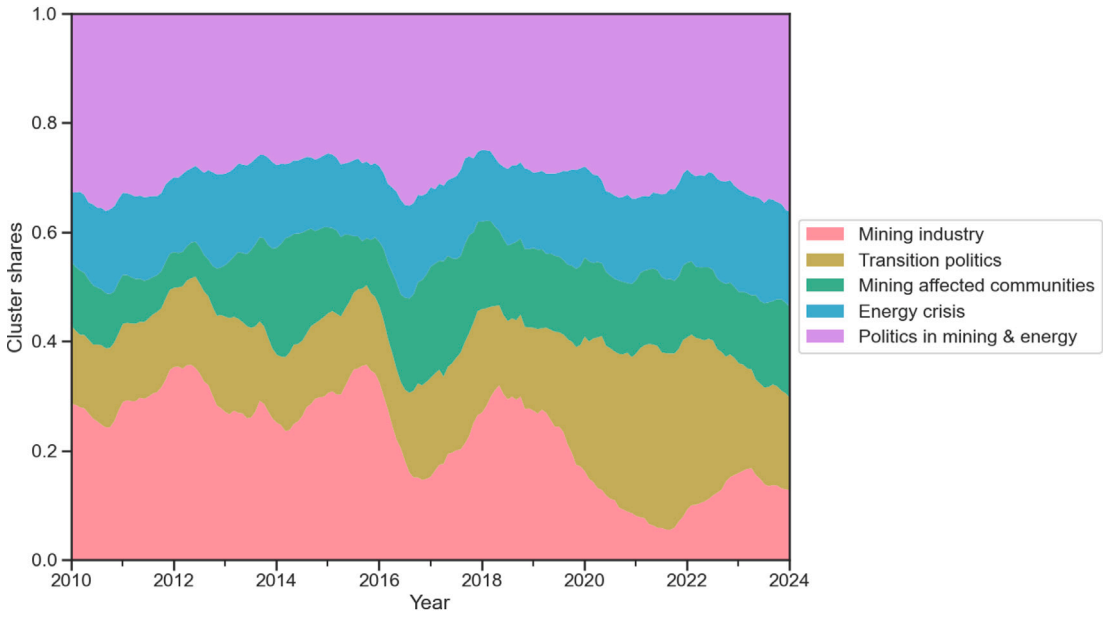


Fig. 3. Shares of topic clusters over time. The shares represent the sum of individual topic shares within each topic cluster and are smoothed over a 3-month rolling average. Refer to the legend for the topic cluster labels.

a topic repeatably peaking in importance due to the opposition of local environmental and conservation organizations against mining projects in protected areas.

Interestingly, the topic *Energy policy & renewables*, while already relatively important, experienced a significant increase starting from 2019. This rise coincides with some recent policy reforms facilitating the introduction and growth of renewable energy capacity. Notably, however, the effort in increasing renewable capacity is not primarily described as a replacement of coal as the major source of electricity generation, in line with [36]. Rather, next to their role in climate mitigation, renewables are discussed regarding their ability to take pressure from the grid amidst the electricity crisis, for instance, offering off-grid solutions.

4.2. Transition

This Section dives deeper into the transition topic, revealing its topical scope and later giving insights into its embeddedness vis-à-vis associated topics via heatmaps. The topic *Transition politics & climate finance* appears prominently in connection by the announcement of the JETP in November 2021 (again, see Section 2 for context), as the topic received a notable surge in relative attention which only lasted for a bit more than a year. The association between the JETP announcement and the observed increase in prominence is supported by an examination of representative articles and the emergence of JETP-related high-scoring words within the topic for these years, e.g., "support", "plan".[12] Particularly in the beginning, the announcement was mostly perceived positively by the media, highlighting the justice consideration behind the JETP, the opportunities for South Africa's energy transition, and its role as a potential blueprint for future climate cooperation across other countries. However, skepticism soon emerged around the yet-to-be-determined terms of the deal. There were concerns that South Africa might be forced into an unfavorable agreement by the Global North partners, potentially harming workers and the country's growth trajectory and setting energy security at risk (compare representative quotes in Table B2 in the Appendix B).

Generally, looking at the topic cluster of *Transitions politics*, the data shows that it is strongly driven by the topic *Energy policy & renewables*, while *Transition politics and finance* – except for the spike in 2021/2022 – is less dominant in this regard. Furthermore, the topic *Climate change impact*, which primarily discusses extreme weather events like droughts or floodings and their consequences seems to play a constant but minor role.

Importantly, we were unable to identify topics directly connected to the phase-out of coal and, for example, the associated impacts on workers and communities. In fact, besides the topic on *Transition politics & climate finance* (see respective high-scoring words of topic 1 in Table B1 in Appendix B), the model does not identify coal phase-out as a topic of its own. The link between coal transition and the issues of coal labor and energy policy & renewables is apparent. However, considering the respective high-scoring topic words (compare topics 5 and 13 in Table B1) along with a qualitative in-depth examination of the most representative newspaper articles on these topics, reveals the absence of discourse on expected coal phase-outs.

Further, the topic of *Coal labor issues* is primarily focused on unions and working conditions within the mining and energy sectors. This encompasses discussions on strikes and wage negotiations. It lacks specific references to the broader implications of coal plant closures or strategies for managing worker transitions in the context of a shift away from coal.

The topic of *Transition politics & climate finance* includes some discussion on phasing out coal, but these references are generally broad and are often framed within the context of international events such as the annual COP summits. The most specific discourse within this topic pertains to the JETP announcement at COP26, which could have been – for a short period of time – a significant driver of media attention within the general coal discourse.

Similarly, the topic of *Energy policy & renewables* focuses more on policy initiatives and the growing renewable energy infrastructure [see also 61]. While it highlights important policy reforms and programs aimed at increasing renewable energy capacity, it generally does not directly address the phase-out of existing coal infrastructure. This is in

[12] DTM can capture intra-topic changes in language. Over time, as specific events like the JETP announcement occur, the language and focus within a topic can shift, reflecting new developments and priorities in the discourse.

Fig. 4. Individual topic share development over time. The colors indicate the topic cluster the individual topic is assigned (compare with Fig. 3). Topic shares are smoothed over a 3-month rolling average. The cluster *Politics in mining & energy* (in purple) encompasses the topics *State capture & corruption*, *State intervention in mining*, *Daily political business*, and *Trade relations*, the topic cluster *Energy crisis* (in blue) contains *Eskom crisis*, *Oil & gas alternatives*, and *Electricity crisis*. Moreover, the cluster *Mining affected communities* (in green) comprehends the individual topics of *Safety & illegal activities*, *Coal labor issues*, *Air pollution*, and *Anti-mining activism*. Finally, the *Transition politics* cluster (in yellow) groups the topics *Transition politics & climate finance*, *Energy policies & renewables*, and *Climate change impact*; while the cluster *Mining industry* (in red) contains the topics *Coal transport*, *Mining market dynamics*, and *Mining investments*. Refer to Table B2 in Appendix B for representative quotes that offer further insights into the content and scope of each topic discussed.

line with [36]'s findings on resistance and lock-in. Rather, the ongoing electricity crisis and the need for more renewable capacity in the electricity mix to improve energy security are discussed.

We now proceed to examine the embeddedness of the transition discourse within the broader maps topic landscape. DTM enables us to identify the main topics and trends in the South African coal discourse. It highlights that concrete coal phase-out topics receive limited attention. As energy transition extends beyond the mere phase-out of coal, encompassing extensive social, economic, and political transformations, we are interested in the embeddedness of transition topics in

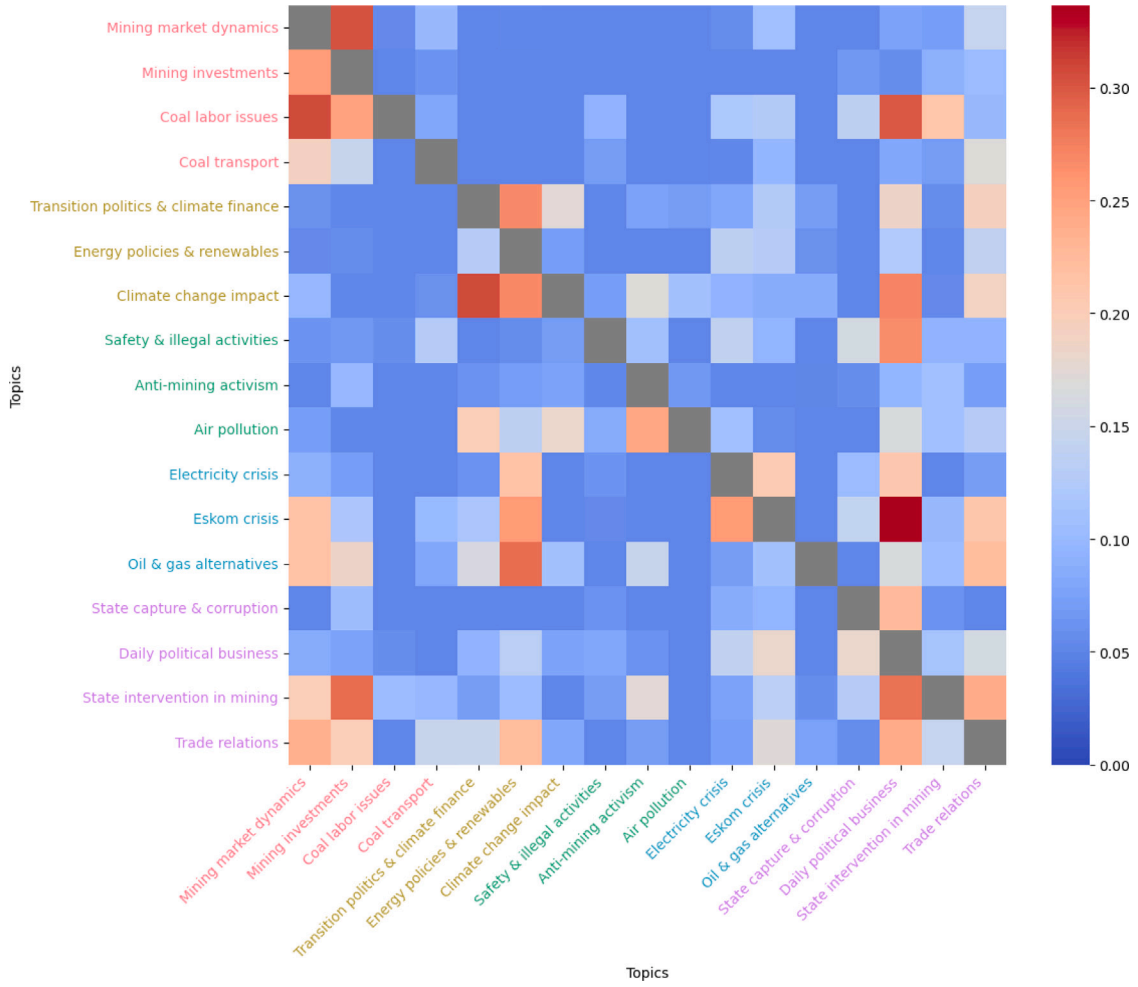


Fig. 5. Semantic similarity between topics identified via DTM and grouped by topic cluster, indicating distributional co-occurrences. The values indicate how much a topic on the vertical axis co-occurs with a different topic on the horizontal axis, evaluated over its share of newspaper articles. The gradient from blue to red visually highlights topics with weaker to stronger similarities.

the overall topic landscape of the coal discourse. Hence, we calculate co-occurrence matrices visualized as a heatmap below (see Fig. 5).

We benchmark BERTopic against DTM to visualize inter-topic relations, as shown in Figure B1 in Appendix B. BERTopic identifies a very similar topic mix, with topics involving the mining industry, politics, and topics regarding the energy crisis, all of them being prominently featured in the coal discourse.

Importantly, we see how the topic of *Transition politics & climate finance* (in green color and encircled) is embedded within the overall discourse landscape on coal. The topic is located in close proximity to the topic *Renewables & nuclear energy*[13] (dark green) with a smooth transition between the two clusters of articles. Furthermore, it is strongly related to the relatively small topic *Air pollution* and, to a lesser extent, to the topic *Oil and gas*, highlighting their significance – of gas in particular – in the energy transition. Overall, the transition (*Transition politics & climate finance*) seems to be discussed mainly in relation to a technological perspective, putting a focus on the growth of low-carbon electricity production, while showing little proximity to labor-focused issues (in purple) (see 4.3 separately on the role of carbon management technology). Confirming the DTM, the BERTopic model does not detect topics explicitly related to coal phase-out (see Figure B1 in Appendix B).

We now turn to the strength of semantic connections between transition topics and the broader set of themes. Fig. 5 presents a heatmap of topic similarities, where, for example, roughly 33 percent of articles containing the topic Eskom crisis (vertical axis) also include the topic Daily political business (dark red area). While such overlaps illustrate general inter-topic relations, our primary focus lies on the embeddedness of the transition topic within this wider landscape. Indeed, *Transition politics & climate finance* strongly co-occurs with *Energy policies & renewables*. Other notable, though somewhat weaker, co-occurrences of the transition cluster exist with *Climate change impact*, *Daily political business*, and *Trade relations*. The strong co-occurrence with *Climate change damages* is further highlighted by its role as the most frequently co-occurring topic in articles discussing *Climate change damages* (*Climate change impact* on the vertical axis and *Transition politics & climate finance* on the horizontal axis. *Transition politics & climate finance* seems to be discussed mainly with a technological perspective, focusing on the expansion of (low-carbon) electricity production (again, see 4.3 below). The comparatively large topic *Energy policies & renewables* itself is furthermore closely linked to the topics of South Africa's energy crisis, which supports the findings in Section 4.1 regarding the dual role of renewable energy for climate change mitigation and energy security.

Transition politics & climate finance also frequently co-occurs with *Daily political business* and *Trade relations*. However, these latter two topics tend to overlap broadly with most other topics. While *Climate change impact* is closely connected to the transition discourse, *Air*

[13] Renewables & nuclear energy corresponds closely to the topic *Energy policy & renewables identified* via DTM.

pollution appears less related. This may be because *Air pollution* mainly addresses local pollution from mining activities, whereas, as noted earlier, the transition discourse is generally not linked to mining-related discussions.

On a theoretical-conceptual level, strong links can be made between transition politics and workers' safety and anti-mining activism by addressing the hazardous conditions and health risks workers face in extractive industries, especially when mining activities are illegal. These perspectives could be central to the transition discourses. Additionally, air pollution concerns from mining activities would align with local environmental with global climate change concerns (compare with results found in [6] based on environmental and labor movements' voices). Our results, however, show that the transition discourse fails to make these conceptual connections, as indicated by a very low co-occurrence with the topics that belong to *Mining affected communities* (in green).

Summing up, the coal discourse in South Africa seems to be dominated by topics involving the mining industry, the relationship between politics with the mining and energy sectors, and the energy crisis. While renewable energies seem to be gaining traction in the discourse, we did not identify a clear shift to topics involving the concrete phase-out of coal usage. The only topic explicitly containing reference to an energy transition away from coal is predominantly connected to the JETP agreement. However, it is important to note that we are discussing the relative importance of topics within the coal discourse – our primary focus in this context – rather than the absolute level of attention to coal-related topics or their relative share within the entire South African news landscape. We discuss such limitations in Section 5.

4.3. Technology

While the political economy literature on South Africa's energy transition explores the strategies for maintaining the hegemony of coal, little has been said about the specific role of carbon management technologies. [5] for instance mention the potential future role of hydrogen while questioning its commercial viability (p.168). Despite increasing discourse on decarbonization and clean energy, coal companies remain dependent on fossil fuel-based business models. Prior research has shown how rhetoric prevails over concrete action [62] and how a focus on technological innovation delays more transformative policies, sometimes labeled as 'fossil fuel solutionism' [45]. We identify a coherent picture of the continued importance of coal in South Africa, sustained by topics 4.1 and their embeddedness 4.2. In line with these findings, this subsection examines how coal continues to co-exist with the promotion of carbon management technologies that some label "false" [46], featured in the transition cluster. This dynamic reflects a crowding-in effect, with these technologies being pivotal to different actors' climate communication strategies [47] while partly being technologically immature or economically unproven. We lay out the discourse representation of four such technologies: carbon capture and storage, clean coal, offsets, and hydrogen. we hypothesize that carbon management technologies appear in the energy transition discourse and are framed – by industry and government voices – in ways that sustain fossil fuel dependence. Our analysis reveals five distinct discourse categories through which these technologies are framed: investment and finance decisions, policy support, academic collaboration, doublespeak, external criticism, and narrative promotion (compare the first column in Table B3 in Appendix B). These categories emerge inductively from the news article content, offering a structured view of how potentially contested technologies are legitimated in South Africa's energy transition. The investment and finance discourse positions these technologies as commercially viable and forward-looking. Eskom's deployment of retrofitting measures for carbon capture and storage, Sasol's planned shift from gray to green hydrogen, and the launch of a flagship hydrogen-based aviation fuel project with support from President Ramaphosa all speak to this framing (compare the second column in Table B3 for the precise statements). Similarly, Standard Bank's openness to financing clean coal facilities reflects how private capital is mobilized. In parallel, the policy support category highlights how government actors actively facilitate such projects through institutional endorsement and regulatory leeway. Examples put forward are the ANC's inclusion of clean coal in planning frameworks and the Department of Environmental Affairs' use of biodiversity offsets to accommodate polluting infrastructure. The discourse of academic collaboration is exemplified by Exxaro's funding of university research on carbon capture, and Nelson Mandela University's innovation in converting coal dust into clean coal demonstrates how knowledge partnerships are leveraged to frame fossil-aligned solutions as research-driven and sustainable. These alliances underscore the function of academic credibility in validating technologies that might otherwise be seen as obstructionism [63]. In contrast, the doublespeak and criticism discourse captures Sasol's internal acknowledgment that carbon capture and storage lacks the potential to meaningfully reduce emissions, thus contradicting its broader public messaging. NGOs such as GroundWork and Greenpeace reject both offsets and the framing of coal as potentially "clean", thus attempting to disrupt dominant discourses. Finally, the narrative promotion category captures the role of politicians and lobbyists in actively constructing and amplifying pro-technology discourse. Political leaders have publicly backed carbon capture demonstrations, while consultants and industry representatives celebrate advancements in clean coal.

Together, these emergent categories reflect how technologies positioned as climate solutions are discursively framed to align with industrial, financial, and political agendas, hence confirming our initial hypothesis. While corporate and state actors use these discourses to institutionalize potentially contested technologies, critical voices remain active in exposing contradictions and redirecting attention to their environmental and social implications. These findings contribute to broader understandings of fossil-fuel solutionism and the evolving political economy of energy transitions.

A comprehensive analysis of how these technologies are framed as business and green growth opportunities within energy transitions is beyond the scope of this paper. We expect academic research on fossil-fuel solutionism to rapidly expand in the future, especially when linked to fossil fuel incumbents' discourses around coal.

4.4. Key actors

Understanding the relative importance of actors is essential for grasping who influences the various coal-related topics. To address this gap, we use automated NER to systematically identify key actors within the South African coal discourse (see 3.2 for methodological details). We group the actors identified by NER into the following categories: the state-owned electricity provider Eskom and its representatives, private companies primarily active in the mining sector, political actors (notably government officials and members of the largest party), companies in sectors other than mining, unions and their representatives, other societal actors (especially environmental organizations), and international (political) actors. Fig. 6 depicts the top five actors of each category in histogram format, where the y-axis indicates the number of sentences that mention a specific actor.

Eskom (in petrol) and mining companies (in pink) are the predominant actors of the discourse, followed by political actors (in orange).[14] Eskom's overwhelming prevalence aligns with its connection to the majority of topics identified in the topic model. The most frequent actors in the mining business are the leading companies involved in coal mining in South Africa, i.e., Glencore, Exxaro, Sasol, Optimum,

[14] Also refer to Figure B3 in Appendix B for an aggregate visualization and ranking of actor categories.

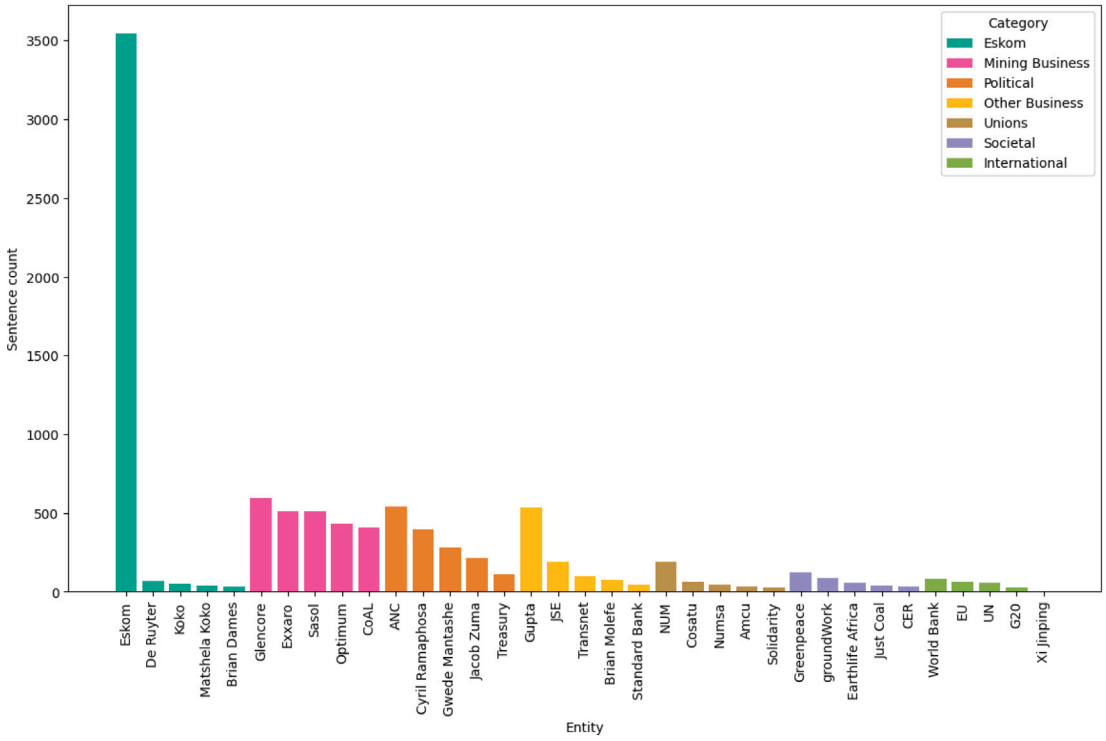


Fig. 6. Top five actors grouped by actor category as detected via NER. The legend in the top right corner indicates the colors corresponding the respective actor categories. De Ruyter, Matshela Koko, and Brian Dames are all current or former Eskom officials. Abbreviations are listed in the order of their appearance (from left to right): CoAL (Coal Africa Limited), ANC (African National Congress), JSE (Johanissburg Stock Exchange), NUM (National Union of Mineworkers), Cosatu (Congress of South African Trade Unions), Numsa (National Union of Metalworkers of South Africa), Acmu (Association of Mineworkers and Construction Union), CER (Centre for Environmental Rights), EU (European Union), UN (United Nations).

CoAl. Looking at the political actors, our results highlight the particular importance of the African National Congress (ANC) which has been the long-time ruling party until the 2024 election, with close ties to the state-owned Eskom company [5], as well as South Africa's current President, Cyril Ramaphosa, and the ex Minister of Resources and Energy, Gwede Mantashe.[15]

The category Other business groups together companies adjacent to the coal sector, such as the Johannesburg Stock Exchange or the national train provider Transnet. The most prominently identified actor in this category is the Gupta family. This influential Indian business family has been involved in a series of corruption cases within the coal mining industry and in cases of state capture, with close ties to former President Jacob Zuma (compare topic *State capture & corruption* in 4.1). The example of the Gupta family illustrates the profound intersection between private business interests and political power. The prevalence of union actors is comparatively low (in brown), with the National Union of Mineworkers (NUM) showing the highest relative importance in the discourse, likely due to its status as the country's largest union for coal workers. Taken together, civil society organizations (in blue) and international actors (in light green) are identified as the least dominant actor groups in the coal discourse. Non-governmental organizations, such as Greenpeace, and international entities, such as the World Bank, the European Union, and the United Nations are part of the data corpus, but their role is marginal.

NER analysis reveals that established energy and mining players dominate the coal discourse, leaving civil society largely sidelined. Eskom and major coal companies take center stage — mirroring our topic clusters (see 4.1) - and ultimately, the primary beneficiaries of coal usage overshadow niche voices like labor unions and environmental organizations.

Finally, inspired by [53], we link actors and topics to provide a clearer picture of actor involvement in specific topic clusters of the discourse.[16] Specifically, we link the actors on a sentence level to the most dominant topic of each article identified via dynamic topic modeling. Fig. 7 shows the normalized shares of actor categories over the topic clusters identified in 4.1. The width of the bars corresponds to the average topic trend shares (compare to Fig. 3). The stacked bar shows heterogeneous roles across actor categories and quasi-monopolies of specific actor groups, e.g. of Eskom in the energy crisis topic cluster (blank in blue bar), and of the mining business in the mining industry cluster (dots in red bar). These two observations are consistent with the results above.

Notably, our findings indicate that Eskom plays a significant role in transition politics (olive bar). Indeed, Eskom is both a target of transition policies, including the JETP and renewable energy policies, and an active player with its own interests in the energy transition. Actor representation in the transition cluster is uneven, which may influence the social acceptability and equity outcomes of a fossil fuel transition. Likewise, societal actors are minimally represented in clusters directly relevant to them, such as those concerning affected mining communities and transition politics (diagonal stripes in green and olive bars, respectively).

[15] Following the 2024 elections, the Ministry was split, with Mantashe retaining responsibility for Mineral Resources while the Energy portfolio was reassigned. Mantashe is considered to be a strong supporter of the coal mining industry within the government.

[16] [53] offer a novel and scalable pipeline for discourse analysis that leverages several NLP methodologies, including topics-entities linkages. As of today, robust mixed-methods approaches featuring an sequential process of NLP methodologies are actively developed by scholars to push the boundaries of research in computational social sciences.

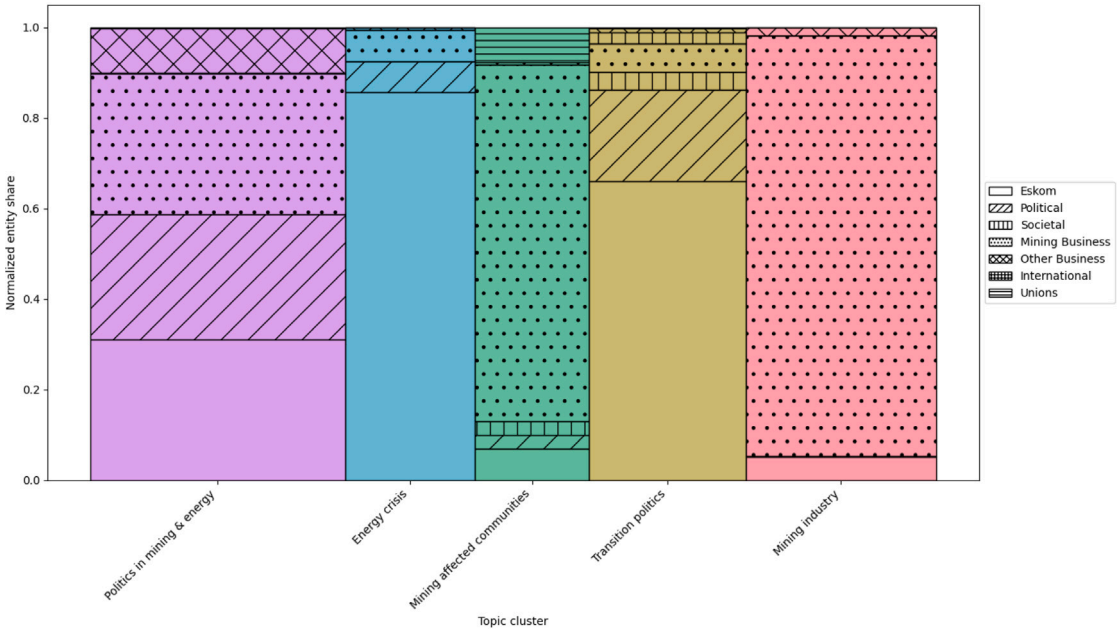


Fig. 7. Stacked bar chart depicting actor category shares per topic cluster, normalized to 1. The width of the topic cluster bars corresponds to the overall cluster share in the corpus. The colors of the bars correspond to the colors of the topic clusters in Figs. 3 and 4. The patterns correspond to the different actor categories, e.g., dots for the mining business (see legend on the right).

4.5. Tones

Next, we turn to the tones that surround the different actors and topics, which can influence public opinion and policymaking, and gives more depth to the previous analysis. In fact, its understanding is essential for gauging the potential fragmentation of attitudes across actor groups and topic clusters. We use sentiment analysis (see 3.2 for methodological details) to gauge the average sentiment by actor (Fig. 8(a)) and by topic cluster (Fig. 8(b)). The two visualizations collectively highlight pronounced heterogeneity in sentiment. The actor-focused sentiment graph (a) shows that within the Mining Business category, articles carry the most positive connotations, consistent with the topic-focused graph (b), where the *Mining industry* cluster exhibits the highest average sentiment. This correlation suggests a favorable discourse around mining activities, potentially reflecting positive economic or operational aspects associated with the sector.

The *Transition politics* cluster, which is represented with a moderately positive sentiment in the topic cluster graph, is linked to the Political entity, which also displays a balanced sentiment in the entity-specific chart. This suggests that political discourse surrounding energy and mining transitions is generally perceived in a constructive manner, although it remains subject to variability in tone.

The energy crisis and mining-affected communities exhibit low average sentiment scores in the topic graph, which parallels the relatively negative sentiment observed in the Eskom and Societal actor categories. This reflects a more general sense of discontent or critical discourse related to systemic issues affecting energy provision and societal impacts. These interlinked results demonstrate the existence of a complex web of sentiment, whereby specific entities and topics (e.g., *Mining Business* and the Mining industry are perceived favorably, whereas others (e.g., Eskom, societal impacts, and the energy crisis) are characterized by a more critical framing. This dynamic illustrates the convergence of industry performance, political discourse, and societal implications in influencing public sentiment.

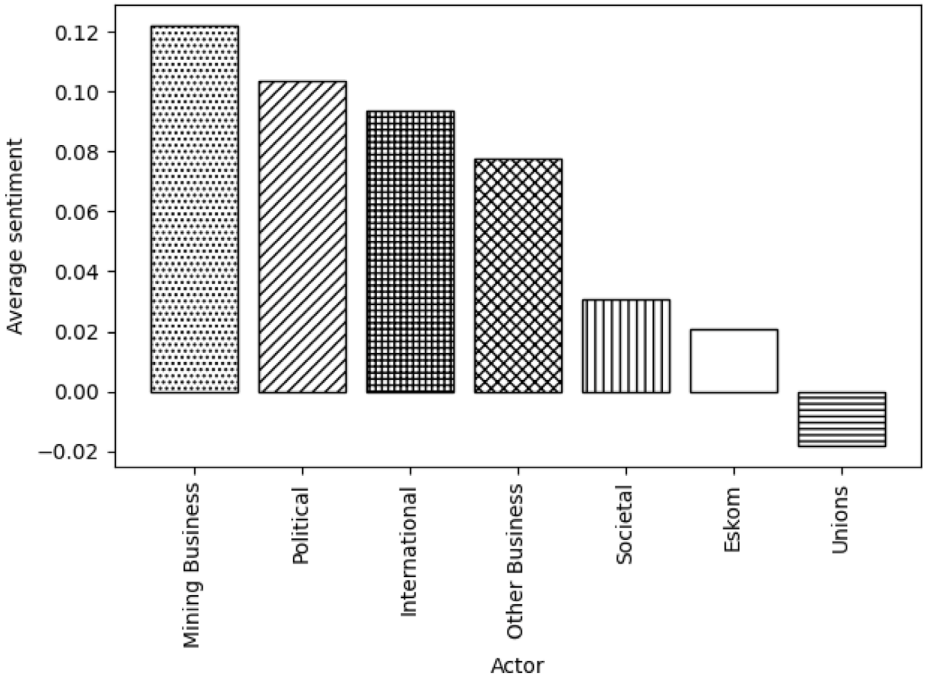
5. Discussion

This paper examines the dynamics of coal-related discourse within South African newspapers, with a particular emphasis on discourse surrounding the energy transition. It introduces and leverages a novel methodological approach that combines text-as-data analysis techniques, including dynamic topic modeling, named entity recognition, sentiment analysis, and keyword search to identify and analyze underlying patterns and trends. We see our methodological approach as a means to our empirical contribution with high policy relevance.

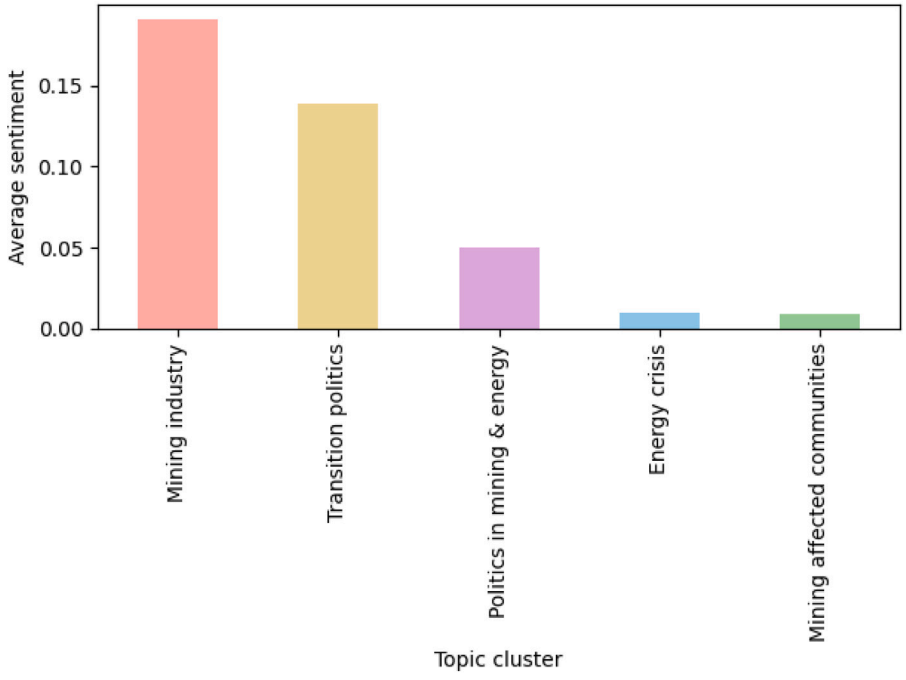
This approach allows us to uncover the key clusters within South Africa's public discourse on coal and track their evolution over time. It also reveals how those clusters are represented in the news and which actors are mentioned in relation to specific topics. Based on this, we can explore how coal is positioned within public discourse, assess the current state of the energy transition, and examine the influence of international climate policy. Judging from the evolution of public discourse, South Africa's coal mining sector does not lose political significance. Rather, we identify an ongoing linkage between South Africa's coal mining sector and politics. Our analysis of media coverage underscores how discourse not only reflects but actively sustains the entanglement between political decision-makers and the coal sector, thereby reaffirming and extending earlier findings [38,39].

The media discourse around coal in South Africa is dominated by incumbents such as Eskom and major coal mining companies, which hold close connections to political elites. This might be a pertinent reflection of vested interests in the continuation of coal use [14,38,39,64]. In contrast, trade unions, environmental groups, and international actors have a small presence in the discourse. Even within the transition politics cluster, where one might expect a broader range of voices strongly advocating – or opposing – the transition, these groups' importance is minor, as reflected by the media. This finding contrasts with previous research, describing their influential role, in particular of unions, in coal and the energy transition [e.g., 5,42]. Our findings suggest that the media landscape reflects and reinforces the power of incumbents, while marginalizing niche actors whose views challenge the dominant coal narrative, thereby limiting the visibility and perceived legitimacy of alternative transition visions.

It is hence less surprising, looking at the results of our sentiment analysis, that the mining industry and transition politics are framed in positive terms overall. We find two major explanations for this. First, coal is framed to play a pivotal role in solving South Africa's



(a) Actors



(b) Topic clusters

Fig. 8. Normalized histogram of sentiments expressed as compound scores, averaged, for (a) actors and (b) topic clusters. Patterns in (a) and colors in (b) are in line with the representation in Fig. 7 above. Each snippet is assigned a compound score (–1 = most negative, +1 = most positive), and scores are then averaged to yield the average sentiment.

ongoing structural challenges in the energy sector and the persistent electricity capacity shortfall [5,34]. Second, Eskom and other incumbents advance discourses of carbon management technologies, including hydrogen, clean coal, offsets, and carbon capture and storage, which function not only as proposed climate strategies but also as powerful narratives that justify the continued reliance on coal [46]. While Eskom's case is shaped by its specific context, it shares patterns with other coal-dependent economies, where energy transition discourses are aligned with incumbent interests [49]. In the case of South Africa, this framework demonstrates how the potential for systemic disruption and destabilization of fossil-fuel-based regimes is neutralized [compare also 50].

Indeed, the energy transition in South Africa is not discussed through a lens of phasing out coal in national media. We find no evidence of a decisive shift toward topics addressing the concrete phase-out of coal. The only topic explicitly referencing an energy transition away from

coal is primarily linked to international climate summits and the 2021 JETP agreement. This topic experienced a brief but notable spike in prominence following the announcement of the agreement. On the one hand, this lack of focus on coal phase-out is unexpected, particularly in light of the extensive body of literature on South Africa's energy transition that explicitly addresses the coal phase-out [e.g., 5,18,19,40]. On the other hand, it confirms how elite power succeeded in maintaining the legitimacy of carbon-intensive systems [10], suggesting that coal and renewables are not portrayed as competing paths. At the same time, debates over contested technologies in South Africa's energy transition also fail to center the issue of coal phase-out.

By contrast, we identify a strong association between the transition cluster and renewable energy, which is framed not only as essential for a low-carbon transition but also as a means to diversify the energy mix and strengthen energy security. Despite the growing focus on renewables, the future role of coal remains largely unquestioned, with business-as-usual narratives prevailing over explicit coal phase-out discourse. This suggests that renewables and coal are not framed as direct competitors. The same finding applies to carbon management technologies. CCS, for instance, explicitly prolongs fossil fuel use, thereby justifying continued extraction and combustion as part of the mainstream transition narrative.

Lastly, any connection between labor issues and transition-related topics is absent. This is surprising, as job security is consistently highlighted as a central driver of South Africa's just energy transition in the literature [e.g., 19,65]. This disconnect likely stems from the lack of explicit discourse on phasing out coal, as attention to the potential implications for workers and communities directly affected by the transition may also be limited. Opposition to coal mining and, to a lesser extent, coal combustion, is predominantly rooted in localized concerns, such as air pollution and environmental harm (e.g., risks to groundwater and mining activities in protected areas). These concerns are not framed as part of a broader call for a low-carbon energy transition featured in national news, a phenomenon also known from mining-affected communities and the local focus of their opposition to mining [66].

We acknowledge several limitations of our analysis. A primary limitation is the restriction to English-language articles, which account for 99.2% of LexisNexis.[17] This constraint may limit the validity of the findings on the overall South African public coal discourse, potentially skewing the results toward the perspectives of political and economic elites while underrepresenting regional and societal voices. Similarly, the exclusion of regional and local newspapers from the analysis might bias our results considering the highly localized nature of coal-related issues such as labor conflicts, pollution, or community protest. These topics may be less visible in national outlets. This limitation could partly explain the observed absence of discourse around coal phase-out. Our analytical approach should hence be extended to regional levels in future work, provided more comprehensive and balanced data is available. Second, as highlighted in 4.1, this paper mainly focuses on the relative importance of topics within the coal discourse, rather than on the absolute level of media attention to coal. The significant variability in monthly articles collected by LexisNexis (compare Fig. 1 in 3.1) deserves further investigation to determine whether this variability over time reflects an overall change in attention to coal-related topics.[18] Third, this paper faces limitations concerning the potential lack of comprehensive coverage across all newspapers in LexisNexis and potential biases toward specific outlets, e.g., more print than online media [67]. This could partly explain the prevalence of business outlets, especially in the first years of the study period. In this regard, [68] stress the advantages of a broader data approach that incorporates publications from NGOs, social movements, corporations, and governmental bodies, whereas our methodology is more structured and offers greater transparency.

A general limitation of topic models is their restricted ability to capture context and nuances, sometimes leading to oversimplified interpretations [12]. For actor identification, using NER we cannot differentiate between the semantic roles of actors. For instance, Eskom might be portrayed as an active agent driving a discourse or as the target of a policy. Future research could enhance NER with semantic role labeling to better capture such nuances, an approach that has yet to be implemented in the energy transition scholarship. For sentiment analysis, rule-based models like VADER face challenges with nuanced language, such as negation or sarcasm, which, however, are more common in social media data than in news articles. Focusing specifically on text segments around the word *coal*, we argue that our approach can enhance sentiment analysis by reducing unnecessary noise.

Related, results aligned with existing knowledge can be perceived as validating, while counterintuitive findings can be dismissed as modeling inadequacies [12,22]. Despite efforts to mitigate these limitations by complementing quantitative approaches with manual validation of the models, some degree of confirmation bias will persist. In line with this, [68] stress how "expert" knowledge remains indispensable for interpreting problems and understanding the contextual nuances of a discourse, hence pointing to the importance of contextualization [see also 69]

6. Conclusion

This paper employs – and explores the possibilities of – a combination of text-as-data methods on a large set of South African newspaper articles published between 2010 and 2024 to analyze the dynamics and scope of public coal discourses. We show that South African newspaper discourse is largely shaped by clusters emphasizing coal's ongoing economic significance and the strong influence of the mining and energy sectors, and political elites. This is further evident in the overwhelming dominance of identified key players like the state-owned electricity provider Eskom and major coal mining companies. The discourse also heavily focuses on the immediate challenges regarding the energy crisis. While there has indeed been a shift in recent years toward energy transition topics, namely the expansion of renewable energy, the phase-out of coal is not actively discussed in national news. In fact, renewables are not predominantly framed as a replacement for coal, but rather discussed in the context of addressing the energy crisis, maintaining a business-as-usual stance on coal.

From a methodological perspective, this research illustrates how economics and social science studies can leverage text-as-data techniques to analyze large volumes of text, offering valuable insights into discourses, which is of crucial importance for policy sequencing and as these shape societal perceptions. We present a robust text-as-data pipeline that provides comprehensive insights into the national coal discourse, highlighting key topics and actors, and the tones expressed toward both. This exploration paints a detailed picture of how coal and transition politics are intertwined, thereby undermining efforts for a just energy transition.

This analysis underscores that media discourses are not neutral reflections of societal opinion but active sites of knowledge production, shaping narratives, actor positions, and the opening of policy windows. In this sense, the observed framing reveals not only media priorities but

[17] English is the language that is predominant not only in the media but also in government and business, and is spoken and understood in urban areas and among higher-educated population groups. However, it is not the first spoken language of the majority of South Africans.

[18] This research contributes to the literature by quantitatively mapping the coal discourse in South Africa and demonstrating the combination of various text-as-data methods to address questions in the field of energy transitions. We emphasize that this approach is descriptive, with no ambition to engage in causal claims.

also the discursive power of dominant actors in structuring transition pathways.

CRediT authorship contribution statement

Charlotte S. Bez: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Giacomo Raederscheidt:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Jan C. Steckel:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.erss.2025.104398.

Data availability

Data will be made available on request.

References

[1] Global Coal Mining Industry - Market Research Report, 2024. Last retrieved on 25/01/2025.
[2] Tradeimex, Coal exports statistics, 2023, Last retrieved on 25/01/2025.
[3] IEA, Sources of electricity generation, 2025, Last retrieved on 27/01/2025.
[4] UNDP, Climate promise, 2023, Last retrieved on 15/01/2025.
[5] Jonathan Hanto, Akira Schroth, Lukas Krawielicki, Pao-Yu Oei, Jesse Burton, South Africa's energy transition–unraveling its political economy, Energy Sustain. Dev. 69 (2022) 164–178.
[6] Almut Mohr, Transition or not? Assessing competing narratives in South Africa's just energy transition, Energy Res. Soc. Sci. 125 (2025) 104099.
[7] Michael Pahle, Dallas Burtraw, Christian Flachsland, Nina Kelsey, Eric Biber, Jonas Meckling, Ottmar Edenhofer, John Zysman, Sequencing to ratchet up climate policy stringency, Nat. Clim. Chang. 8 (10) (2018) 861–867.
[8] Geoffroy Dolphin, Michael Pahle, Dallas Burtraw, Mirjam Kosch, A net-zero target compels a backward induction approach to climate policy, Nat. Clim. Chang. 13 (10) (2023) 1033–1041.
[9] Michael Jakob, Christian Flachsland, Jan Christoph Steckel, Johannes Urpelainen, Actors, objectives, context: A framework of the political economy of energy and climate policy applied to India, Indonesia, and Vietnam, Energy Res. Soc. Sci. 70 (2020) 101775.
[10] Benjamin K. Sovacool, Marie-Claire Brisbois, Elite power in low-carbon transitions: A critical and interdisciplinary review, Energy Res. Soc. Sci. 57 (2019) 101242.
[11] Hanna-Mari Husu, Rethinking incumbency: Utilising Bourdieu's field, capital, and habitus to explain energy transitions, Energy Res. Soc. Sci. 93 (2022) 102825.
[12] Finn Müller-Hansen, Max W. Callaghan, Yuan Ting Lee, Anna Leipprand, Christian Flachsland, Jan C. Minx, Who cares about coal? Analyzing 70 years of german parliamentary debates on coal with dynamic topic modeling, Energy Res. Soc. Sci. 72 (2021) 101869.
[13] Lukáš Lehotskỳ, Filip Černoch, Jan Osička, Petr Ocelík, When climate change is missing: Media discourse on coal mining in the Czech Republic, Energy Policy 129 (2019) 774–786.
[14] Jan Osička, Jörg Kemmerzell, Maksymilian Zoll, Lukáš Lehotskỳ, Filip Černoch, Michèle Knodt, What's next for the European coal heartland? Exploring the future of coal as presented in German, Polish and Czech press, Energy Res. Soc. Sci. 61 (2020) 101316.
[15] Karoliina Isoaho, Kamilla Karhunmaa, A critical review of discursive approaches in energy transitions, Energy Policy 128 (2019) 930–942.
[16] Jochen Markard, Karoliina Isoaho, Linda Widdel, Discourses around decline: Comparing the debates on coal phase-out in the UK, Germany and Finland, in: Technologies in Decline: Socio-Technical Approaches to Discontinuation and Destabilisation, Routledge, 2023, pp. 119–144.
[17] Aleksandra Wagner, The role of media influence in shaping public energy dialogues, in: The Oxford Handbook of Energy and Society, Oxford University Press, 2018, p. 381.
[18] Pegah Mirzania, Joel A. Gordon, Nazmiye Balta-Ozkan, Ramazan Caner Sayan, Lochner Marais, Barriers to powering past coal: Implications for a just energy transition in South Africa, Energy Res. Soc. Sci. 101 (2023) 103122.
[19] Claudia Strambo, Jesse Burton, Aaron Atteridge, The End of Coal? Planning a "Just Transition" in South Africa, Stockholm Environment Institute, Stockholm, 2019.
[20] Matthew Gentzkow, Bryan Kelly, Matt Taddy, Text as data, J. Econ. Lit. 57 (3) (2019) 535–574.
[21] Sally Hewitt, Discourse analysis and public policy research, Cent. Rural. Econ. Discuss. Pap. Ser. 24 (2009) 1–16.
[22] Karoliina Isoaho, Daria Gritsenko, Eetu Mäkelä, Topic modeling and text analysis for qualitative policy research, Policy Stud. J. 49 (1) (2021) 300–324.
[23] Anna Leipprand, Christian Flachsland, Regime destabilization in energy transitions: The German debate on the future of coal, Energy Res. Soc. Sci. 40 (2018) 190–204.
[24] Ian Wash, Interpreting public policy dilemmas: Discourse analytical insights, Humanit. Soc. Sci. Commun. 7 (1) (2020) 1–12.
[25] Kimberley Vandenhole, Kristijan Garic, Philip Leifeld, When does discursive change happen? Detecting phase transitions in discourse networks of sustainability transitions, Energy Res. Soc. Sci. 122 (2025) 104020.
[26] Xinxin Wang, Kevin Lo, Just transition: A conceptual review, Energy Res. Soc. Sci. 82 (2021) 102291.
[27] Justin Grimmer, Brandon M. Stewart, Text as data: The promise and pitfalls of automatic content analysis methods for political texts, Politi. Anal. 21 (3) (2013) 267–297.
[28] David M. Blei, Probabilistic topic models, Commun. ACM 55 (4) (2012) 77–84.
[29] Paul DiMaggio, Manish Nag, David Blei, Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding, Poetics 41 (6) (2013) 570–606.
[30] Niccolò Manych, Finn Müller-Hansen, Jan Christoph Steckel, The political economy of coal across 12 countries: Analysing qualitative interviews with topic models, Energy Res. Soc. Sci. 101 (2023) 103137.
[31] Finn Müller-Hansen, Yuan Ting Lee, Max Callaghan, Slava Jankin, Jan C. Minx, The german coal debate on Twitter: Reactions to a corporate policy process, Energy Policy 169 (2022) 113178.
[32] Joris Dehler-Holland, Marvin Okoh, Dogan Keles, Assessing technology legitimacy with topic models and sentiment analysis–The case of wind power in Germany, Technol. Forecast. Soc. Change 175 (2022) 121354.
[33] Benjamin K. Sovacool, Jonn Axsen, Steve Sorrell, Promoting novelty, rigor, and style in energy social science: Towards codes of practice for appropriate methods and research design, Energy Res. Soc. Sci. 45 (2018) 12–42.
[34] Jesse Burton, Andrew Marquard, Bryce McCall, Socio-economic considerations for a Paris agreement-compatible coal transition in South Africa, Energy Res. Cent. Univ. Cape Town (2019).
[35] Etienne Nel, Lochner Marais, Zolile Mqotyana, The regional implications of just transition in the world's most coal-dependent economy: The case of mpumalanga, South Africa, Front. Sustain. Cities 4 (2023) 1059312.
[36] Marie Blanche Ting, Rob Byrne, Eskom and the rise of renewables: Regime-resistance, crisis and the strategy of incumbency in South Africa's electricity system, Energy Res. Soc. Sci. 60 (2020) 101333.
[37] Melanie Wiese, Liezl-Marié van der Westhuizen, Impact of planned power outages (load shedding) on consumers in developing countries: Evidence from South Africa, Energy Policy 187 (2024) 114033.
[38] Lucy Baker, Peter Newell, Jon Phillips, The political economy of energy transitions: the case of South Africa, New Politi. Econ. 19 (6) (2014) 791–818.
[39] Lucy Baker, Jesse Burton, Catrina Godinho, Hilton Trollip, The political economy of decarbonisation: Exploring the dynamics of South Africa's electricity sector, Energy Res. Cent. (2015) https://sro.sussex.ac.uk/id/eprint/58502.
[40] Becca Wilgosh, Alevgul H. Sorman, Iñaki Barcena, When two movements collide: Learning from labour and environmental struggles for future just transitions, Futures 137 (2022) 102903.
[41] Iain Todd, Darren McCauley, Assessing policy barriers to the energy transition in South Africa, Energy Policy 158 (2021) 112529.
[42] Tobias Kalt, Agents of transition or defenders of the status quo? Trade union strategies in green transitions, J. Ind. Relations 64 (4) (2022) 499–521.
[43] Lina Lefstad, Jonas Allesson, Henner Busch, Wim Carton, Burying problems? Imaginaries of carbon capture and storage in Scandinavia, Energy Res. Soc. Sci. 113 (2024) 103564.
[44] Lena Faber, Henner Busch, Lina Lefstad, A trojan horse for climate policy: Assessing carbon lock-ins through the carbon capture and storage-hydrogen-nexus in Europe, Energy Res. Soc. Sci. 120 (2025) 103881.
[45] William F. Lamb, Giulio Mattioli, Sebastian Levi, J. Timmons Roberts, Stuart Capstick, Felix Creutzig, Jan C. Minx, Finn Müller-Hansen, Trevor Culhane, Julia K. Steinberger, Discourses of climate delay, Glob. Sustain. 3 (2020) e17.
[46] Sarah Eckstein, Marcel Llavero-Pasquina, Freddie Daley, False solutions: How do fossil fuel companies reproduce their hegemony through the energy transition? Energy Res. Soc. Sci. (2025) (Forthcoming).

[47] Yutong Si, Dipa Desai, Diana Bozhilova, Sheila Puffer, Jennie C. Stephens, Fossil fuel companies' climate communication strategies: Industry messaging on renewables and natural gas, Energy Res. Soc. Sci. 98 (2023) 103028.
[48] Carl-Friedrich Schleussner, Gaurav Ganti, Quentin Lejeune, Biqing Zhu, Peter Pfleiderer, Ruben Prütz, Philippe Ciais, Thomas L. Frölicher, Sabine Fuss, Thomas Gasser, et al., Overconfidence in climate overshoot, Nature 634 (8033) (2024) 366–373.
[49] George Ferns, Kenneth Amaeshi, Fueling climate (in) action: How organizations engage in hegemonization to avoid transformational action on climate change, Organ. Stud. 42 (7) (2021) 1005–1029.
[50] Peter Newell, Trasformismo or transformation? The global political economy of energy transitions, Rev. Int. Politi. Econ. 26 (1) (2019) 25–48.
[51] David M. Blei, John D. Lafferty, Dynamic topic models, in: Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 113–120.
[52] Joris Dehler-Holland, Kira Schumacher, Wolf Fichtner, Topic modeling uncovers shifts in media framing of the German renewable energy act, Patterns 2 (1) (2021).
[53] Rrubaa Panchendrarajan, Geri Popova, Tony Russell-Rose, Crisis talk: analysis of the public debate around the energy crisis and cost of living, Soc. Netw. Anal. Min. 14 (1) (2024) 74.
[54] Clayton Hutto, Eric Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, 2014, pp. 216–225.
[55] Paulo Rita, Sérgio Moro, Gabriel Cavalcanti, The impact of COVID-19 on tourism: Analysis of online reviews in the airlines sector, J. Air Transp. Manag. 104 (2022) 102277.
[56] Karla Cepeda, Rajesh Jaiswal, Sentiment analysis on covid-19 vaccinations in Ireland using support vector machine, in: 2022 33rd Irish Signals and Systems Conference, ISSC, IEEE, 2022, pp. 1–6.
[57] Maarten Grootendorst, BERTopic: Neural topic modeling with a class-based TF-IDF procedure, 2022, arXiv preprint arXiv:2203.05794.
[58] Antonio Moreno-Ortiz, Making Sense of Large Social Media Corpora: Keywords, Topics, Sentiment, and Hashtags in the Coronavirus Twitter Corpus, Springer Nature, 2024.
[59] Gabriel Lindelöf, Talayeh Aledavood, Barbara Keller, et al., Dynamics of the negative discourse toward COVID-19 vaccines: topic modeling study and an annotated data set of Twitter posts, J. Med. Internet Res. 25 (1) (2023) e41319.
[60] Hajar Zankadi, Abdellah Idrissi, Najima Daoudi, Imane Hilal, Identifying learners' topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques, Educ. Inf. Technol. 28 (5) (2023) 5567–5584.
[61] Lucy Baker, Benjamin K. Sovacool, The political economy of technological capabilities and global production networks in South Africa's wind and solar photovoltaic (PV) industries, Politi. Geogr. 60 (2017) 1–12.
[62] Mei Li, Gregory Trencher, Jusen Asuka, The clean energy claims of BP, Chevron, ExxonMobil and shell: A mismatch between discourse, actions and investments, PLoS One 17 (2) (2022) e0263596.
[63] Sofia Hiltner, Emily Eaton, Noel Healy, Andrew Scerri, Jennie C. Stephens, Geoffrey Supran, Fossil fuel industry influence in higher education: A review and a research agenda, Wiley Interdiscip. Rev.: Clim. Chang. 15 (6) (2024) e904.
[64] Jan C. Steckel, Michael Jakob, The political economy of coal: Lessons learnt from 15 country case studies, World Dev. Perspect. 24 (2021) 100368.
[65] Mark Swilling, Josephine Kaviti Musango, Jeremy J. Wakeford, Developmental states and sustainability transitions: Prospects of a just transition in South Africa, J. Environ. Policy Plan. 18 (2016) 650–672.
[66] Jacklyn Cock, 'Beware of the crocodile's smile': Labour-environmentalism in the struggle to achieve a just transition in South Africa, in: The Palgrave Handbook of Environmental Labour Studies, Springer, 2021, pp. 177–197.
[67] Noah Buntain, Carol M. Liebler, Kyle Webster, Database use, database discrepancies: Implications for content analyses of news, Newsp. Res. J. (2023) 07395329231155193.
[68] Lira Luz Benites-Lazaro, Leandro Giatti, Angélica Giarolla, Topic modeling method for analyzing social actor discourses on climate change, energy and food security, Energy Res. Soc. Sci. 45 (2018) 318–330.
[69] James Goodman, Jonathan Paul Marshall, Problems of methodology and method in climate and energy research: Socialising climate change? Energy Res. Soc. Sci. 45 (2018) 1–11.