

USING ECO-LDA AS A FRAMEWORK FOR THE COMPUTATIONAL ECOLINGUISTIC ANALYSIS OF PAKISTANI MEDIA COVERAGE OF THE LAHORE SMOG (2024-25): A RECENT SNAPSHOT OF THE CRISIS

Shahid Nawaz

*MS English Research Scholar, Humanities, COMSATS University Islamabad
(Islamabad Campus), Pakistan.*

Email: shahidkhan966888@gmail.com

Abstract

Climate change is a long-lasting environmental issue in Southeast Asian countries, specifically Pakistan. This study investigates how Pakistani media has framed the Lahore Smog crisis in recent years (2024-25). In addition, the study aims to understand the ecological relationship framed by the Pakistani media, and how has the framing itself evolved over the last two years. Two major Pakistani News outlets, including Dawn and The News International, were selected for data collection. 200 articles were purposively collected for the corpus 2024, while 160 articles accounted for the 2025 corpus. Moreover, the study proposed a novel and replicable Eco-LDA framework for data analysis, which blended Latent Dirichlet Allocation (LDA), Frame Semantics, and Stibbe's (2015) Ecolinguistic framework. The Eco-LDA framework was specifically designed to identify latent frames in the corpora and assess the ecological visibility in media discourse.

Results reflect a shift in media framing in the recent years. In 2024, media framed the Lahore smog as anthropocentric and episodic issue; it highlighted immediate health risks and civic disruptions while backgrounding the root-causes of the ecological crisis. In 2025 media framing, on the other hand, industrial emissions, crops burning, climate change, and policy interventions were foregrounded. The findings reveal an increasing ecological literacy in the news media discourse in Pakistan over 2024 and 2025.

The research provides latest insights into the evolving media framing of the Lahore smog. It may assist environmental communication, climate policies, governance, and public ecological understanding. Moreover, methodologically, this research contributes to computational ecolinguistic inquiry by introducing a replicable Eco-LDA framework.

Keywords: *Eco-LDA Framework, NLP (Natural Language Processing), Frame Semantics, Media Framing, Lahore Smog, Topic Modeling, LDA, Ecolinguistics, Environmental Discourse, Climate Change, Antismog, Wellbeing, Sustainability.*

Introduction

Climate change is a universal problem impacting natural resources, food security, and environmental sustainability. Some major causes of climate change and environmental degradation include deforestation, industrial growth, and unsustainable agricultural practices (Adnan et al., 2024). In Pakistan, these activities have accelerated pollution levels, aggravated water scarcity, and intensified the loss of biodiversity across the country.

South Asia is currently experiencing a grave cross-border air pollution crisis. This climate issue stems from a rapid industrialization, deforestation, population growth, and other unsustainable agricultural practices. "South Asia is home to 9 of the world's 10 cities with the worst air pollution, which causes an estimated 2 million premature deaths across the region each year. Air pollution can also have significant effects on productivity and economic costs (Awan, 2025). Therefore, Lahore once celebrated as the "City of Gardens" has now turned as one of the most polluted cities in the world.



Figure 1.1 Source: (Swiss Air Quality Index [IQAir], 2025)

The figure above shows ranking of world's top four polluted cities as per October 2025. All those cities are from the South Asia. This recent air quality data identifies Lahore as the most polluted city globally, with an Air Quality Index (AQI) of 328, which is categorized as hazardous. Eventually, it leads to environmental crisis in Pakistan such as the Lahore Smog.

Smog is a pressing environmental issue in Pakistan; specifically, at peak in Punjab in the shape of Lahore Smog. Some major causes of the Lahore Smog include industrial emissions and crop burnings (Razzaq et al., 2024). The crisis itself is a major threat to the ecological sustainability; it negatively impacts human health, air quality, and agricultural productivity. In addition, Lahore's recurrent smog cycle is a unique "fifth season," caused by regional crop burning and unsatisfactory environmental conditions (Majeed et al., 2024). A coordinated regional climate action may reduce Lahore's smog intensity. Furthermore, according to Akbar et al. (2025), the success against the crisis new episode depends on whether these efforts shape the public narratives, influence behavior, and foster resilience, or results in cautionary story of inaction.

Mounting public awareness and encouraging public perception is key to driving environmental initiative and policy change as stated by Gorczyca and Zozulia (2025) in their article "We have the right to breathe clean air" – Mobilising communities in the fight for good air quality" as:

In the case of Krakow, a high ecological awareness of residents and its further enhancement is of great importance in implementing bold environmental solutions. The role of disseminating information about air pollution in the mass media and social media was outlined by the informants as a crucial tool for raising awareness and achieving real changes in policy and the inhabitants' behaviour. (p. 14)

One example of media framing role in shaping and altering public perceptions of the public is a research conducted by Setiadi (2025) on the framing of air pollution in *The New York Times*. His findings reflected that *The New York Times* frames air pollution in Jakarta, describing the city as "sinking under the weight of dysfunction," to highlight government negligence, while similar crises in countries like Germany are framed more neutrally. This selective framing demonstrates how media language shapes perceptions of responsibility and assigns blame differently across the regions. Moreover, a similar a study was conducted by Minnema et al. (2021), who utilized a Framenet-based semantic framework to investigate responsibility framing in Italian news reports

on femicides. Their analysis demonstrated how linguistic structures contribute to the backgrounding of perpetrator agency and the foregrounding of victim roles. The results revealed implicit biases in media discourse. This work highlights the potential of frame semantics as a methodological tool for examining socially significant language use and for advancing research in computational social science and critical discourse studies.

This research attempts to propose an integrated and replicable methodological framework which has the potential to analyze any corpus and interpret the quantitative results from an ecolinguistic lens. The framework is named as Eco-LDA framework, consisting of LDA (Latent Dirichlet Allocation) from computational linguistics, framing from frame semantics, and ecolinguistic perspective from Stibbe's (2015) *The Stories We Live By*. It operates in three steps: firstly, LDA is run to extract top keywords and suggested topics; secondly, the topics are treated as frames while the keywords act as slots of the same frame; thirdly, the frames are interpreted from ecolinguistic perspective for deeper ecological insights.

Problem Statement

Climate change and air quality degradation in South Asia have emerged as a major global environmental challenge. It has intensified ecological issues across the region. As a result, the Lahore Smog crisis has become a persistent ecological issue in Pakistan. A general observation of media coverage, during the peak of the crisis, indicates that news discourses may silence ecological relationships and normalize serious ecological crises; they may over-emphasize the short-term goals. A limited scholarly attention has been paid to investigate how media language constructs or silences ecological relationships, specifically in the context of Lahore Smog discourse. This research analyses Pakistani news coverage of the Lahore smog over 2024 and 2025 to determine how has the issue framed and evolved. The study uses a novel and replicable, an Eco-LDA, framework, which integrates Ecolinguistics, Frame Semantics, and computational NLP methods. Moreover, the study aims to determine the ecological relationship depicted in media coverage of the issue. The results may enhance the understanding of how media discourse frame and influence ecological perceptions. Additionally, it may encourage Pakistani news media to build more ecologically conscious and inclusive narratives.

Research Objectives

1. To determine the latent themes depicted in Pakistani news coverage on Lahore smog crisis.
2. To examine how do these latent themes frame the Lahore smog crisis and construct ecological relationships.
3. To investigate how has the framing of Lahore smog issue evolved over 2024 and 2025.

Research Questions

1. What latent themes are depicted in Pakistani news coverage on Lahore smog crisis?
2. How do these latent themes frame the Lahore smog crisis and construct ecological relationships?
3. How has the framing of Lahore smog issue evolved over 2024 and 2025?

Significance of the Study

This research contributes significantly to field of ecolinguistics. Specifically, it proposes a novel and replicable Eco-LDA framework. The framework integrates frame semantics, computational topic modeling, and Stibbe's (2015) ecolinguistic framework. The newly proposed framework could be used to investigate environmental crises (i.e., climate change) over a long period of time. It could also be utilized to track evolution in the framing of environmental issues

in various channels and contexts. Moreover, the results offer valuable insights for government, policy makers, climate advocacy groups, and journalists; it encourages them for a sustainable policy implementation. In addition, the research emphasizes the capability of media discourse to foster a greater ecological awareness and responsibility within a society.

Delimitation of the Study

Since there are limited facilities for an independent researcher in Pakistan, therefore, the data is collected manually only from two leading news channels such as the *Dawn* and *The News International*. Moreover, the research is limited to a latest snapshot of the Lahore smog crisis in Pakistani news media (2024-25). Yet, a novel and replicable Eco-LDA Framework is designed and tested which contributes significantly to interdisciplinary fields such as ecolinguistics, NLP (Natural Language Processing), and computational ecolinguistic inquiry.

Literature Review

This section discusses key concepts, represents the recent literature on Lahore smog, and identifies the gap in existing literature.

Frames are cognitive structures that enhance the understanding of social interaction among the people. Additionally, it also assigns meaning to happenings that would otherwise be too ambiguous to understand (Goffman, 1974). Frame analysis, on the other hand, determines how media discourse creates meaning through different associations, and how they are understood or received by the audience (David et al., 2011). According to Fruchtmann (2020), frame analysis reveals how information foregrounds some aspects of life while downplaying others; it shapes public understanding of reality. Moreover, Caiani (2023) argues that frame analysis also helps in the understanding of how actors construct frames, creates meaning (or a form of reality), and encourages collective action. Furthermore, automatically induced linguistic constructs could be used as a frame analysis tool in corpus-based frame analysis; it helps the researcher systematically identify and interpret frames within a discourse (Touileb & Salway, 2014).

Corpus analysis is a heuristic and reflexive method in discourse studies which provides a macro-level perspective on large textual data and assists researchers to extract the underlying patterns (Kutter, 2017). According to Church and Mercer (1993), frequency-based preferences, including collocations and word associations, are pivotal in language processing. Though often ignored in computational linguistics since they are hard to model with traditional parsing techniques; they can have a positive impact in language processing. In addition, they argue that entropy and evaluation measures have become increasingly significant in the recent years. This further emphasizes the growing trend of quantitative assessment in language modeling (Church & Mercer, 1993).

Topic modeling is a statistical method for identifying latent thematic structures within a corpus. A word is the basic unit of data, a document is a sequence of words, and a corpus is a collection of documents (Vayansky & Kumar, 2020). In addition, a vocabulary is the set of all unique words, while a topic is a probability distribution over the vocabulary; topics are inferred using probabilistic or generative models which reveals patterns of words co-occurrence across the corpus. Additionally, topic models are widely employed in research to determine dominant topics across the corpora (Churchill & Singh, 2022).

Natural language processing (NLP) is a fundamental area of computer science concerned with extracting meaning from textual data. However, according to Jelodar et al. (2019), among NLP techniques, topic modeling, specifically Latent Dirichlet Allocation (LDA), has been widely used

to identify latent semantic structured in large collections of unstructured documents. Moreover, LDA has previously been applied to investigate large text corpora, including business intelligence literature, artificial intelligence (AI) trends, disaster-related discourse, and public awareness of sustainable development issues (Chang et al., 2021).

Advances in Natural Language Processing (NLP) rely on effective lexical organization, which remains a key challenge. Lexical resources such as WordNet and FrameNet apply distinct theoretical frameworks hierarchical relations and frame semantics respectively, to structure lexical information for the diverse NLP applications (Boas, 2005). Moreover, frame semantics emphasizes that “to understand any one [concept], you have to understand the whole structure in which it fits” (Fillmore, 2006, p. 373). Furthermore, according to Gawron et al. (2008), classical semantics simplifies meaning by limiting it to words level, while frame semantics, in contrast, understands words within broader conceptual structures.

Lahore is identified as a serious environmental and public health problem; this issue is primarily caused by fuel and crop burning, industrial pollution, and other construction activities (Hussain et al., 2025). Employing a cross-sectional design, Qusien (2025) analyzed 365 articles on Lahore smog from six distant Pakistani newspapers. The newspapers covered two episodes of the issue such as 2017 and 2019. The findings revealed that the Urdu newspapers covered the issue more extensively than the English newspapers. The research also identified dominant frames such as disaster response, public health, and responsibility. Additionally, Dawn, The Nation, and The News International (January-March, 2023) employed structured language and narrative strategies that highlighted the government’s proactive role in addressing smog and sustainability concerns in Lahore (Gul et al., 2025). Moreover, editorials published in *Dawn* and *The News* (2016 and 2017) showed elements of solution-based journalism, major policy initiatives such as the Pakistan Climate Change Act 2017 were largely neglected (Javed et al., 2023).

Research Gap

Although previous studies have examined the causes, impacts, and media coverage of Lahore smog, most of them have relied on limited timeframe, specific newspapers, or traditional content analysis (Javed et al., 2023; Gul et al., 2025; Qusien, 2025). There is a lack of research employing computational approaches to systematically uncover latent themes, linguistic frames, and ecological relationships across consecutive years (when the crisis is at the peak). Existing studies also focus on crisis-driven or policy-oriented coverage; they overlook the evolution of ecologically informed and systematic framing. This gap is especially important as Lahore, one of South Asia’s most climate-vulnerable cities and ranked among the world’s most polluted cities in late 2025 (Hussain et al., 2025), faces severe environmental and public health challenges.

Materials and Methods

This research utilized mixed method research design while drawing on computational methods in accordance with Frame Semantics and Ecolinguistics. The nexus was innovative which was named as Eco-LDA framework. The following Figure 3.1 illustrates the data collection process for the two corpora (representing the years 2024 and 2025).

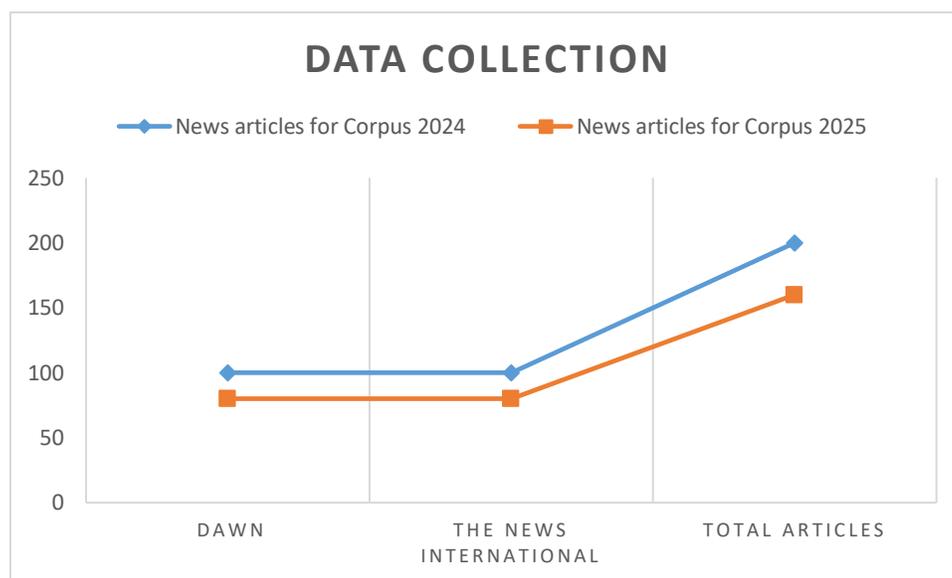


Figure 3.1 Data Collection (Corpora Construction for the Year 2024 and 2025)

Figure 3.1 represents the data collection from two leading news outlets in Pakistan, the *Dawn* and *The New Internationalist*. These two news outlets were selected purposively since their search engine worked well for manually collecting our corpus (e.g., searching through specific keywords “Lahore Smog 2024” and “Lahore Smog 2025”). Moreover, from January 2024 to December 2024, 100 articles were collected from the official website of each news outlet. In the same way, from January 2025 to November 2025, 80 articles were purposively extracted from each source. The 2024 corpus (Lahore Smog 2024.txt) consisted 200 articles, while the 2025 corpus (Lahore Smog 2025.txt) was comprised of 160 articles. The mismatch in total articles in each year is because of the purposive sampling technique; only those articles were gathered which explicitly discussed the Lahore smog crisis.

Data Preprocessing

Prior to data analysis, the corpora were preprocessed to ensure consistency and reduce noise. Operations included lowercasing, removal of URLs, punctuation, special characters, and numbers. In addition, stopword elimination, tokenization, and lemmatization were also carried out. These steps were implemented using Python 3.12 libraries including pandas for data handling, while NLTK and SpaCy for tokenization, stopword removal, and lemmatization. The resulting cleaned corpus provided a standardized dataset suitable for subsequent computational analysis and visualization.

Data Analysis

Data was analyzed through an analytical framework proposed in this article. The framework was conceptual from a nexus of Computational Grounded Theory (Nelson, 2020) in accordance with topic modeling, frame semantics, and Stibbe’s (2015) ecolinguistic framework. The framework has three steps to carry the whole research process. Moreover, Plotly and Matplotlib (Python) libraries were used for visualization of data (e.g., LDA topic ranking and frames comparison). The following figure represents the Eco-LDA Framework.

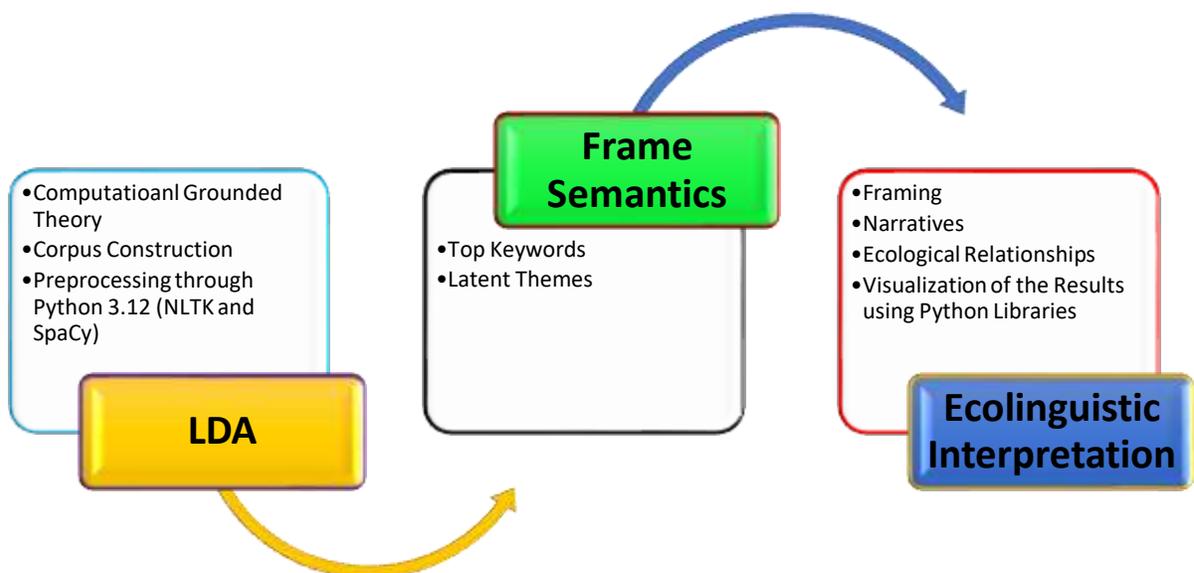


Figure 3.2 Eco-LDA Framework

Figure 3.2 illustrates the proposed framework which is a blend of Frame Semantics (Fillmore, 1976), Computational Grounded Theory (Nelson, 2020), topic modeling, and Stibbe’s ecolinguistic framework (2015). This newly introduced framework worked well in identifying frames in corpus and then interpreting them through ecolinguistic perspective. Fillmore introduced Frame Semantics, in his work *Frame Semantics and the Nature of Language (1976)*, as a theory positing that language meaning is grounded in structured cognitive frames. These frames are mental schemes that guide how individuals perceive, interpret, and communicate experience.

A ‘frame’, as the notion plays a role in the description of linguistic meanings, is a system of categories structured in accordance with some motivating context. Some words exist in order to provide access to knowledge of such frames to the participants in the communication process, and simultaneously serve to perform a categorization which takes such framing for granted. (Fillmore, 2006, p. 381)

Frame is a story about an area of life which is brought to human mind through trigger words (Stibbe, 2015). He exemplifies it as “when people hear the word ‘buy’ this triggers a transactional frame” (p. 47). Moreover, Nelson (2020) posits that computational grounded theory is a three-steps methodological framework that integrates computational techniques (e.g., machine learning and natural language processing) with qualitative interpretation to identify, refine, and validate patterns in large text datasets. Stibbe (2015) states that certain types of framing may put the short-term interests of business (e.g., self-centered or extrinsic values) ahead of environmental considerations. Furthermore, he posits that “the aim is to contribute to efforts to raise awareness of – and resist – ecologically destructive stories, and promote use of language which tell different stories that encourage people to protect the systems that life depends on” (p. 16).

Therefore, in our context, firstly, the corpus 2024 was loaded into Python 3.12 and LDA (Latent Dirichlet Allocation) was applied. Firstly, the clear corpora were loaded into Python 3.12 to extract the top keywords and latent topics. This computational method extracted the top keywords along with suggested topics (or themes). Ylä-Anttila et al. (2013) argue that LDA topics

can be treated as frames. Moreover, Ylä-Anttila et al. (2018), in their study on media coverage of climate change, have treated the LDA topics as frames. Therefore, in the similar way, in our context, since these keywords (under each topic) were interrelated, that is why the topics were treated as frames while the top keywords as the different slots in a frame. Furthermore, these frames and their respective interconnected slots were analyzed through ecolinguistic lens. Ecolinguistic interpretation helped to investigate how do the Pakistani News media frame the Lahore smog crisis and reflect ecological relationships. Finally, the findings of both corpora (2024-25) were compared to determine the evolution in the media framing of the smog crisis over 2024 and 2025.

Results

This section presents the extracted topics from two corpora on Lahore Smog crisis and their frame analysis assisted by ecolinguistic interpretation. The following table shows the suggested topics along with top ten keywords for the corpus 2024.

Table 4.1 LDA Extracted Top Keywords and Suggested Topics (Frames) in 2024 Corpus

Topic	Keyword with Weight	Weight	Rank	Suggested topics
1.	0.014*"rain" + 0.010*"march" + 0.009*"survey" + 0.007*"people" + 0.007*"province" + 0.007*"forecast" + 0.006*"expect" + 0.006*"snowfall" + 0.006*"report" + 0.006*"weather"	0.078	2	Weather Forecast and Public Impact
2.	0.016*"aqi" + 0.012*"government" + 0.012*"city" + 0.011*"level" + 0.010*"school" + 0.009*"vehicle" + 0.007*"child" + 0.007*"environmental" + 0.007*"protection" + 0.006*"day"	0.097	1	Air Pollution and Public Health
3.	0.014*"road" + 0.008*"area" + 0.008*"green" + 0.008*"district" + 0.006*"smoke" + 0.006*"lockdown" + 0.006*"public" + 0.006*"effort" + 0.005*"police" + 0.005*"impose"	0.072	3	Local Disruptions and Community Response
4.	0.009*"minister" + 0.009*"measure" + 0.009*"anti" + 0.007*"health" + 0.007*"industrial" + 0.006*"government" + 0.006*"plant" + 0.005*"department" + 0.005*"industry" + 0.005*"environment"	0.068	4	Government Measures and Industrial Regulation

5.	0.011*"pakistan" + 0.006*"emission" + 0.005*"plan" + 0.005*"term" + 0.005*"long" + 0.005*"burn"	+	0.007*"policy" + 0.005*"crisis" + 0.005*"issue" + 0.005*"action"	+	0.059	5	Environmental Policy and National Challenges
----	---	---	--	---	-------	---	--

Table 4.1 represents the extracted top keywords along with suggested topics respectively. These topic (or themes) are dealt with as frames, since the keywords fills the slots of the suggested frames. These five frames are discussed in the following section.

The frame ‘Weather Forecast and Public Impact’ draws heavily from metrological discourse, with keywords including “rain (0.014),” “march (0.010),” “survey (0.009),” “people (0.007),” and “province (0.007)” combine with forecast-oriented slots such as “forecast (0.007),” “expect (0.006),” “snowfall (0.006),” “report (0.006),” and “weather (0.006).” Together, these terms convey atmospheric conditions as cyclic and predictable phenomenon. By incorporating smog within the narrative of logic weather cycles, the discourse attributes the issue to natural atmospheric variation rather than human-induced ecological degradation. These top keywords represent the smog as a seasonal expected cycle instead of a symbol of clear ecological symptom. In addition, the origins of the pollution are backgrounded completely. This frame, consequently, reduces the ecological visibility within the Lahore smog media discourse in Pakistan.

As the most dominant frame in the 2024 corpus, Air Pollution and Public Health becomes the principal interpretive lens for understanding the smog crisis in Lahore, Pakistan. The highest-weighted slots include “aqi (0.016),” “government (0.012),” “city (0.012),” “level (0.011),” and “school (0.010).” These top keywords highlight measurable indicators, administrative responsibility, and the direct consequences for urban population. In addition, supporting slots such as “vehicle (0.009),” “child (0.007),” “environmental (0.007),” and “protection (0.007)” construct a story focused on human vulnerability; children and school-going populations are represented as the first victim of the crisis. Since this frame dominates the dataset, smog is primarily framed as a matter respiratory risk, school closures, or urban discomfort rather than as an ecological (human-induced) disturbance. Moreover, the environment is often mentioned only through institutional vocabulary (e.g., environmental protection), which prioritizes regulatory action over ecological interdependence. Ultimately, this frame naturalizes as anthropocentric perspective of the Lahore smog: the smog issue matters the most when it impacts human bodies, while degradation of ecological systems or impacts on other species remains comparatively absent.

The mid-ranking frame, Local Disruptions and Community Response, consists top keywords that highlight spatial and civic disruptions, such as “road (0.014),” “area (0.008),” and “district (0.008).” Other additional slots including “smoke (0.006),” “lockdown (0.006),” “public (0.006),” “effort (0.006),” “police (0.005),” and “impose (0.005)” represent smog as a logistical and administrative issue. This frame gives more emphasis to restricted mobility, government interventions, and other responses taken to enhance air quality. The word “green” has been appeared in the context of administrative response rather than the ecological frame itself. This frame is a mid-ranked frame since it neither challenges the dominant narratives nor does it provide a strong ecological alternative. Ultimately, it presents the smog as an issue which requires civic management enforcement.

Government Measures and Industrial Regulation is the frame with lower weight in the discourse; it connects, though not dominantly, smog to the sources of pollution. The top keywords of the frame are comprised of “minister (0.009),” “measure (0.009),” “anti (0.009),” “health (0.007),” “industrial (0.007),” “industry (0.005),” and “environment (0.005).” This frame opens the discourse by discussing industrial facilities, environmental agencies, and regulatory departments. Nevertheless, despite this ecological presence, the linguistic space is highly diplomatic and strategic. Moreover, since this frame is less dominant, its capacity to highlight and build ecological relationship remains limited. Though it contributes to the public understanding of the Lahore smog. Its diminished weight, however, restricts its influence on the smog framing.

The least-dominant frame, Environmental Policy and National Challenges, comprises the most ecological keywords, yet it exerts least discursive influence. Slots such as “pakistan (0.011),” “policy (0.007),” “emission (0.006),” “crisis (0.005),” “plan (0.005),” “issue (0.005),” “term (0.005),” “action (0.005),” “long (0.005),” and “burn (0.005)” points to the systemic origins of smog; it includes emissions, crop burning, and long-term environmental planning. However, although this frame offers the most direct linguistic access to ecological causation, its low ranking confirms that such perspectives are least prevalent in public discourse. The ecological narrative is therefore present but quiet. It appears largely as an abstract policy discussion rather than as a lived ecological relationship involving land, air, plant life, and human responsibility. The Following figure demonstrates the topic (frames) hierarchy or ranking.



Figure 4. 1 Interactive LDA Topic Ranking (Vertical Layout) in 2024 Corpus

Figure 4.1 presents the interactive topic ranking in a vertical bar layout; it illustrates the relative weight of each frame in the 2024 Pakistani news corpus. On the horizontal axis, the five frames are displayed, while the vertical axis reflects their corresponding weight values. The height

of each bar visually shows the prominence of the respective frame. Frame two, Air Pollution and Public Health, exhibits the greatest weight, followed by Frame two, Weather Forecast and Public Impact. These two frames emphasize human health concerns and predictability of the smog as a natural phenomenon. Frame three, Local Disruptions and Community Response, occupies a mid-level position. Frame four and five (e.g., Government Measures and Industrial Regulation and Environmental Policy and National Challenges) are the least prominent. This graphical representation highlights the hierarchy of narrative (or frame) emphasis. Immediate human-centered concerns (e.g., health issues) and metrological normalization (e.g., weather forecast and predictability) dominate the media discourse. Structural and ecological dimensions, on the other hand, are comparatively marginal. Throughout the 2024 corpus, the media presents an anthropocentric worldview (which is focused on humans), while ecological relationships and health impacts of other species are silenced. The Figure 4.1 provides a visual confirmation of the patterns observed in the Eco-LDA frame analysis; it reinforces the interpretation of which aspects of the smog crisis are foregrounded or silenced in Pakistani news coverage.

Table 4.2 LDA Extracted Top Keywords and Suggested Topics (Frames) in 2025 Corpus

Topic	Keywords with Weight	Weight	Rank	Suggested Topics
1.	air: 0.024, smog: 0.020, climate: 0.019, punjab: 0.018, clean: 0.013, change: 0.013, pakistan: 0.012, pollution: 0.011, government: 0.010, policy: 0.008	0.148	1	Climate Change in the Region
2.	pm: 0.016, particulate: 0.011, matter: 0.010, concentration: 0.007, measure: 0.006, aqi: 0.006, guideline: 0.006, fine: 0.006, datum: 0.006, report: 0.006	0.08	4	Monitoring and Assessing Air Quality
3.	say: 0.027, lahore: 0.016, antismog: 0.009, conference: 0.008, gun: 0.006, environmental: 0.006, water: 0.006, issue: 0.006, city: 0.005, area: 0.005	0.094	3	Government Antismog Response
4.	quality: 0.016, level: 0.013, world: 0.011, emission: 0.010, health: 0.010, year: 0.009, pollute: 0.008, month: 0.007, high: 0.006, environment: 0.006	0.096	2	Health Impacts of Environmental Pollution

Table 4.2 represents the four dominant frames extracted from the 2025 corpus of Pakistani news coverage on the Lahore smog crisis. These frames reflect just how Pakistani media discusses the Lahore smog issues.

The first frame, as suggested by the model output, is Climate Change in the Region (0.148). The frame is comprised of the keywords such as “air (0.024),” “smog (0.020),” “climate (0.019),” “punjab (0.018),” and “clean (0.013)” with supporting keywords including “change (0.013),” “pakistan (0.012),” “pollution (0.011),” “government (0.010),” and “policy (0.008s).” While framing the issue of smog as an ecological crisis, the frame links it to human activity and local governance. The frame further illustrates the smog as a long-term environmental disruption. Unlike 2024 media coverage, the issue as a result of the large scale environmental transformation

rather than an episodic weather phenomenon. Moreover, “government (0.010)” and “policy (0.008s)” slots of the frame highlight the local government responsibility for the possible ecological outcomes. The top keywords, altogether, makes human-nature inter-relationship visible; it illustrates how human-induced actions and government responses can affect the overall ecological sustainability.

Monitoring and Assessing Air Quality (0.080) is the second frame constructed in 2025 corpus which highlights scientific and technical discourse around the crisis. The slots with high weightage include keywords such as “pm (0.016),” “particulate (0.011),” “matter (0.010),” “concentration (0.007),” and “measure (0.006),” with supporting keywords such as “aqi (0.006),” “guideline (0.006),” “fine (0.006),” “datum (0.006),” and “report (0.006).” This frame constructs the smog as a phenomenon which could be quantified and measured (in the shape of “aqi”). It, therefore, highlights data collection and scientific assessment. Moreover, the frame also quantifies human exposure to the risk. Although the root-causes are secondary, the frame implicitly acknowledges ecological harm through specific references to pollution levels and concentrations. The frame, ultimately, approaches the crisis with a scientific lens, focusing on technical evaluation. In addition, the discourse focuses on manageability and control rather than long-term ecological wellbeing. Furthermore, the smog is represented as a scientific problem rather than an ecological crisis.

Government Antismog Response (0.094) is the third frame constructed in the media discourse which foregrounds institutional intervention and policy action. The top keywords are comprised of “say (0.027),” “lahore (0.016),” “antismog (0.009),” “conference (0.008),” and “gun (0.006),” with supporting slots such as “environmental (0.006),” “water (0.006),” “issue (0.006),” “city (0.005),” and “area (0.005).” The frame represents the issue as a challenge to be tackled by the local government (governance and policy action). Although ecological elements are visible, yet they are subordinated to policy action. In addition, the narrative emphasizes regulation and administrative authority over long-term causes of the ecological crisis. The frame, thus, reinforces human-centered solutions while backgrounding the root-causes of the ecological issue (e.g., the Lahore smog). Furthermore, the narrative illustrates the Lahore smog as manageable crisis; it highlights procedural responses rather than underlying ecological interconnections.

Health Impacts of Environmental Pollution (0.096) is the fourth frame constructed in the media coverage of the crisis in 2025. The frame highlights vulnerability in relation to environmental crisis (e.g., the Lahore smog). The top and high-weight slots include “quality (0.016),” “level (0.013),” “world (0.011),” “emission (0.010),” and “health (0.010),” while some other supporting slots are “year (0.009),” “pollute (0.008),” “month (0.007),” “high (0.006),” and “environment (0.006).” The frame represents the smog as a public health concern which is linked to the environmental degradations. The narrative also foregrounds societal impacts, while connecting wellbeing to the air quality and levels of emissions. In addition, ecological degradation is made visible by emphasizing the interdependence of human and environmental systems. Moreover, although the long-term root-causes of the crisis are less dominant, the frame illustrates how media discourses link ecological issues with tangible human consequences. The frame also signals the urgency of environmental stewardship for public health protection.

Taken together, these frames illustrate how the 2025 corpus on Lahore smog negotiates ecological, technical, administrative, and health-related dimensions. The results show greater

ecological visibility than in previous year (2024), yet systemic drivers of degradation remain only partially expressed within media narratives. The following figure shows the hierarchy of these topics.

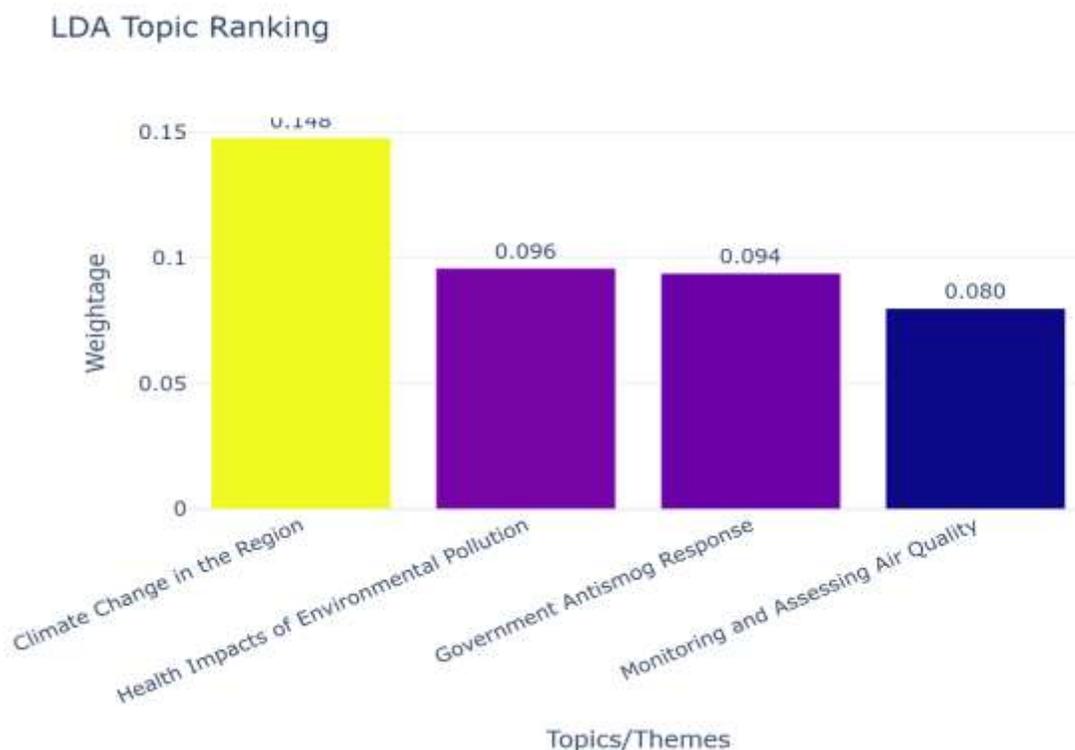


Figure 4. 2 LDA Topic (Frame) Ranking in 2025 Corpus

Figure 4.2 presents the ranked distribution of the 2025 corpus frames and shows a clear reorientation in how the smog crisis is structured in the public discourse. The highest weighting of the “Climate Change in the Region” frame positions ecological processes as the central explanatory lens for 2025 news coverage of the Lahore smog issue; it indicates a discourse that foregrounds environmental driver rather than treating smog as an isolated or cyclic natural event (e.g., as witnessed in the first frame of 2024 corpus). In addition, the “Monitoring and Assessing Air Quality” frame appears with reduced prominence. It suggests that technical measurement is no longer the dominant mode through which the crisis is interpreted. Moreover, the “Government Antismog Response” frame holds a mid-level position. It reflects a continued but less forceful emphasis on administrative intervention. Finally, the “Health impacts of Environmental Pollution” frame occupies the lowest rank; it shows a shift away from health-centered immediacy as the primary anchor. Compared with 2024 corpus, the most apparent change lies in the strengthened ecological visibility of 2025; it has moved climatic and environmental determinants to the forefront, while the earlier metrological or episodic framings (of 2024 corpus) recedes. Luckily, this shift marks a more explicitly ecological narrative structure in the 2025 corpus.

The following figure reflects the significant shifts in media framing of Lahore smog crisis over 2024 and 2025.

Line Plot: Frame Weights 2024 vs 2025

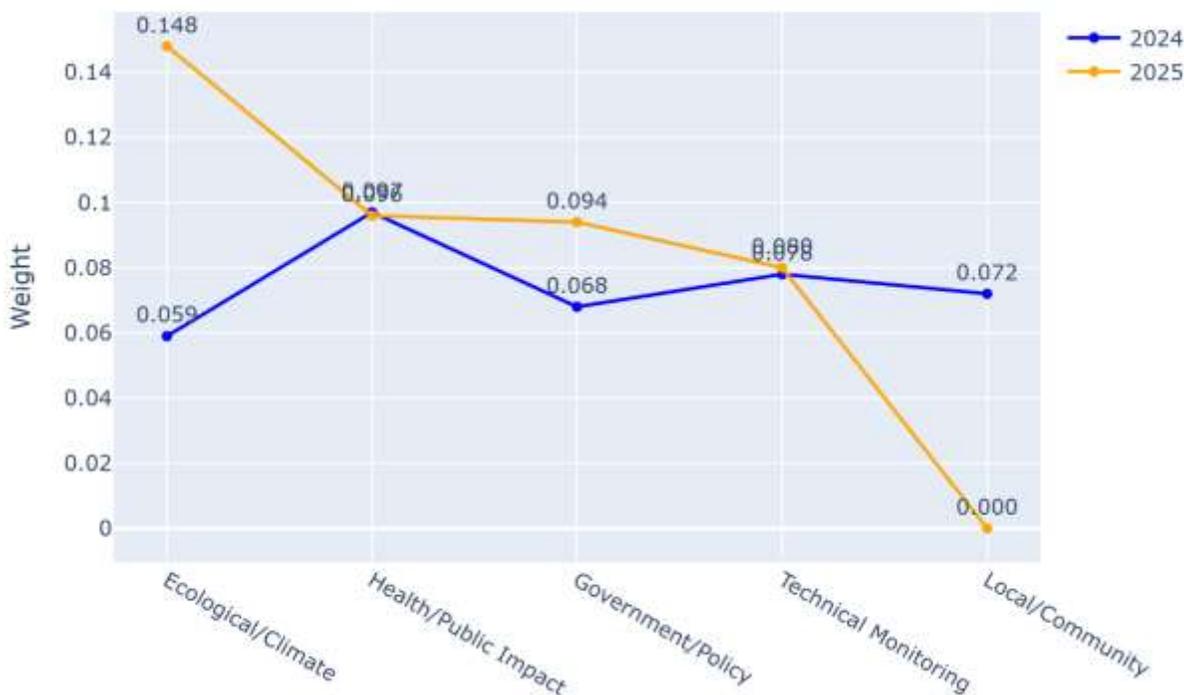


Figure 4. 3 Line Plot: Frame Weights of 2024 and 2025 Corpora

Figure 4.3 illustrates the temporal shift in media framing of the Lahore smog issue between 2024 and 2025, as visualized using Python 3.12 with the Plotly library. In the figure, the blue line represents 2024 framing, while the orange line reflects 2025 framing of the crisis. The visualization illustrates a rise in the ecological awareness in Pakistani news media; the “ecological/ climate” theme has climbed from 0.059 in 2024 to 0.148 in 2025. However, the contrasting theme “health/ public impact” has declined from a weight of 0.097 to 0.096. Other frames, such as “government/ policy” and “technical monitoring,” represents a moderate surge as the media has continually focused on government responses and the monitoring of air quality. Collectively, Figure 4.3 demonstrates a reduction in anthropocentric reporting of the Lahore smog in 2025 as compared to that of 2024. Ultimately, this shift in reporting (2024-25) of the ecological crisis (e.g., Lahore smog) highlights the emergence of a more ecologically informed narrative in Pakistani news coverage of the Lahore smog crisis. Similar results are visualized in the following figure 4.4.



Figure 4.4 Word-cloud of Key Topics (Frames) in 2024 and 2025

The Figure 4.4 illustrates the media transition regarding the subject of Lahore smog over the . Generated through Plotly library in Python 3.12, the above visual reflects the core results of the research. It demonstrates how Pakistani media framing of the environmental crisis has evolved over the last two years. Specifically, it reflects how 2024 media coverage focused more on or highlighted the smog as something cyclic and episodic. It discussed the ecological issue as equivalent to weather which could be predicted. Therefore, the 2024 reporting framed the Lahore smog crisis through an anthropocentric worldview. However, the 2025 media coverage gave more emphasis to the climate change itself, and tried to highlight the core causes of the crisis (e.g., particulate matter, emission, concentration, pollution). In addition, the keyword “antismog” alone reflect the idea that the ecological understanding and awareness in Pakistani news media has improved in the last year (2025). Ultimately, the word-cloud in figure 4.4 visualizes the evolution in media framing of the Lahore smog over the last two years; it represents a latest snapshot of the hazardous and alarming situation, in the shape of Lahore smog, in Pakistan.

Discussion

Utilizing a newly developed framework, named the Eco-LDA Framework, the research analyzed the framing of the Lahore smog crisis in Pakistani news media (2024-2025), including the *News* and *The News International*. The framework is not adopted from prior studies but was conceptualized by researcher from the integration of frame semantics, computational modeling (or LDA), and Stibbe’s (2015) ecolinguistic framework (for an ecolinguistic interpretation). The framework is replicable and can be used in other studies to identify and interpret ecological visibility in environmental crises.

Variation in media framing of climate change issues influences public understanding as well as their ecological behavior. For example, the *People’s Daily* represents climate change as an international conflict (e.g., China-US power dynamics), while the *Sydney Morning* frame the issue as a domestic tension within a country (e.g., Australia).

In our context, the 2024 corpus of the Lahore smog indicated that the Pakistani media primarily focused on human-centered (i.e., anthropocentric) and episodic (or predictive) perspectives. For instance, the most dominant topic (or frame), Air Pollution and Public Health,

emphasized institutional responsibility, AQI levels, and impacts on urban population. Moreover, other frames such as “Weather Forecast and Public Impact” represented the crisis as a predictable seasonal phenomenon, and it further backgrounded possible human-induced causes of the crisis (e.g., Lahore smog). In addition, the less prominent frames were “Local disruptions and Community Response,” “Government Measures and Industrial Regulations,” and “Environmental Policy and National Challenges.” Since these topics were less dominant, therefore, the 2024 corpus offered limited ecological perspective on the crisis. Across the 2024 corpus modeling, frames were anthropocentric as reflected by the results. For example, immediate health and civic concerns were foregrounded while ecological relationships were marginalized. Similarly, according to Boykoff and Boykoff (2007), by prioritizing human-interests, media coverage often normalizes climate change. This framing further sidelines the nature and intensity of the ecological crisis. For example, a similar study of the media coverage of 2025 Pakistan floods revealed that the corpus primarily employed an anthropocentric framing to discuss the issue. Moreover, the media framing highlighted human and infrastructural impacts while it largely marginalized ecological factors such as systemic environmental causes (Nawaz et al., 2025).

According to Tanveer (2024), through strategic engagement digital media framing can shape public perceptions and foster rise climate change awareness. In a similar way, in our study, the 2025 framing of the Lahore smog has shifted towards more ecologically aware narratives. For example, the most prominent frame, “Climate Change in the Region,” has foregrounded the long-term environmental causes (e.g., industrial emissions, crop burning, and regional climate change) of the issue. Health impacts, on the other hand, were still in the cluster however they were not much visible as compared to the 2024 corpus. Ultimately, it witnesses a rising awareness about the ecological perspectives in the reporting of environmental issues (e.g., Lahore smog) in Pakistani media. It also indicates the inter-relationship between environmental systems, good governance, and societal impacts.

The Plot chart or connected line visualization in Figure 4.3’ reflects the same story about the shifting in framing of the crisis over 2024 and 2025. The blue line, specifically, demonstrates the dominance of anthropocentric and episodic concerns in 2024 media framing, while the orange line highlights the rising prominence of ecological narratives in 2025 coverage. These trends are reflected in the following interactive hover bar chart designed through Python 3.12.

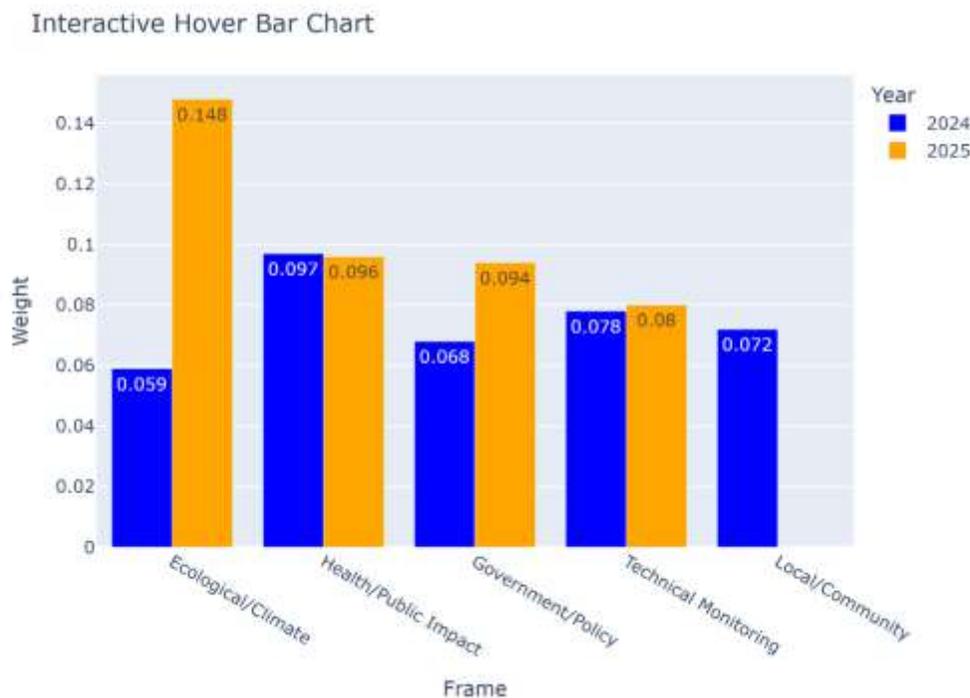


Figure 5.1 Interactive Hover Bar Chart of 2024 and 2025 Frames

Figure 5.1 is yet another visual representation (Interactive Hover Bar Chart) of the framing in the corpora in 2024 and 2025. It reflects a shift from human-centered and short-term solution oriented framing in 2024 towards a more ecological and holistic discussion in 2025 corpus (media framing). This kind of framing in the reporting of environmental issues plays a pivotal role in improving environmental awareness (Jharotia, 2018). In addition, it also fosters government green initiatives and encourages positive ecological behavioral change in the public. Furthermore, Zubair and Almas (2026) argue that any ecocatastrophe in the modern world constitutes an irreversible threat to the ecosystem (e.g., a nuclear disaster); confronting environmental degradation is critical. Moreover, modern fiction, such as *The Light Pirate*, could play a pivotal role in enhancing understanding of human-nature relationship while promoting environmental awareness and responsibility.

The following section briefly answers the three main research questions:

The first research question was to examine the latent themes constructed in Pakistani news discourse on Lahore smog. The results showed that the 2024 corpus had five distant themes as extracted by running LDA. The 2024 latent themes include: Weather Forecast and Public Impact, Air pollution and Public Health, Local Disruptions and Community Response, Government Measures and Industrial Regulation, and Environmental Policy and National Challenges. The LDA of 2025 corpus, on the other hand, extracted four latent themes including Climate Change in the Region, Health Impacts of Environmental Pollution, Government Antismog Response, and Monitoring and Assessing Air Quality. These latent themes reflected a change in how the media discussed and framed the Lahore smog crisis over 2024 and 2025. The shift also highlighted an increased prominence of ecological and climate-related themes (in 2025).

The second research question was to analyze how do the extracted latent themes frame the Lahore smog crisis and reveal underlying ecological relationship? The results showed that in 2024 corpus, the frames focused on short-term human impacts, civic disruptions, and administrative enforcements, while the ecological causes of the issue were largely marginalized. Likewise, in Pakistan, newspapers employ solution-based journalism but provide limited and episodic coverage of environmental issues such as smog (Javed et al., 2023). However, in 2025 corpus, however, the “Climate Change in the Region” frame foregrounded systemic environmental causes; it linked smog to industrial emissions, crop burning, and regional climate change. Moreover, health and policy frames complemented this perspective. It also revealed a growing awareness of human-environment interdependencies in media discourse.

The third research question was to examine how do Pakistani news coverage of Lahore’s smog has evolved between 2024 and 2025? The Eco-LDA framework assisted analysis revealed that the coverage has evolved from anthropocentric, episodic (or cyclic and predictive) framing in 2024 corpus to a more ecologically informed, systemic framing in 2025 corpus. The dominance and transition of ecological frames in 2025 corpus reflects a positive change in terms of ecological sustainability (e.g., integrating environmental frames). In addition, short-term, human-centered (i.e., immediate health issues), and episodic frames are less visible in 2025 reporting. This shift, altogether, indicates a more ecologically aware and sustainability focused understanding of the Lahore smog in 2025. Specifically, the keywords such as “antismog” further highlights the government and climate bodies seriousness.

Conclusion

The research was aimed at investigating the transition in framing of the Lahore smog issue in Pakistani media, specifically the *Dawn* and *The News International*, over 2024 and 2025. The main objectives were to identify latent themes, investigate their framing of the ecological relationships, and to assess the evolution in media framing of the crisis (e.g., The Lahore smog) over the time. To achieve the objectives, the researcher developed and employed the Eco-LDA Framework conceptualized from the nexus of computational topic modeling (e.g., LDA), Semantics, and Stibbe’s (2016) ecolinguistic framework. The corpus 2024 was comprised of 200 articles (100 from each outlet), while the 2025 corpus included 160 articles (80 from each news media). The inconsistent number of article was the result of purposive sampling as only those articles were gathered which explicitly discussed the Lahore smog either in 2024 or 2025.

The results indicated that different framing trends were employed by the Pakistani news media outlets in 2024 and 2025. The 2024 dominant frames, specifically, were largely episodic and anthropocentric in nature. While backgrounding the ecological causes of the crisis and the long-term sustainable solutions, the frames focused highly on human-centered narratives (e.g., health risks, school closures, and administrative responses). In contrast, in 2025 media reporting of the crisis (e.g., the Lahore smog), the framing has transitioned towards a discourse which emphasizes ecological wellbeing and sustainability. The four dominant frames, in 2025 corpus, notably the “Climate Change in the Region,” represented smog as a part of the climate cluster (e.g., the border climatic and ecological processes). This transition indicates Pakistani media’s rising awareness and recognition of the Lahore smog as a serious ecological concern rather than just a cyclic or episodic (e.g., weather predictive) phenomenon with public health impacts.

In conclusion, the research found that the Pakistani media outlets, specifically the *Dawn* and *The News International*, have gradually shifted their focus, in the Lahore smog reporting,

towards more ecologically informed and sustainability-focused narratives over 2024 and 2025. The results, thus, offer implications for public understanding of the ecological crises, national green policies discourse, and environmental communication in Pakistan. In addition, the proposed Eco-LDA Framework was efficient, effective, and replicable. It further indicates that the framework could be employed in relevant computational ecolinguistic enquires (e.g., the understanding of ecological crises). Moreover, throughout the research, it has been confirmed that media framing can evolve to enhance public ecological awareness. Such foregrounding may further encourage a more informed and ecologically responsible behavior.

Limitations and Future Research Directions

This research is not bulletproof; it has several limitations. Firstly, the subject of analysis is limited to only two leading Pakistani news outlets, such as the *Dawn* and *The News International*, which restrict the generalizability of the results across the diverse media environment in Pakistan. Secondly, the corpora are comprised of limited articles because the data has been gathered manually by the researcher; there were limited research facilities and corpus databases availability. Thirdly, the analysis has extensively focused only on the textual data while all other visual modalities were disregarded. In addition, the audience perception has not been dealt with in this research. Finally, instead of long-term big corpora and decade's transitioning signals, the research presents a snapshot of the recent two years' (2024-2025) ecological insights of the contemporary media framing of the Lahore smog crisis. For capturing a more holistic picture of the Lahore smog issue, future researchers may expand the scope of the study to include other leading news outlets as well as other visual modalities for the computational ecolinguistic inquiry of the ecological crises in Pakistan. In addition, expanding the corpus size may also enhance the rigor and robustness of the computational analyses carried out through the Eco-LDA Framework. Moreover, longitudinal studies may further investigate whether the ecologically informed framing observed in 2025 is a context-specific and timely response or it is a sustained discursive shift towards better environmental communication in Pakistani media. Finally, as the proposed Eco-LDA Framework is replicable, the future researchers may employ it in their computational ecolinguistic enquiries (e.g., flooding, water scarcity, SDGs, or the climate change) to assess and enhance understanding of ecological crises across the globe.

References

- Adnan, M., Xiao, B., Bibi, S., Xiao, P., Zhao, P., Wang, H., ... & An, X. (2024). Known and unknown environmental impacts related to climate changes in Pakistan: an under-recognized risk to local communities. *Sustainability*, 16(14), 6108.
- Akbar, M. B., Arif, A., & Khalid, A. (2025). A Social Marketing Perspective on Tackling SMOG—An Example From Lahore. *Social Marketing Quarterly*, 31(3), 317-323.
- Al Jazeera. (2025, September 19). 'Crisis of justice' as floods devastate: Pakistan's climate change minister. Retrieved from <https://www.aljazeera.com/news/2025/9/19/impact-of-climate-change-a-harsh-reality-facing-pakistan>
- Awan, P. (2025, April 7). 'Choking on smog: Lahore's air crisis threatens health and wellbeing'. *Global South Hub*. Retrieved from <https://globalsouth.org/2025/04/choking-on-smog-lahores-air-crisis-threatens-health-and-wellbeing/>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

- Boas, H. C. (2005). From theory to practice: Frame Semantics and the design of FrameNet (pp. 129-160). na.
- Caiani, M. (2023). Framing and social movements. *Discourse Studies*, 25(2), 195-209.
- Chang, I. C., Yu, T. K., Chang, Y. J., & Yu, T. Y. (2021). Applying text mining, clustering analysis, and latent dirichlet allocation techniques for topic classification of environmental education journals. *Sustainability*, 13(19), 10856.
- Church, K., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational linguistics*, 19(1), 1-24.
- Churchill, R., & Singh, L. (2022). The evolution of topic modeling. *ACM Computing Surveys*, 54(10s), 1-35.
- David, C. C., Atun, J. M., Fille, E., & Monterola, C. (2011). Finding frames: Comparing two methods of frame analysis. *Communication methods and measures*, 5(4), 329-351.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1), 20-32.
- Fillmore, C. J. (2006). Frame semantics. *Cognitive linguistics: Basic readings*, 34, 373-400.
- Fruchtman, E. (2020). A media frame analysis of xeno-racism after Brexit.
- Gawron, J. M., Maienborn, C., Von Heusinger, K., & Portner, P. (2008, January). Frame semantics.
- Gorczyca, K., & Zozulia, D. (2025). "We have the right to breathe clean air"—Mobilising communities in the fight for good air quality. *Environment and Planning C: Politics and Space*, 23996544251384532.
- Gul, U., Faraz, H., & Ali, U. (2025). Sustainability Concerns and Government Stance: A Corpus-Based Genre Analysis of the Selected Three Pakistani Newspapers' Articles. *Social Science Review Archives*, 3(3), 2017-2029.
- Huan, C. (2024). Politicized or popularized? News values and news voices in China's and Australia's media discourse of climate change. *Critical Discourse Studies*, 21(2), 200-217.
- Hussain, A., Abbas, M., & Kabir, M. (2025). Smog Pollution in Lahore, Pakistan: A Review of the Causes, Effects, and Mitigation Strategies.
- IQAir. (2025). Live most polluted major city ranking. IQAir. Retrieved October 29, 2025, from <https://www.iqair.com/world-air-quality-ranking>
- Javed, H., Asgher, S., & Bukhari, S. F. (2023). Role of media in promoting environment protection during smog crisis in Lahore. *Journal of Media & Communication*, 4(1), 1-23.
- Jharotia, A. K. (2018, March). Role of media in enhancement of environmental awareness. In Conference: Power of Media: Shaping the Future, At Tecnia Auditorium, New Delhi (pp. 1-11).
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, 78(11), 15169-15211.
- Kutter, A. (2017). Corpus analysis. In *The Routledge Handbook of Language and Politics* (pp. 169-186). Routledge.
- Majeed, R., Anjum, M. S., Imad-ud-din, M., Malik, S., Anwar, M. N., Anwar, B., & Khokhar, M. F. (2024). Solving the mysteries of Lahore smog: the fifth season in the country. *Frontiers in Sustainable Cities*, 5, 1314426.

- Nawaz, M. S., Siddiq, S., & Hakim, S. (2025). A CORPUS-BASED LINGUISTIC ANALYSIS OF DISCOURSE MARKER USAGE IN ENGLISH GRADUATE STUDENTS' ACADEMIC PRESENTATIONS. *Journal of Applied Linguistics and TESOL (JALT)*, 8(3), 2541-2554. <https://doi.org/10.71281/jals.v3i4.485>
- Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological methods & research*, 49(1), 3-42.
- Qusien, R. (2025). Beyond the haze: Media coverage of smog in Pakistan. *Journalism*, 14648849251316250.
- Razzaq, A., Zafar, M. M., Zahra, L. T., Qadir, F., Qiao, F., Ullah, M. H., ... & Jiang, X. (2024). Smog: Lahore needs global attention to fix it. *Environmental Challenges*, 16, 100999.
- Setiadi, S. W. (2025). Metaphor and framing in climate reporting: An colinguistics perspective (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Stibbe, A. (2015). *Ecolinguistics: Language, ecology and the stories we live by*. Routledge.
- Tanveer, Z. (2024). Leveraging Digital Media for Climate Awareness: The Role of Social Platforms in Environmental Advocacy. *Journal of Jurivox*, 1(1), 1-9.
- Touileb, S., & Salway, A. (2014). Constructions: a new unit of analysis for corpus-based discourse analysis. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation* (pp. 634-643). Waseda University.
- Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582.
- Zubair, M., & Almas, N. (2026). Ecological uncanny and fallout of repression: exploring the metaphor of ecocatastrophe in *The Light Pirate*. *Journal of Humanities and Applied Social Sciences*, 1-11. <https://doi.org/10.1108/JHASS-09-2025-0177>