



The Representation of the Word ‘Pollution’: A Corpus-Assisted Ecolinguistics Study

Gusti Ayu Praminatih¹

[gusti.praminatih@ipb-intl.ac.id¹](mailto:gusti.praminatih@ipb-intl.ac.id)

Faculty of Tourism and Business, Institut Pariwisata dan
Bisnis Internasional¹

Abstract

This study examines how the concept of pollution is constructed in Indonesian discourse by combining corpus linguistics with ecolinguistic interpretation. Using data from the Leipzig Corpora Collection, specifically the 2019 news collection, and analysed using AntConc software, nominal collocates of pollution were identified and examined in relation to their nominal collocation and discursive implications. The findings show strong associations with air quality, health risks, governance, and technological solutions. Nominal collocation interpretation highlights discourses of risk and victimhood, often centred on toxic exposure and children, alongside discourses of responsibility and technical solutions, expressed through references to regulation, electric vehicles, etc. The study suggests that the representation of the word 'pollution' varies from risk-centred narratives to solution-oriented framings, reflecting broader sustainability discourses. This study contributes to how language shapes ecological understanding and provides empirical insights into environmental discourse in Indonesia.

Keywords: collocation, ecolinguistics, Indonesian news corpus, Leipzig Corpora Collection, linguistic corpus

Introduction

Pollution has emerged as one of the most urgent ecological challenges facing Indonesia and the world. Rapid industrialisation, urban expansion, and regulatory gaps have intensified air and water contamination, thus threatening biodiversity, accelerating climate impacts, and creating profound risks for human health. While natural and social scientists have extensively documented the material harms of pollution, the role of language in producing, normalising, or contesting these harms remains underexplored. It is here that ecolinguistics becomes essential, as language shapes the mental models and cultural narratives, i.e., the stories we live by that legitimate practices such as unchecked industrial growth and consumerism, while marginalising native and sustainable practices (Stibbe, 2014, 2015). Therefore, examining linguistic patterns is theoretically valuable but also socially urgent, as discourse frames public perception and legitimacy, and can either accelerate or impede mitigation and remediation efforts (Steffensen, 2025; Stibbe, 2014, 2015).

This study addresses that gap by investigating how pollution is represented within Indonesian discourse, employing large-scale corpus methods. Specifically, it analyses patterns, contextual associations, and emergent discursive frames of lemma pollution in the Indonesian section of the Leipzig Corpora Collection (2013). This widely used multilingual resource provides billions of words of authentic text data. The corpus-assisted ecolinguistic approach enables the quantification of representations (Biber, 2011; Biber et al., 2007; Biber & Conrad, 2001), the revelation of otherwise invisible patterns, and the detection of shifts that small-scale qualitative samples might overlook (Baker, 2006, 2010; Baker et al., 2006, 2008).

Precedents from related studies demonstrate the value of such an approach. A study conducted by Liu and Li (2017) in a corpus-assisted discourse analysis of Chinese and Anglo-American newspapers demonstrated how smog was discursively constructed in competing ways, with implications for assigning responsibility and shaping public perceptions of risk. A similar study by Liu and Zhang (2018) highlighted how collocations and semantic prosodies surrounding “smog” and “air pollution” reveal competing framings of uncertainty, blame, and solution narratives. Winter et al. (2021) examined how civil society in Shanghai perceives and frames the issue of air pollution. It explores the discourses, narratives, and attitudes that shape public understanding and responses to environmental challenges.

Within ecolinguistics, Steffensen (2025) has mapped the field’s theoretical and methodological developments, including the rise of corpus-assisted studies that allow more robust generalisations across large datasets. Thus, by combining quantitative corpus analysis with ecolinguistic interpretive depth, the present study contributes novel empirical evidence to environmental communication, ecolinguistics, and discourse studies. It seeks to identify what is said about pollution in Indonesian discourse, as well as how it is said, which actors are framed as responsible, which risks are emphasised or downplayed, and what solutions are legitimised or silenced. Consequently, the study can contribute to communicative interventions, public education campaigns, and policy framing strategies to foster more ecologically sustainable narratives.

The present study proposed three research questions, as follows:

- 1) What collocational patterns accompany the word ‘pollution’ in the Indonesian section of the Leipzig Corpora Collection?
- 2) What semantic and contextual associations emerge from those collocations?
- 3) Which discursive frames (e.g., responsibility, risk, victimhood, technical solutions) are constructed, and how do they vary across genre or over time?

Literature Review

Theoretical insights from ecolinguistics provide a foundation for understanding how language shapes environmental discourse. Stibbe (2014, 2015) has argued that language constructs the “stories we live by”, narratives such as economic growth or consumerism that both reflect and reinforce ecological attitudes and behaviours. Within this framework, ecolinguistic analysis involves identifying discourses that encourage sustainable living and those that legitimise ecological harm. Steffensen (2025) has further mapped the development of ecolinguistics and emphasised the increasing importance of large corpora in addressing environmental issues. The survey highlights how corpus methods enable researchers to detect broad linguistic patterns, while critical ecolinguistic interpretation allows for the exploration of their ideological and ecological implications. These theoretical perspectives establish the rationale for combining corpus linguistics with ecolinguistics in the study of environmental discourse.

Empirical studies on air pollution discourse illustrate the value of this integrated approach. Liu and Li (2017) examined the discursive constructions of “smog” in Chinese and Anglo-American newspapers, identifying competing framings that attributed responsibility, underscored scientific uncertainty, or foregrounded technological solutions. Their study demonstrates how corpus methods can uncover systematic rhetorical choices that shape public understanding. Liu and Zhang (2018) examined the discourses surrounding scientific certainty and uncertainty related to the health risks associated with air pollution. Using corpus-assisted analysis, they showed how collocations and semantic prosody reveal whether risk is amplified or downplayed. These studies demonstrate the ability of corpora to move beyond surface-level frequency counts and to expose rhetorical strategies embedded in environmental discourse. Complementing these corpus-based perspectives, Winter et al. (2021) examined how civil society in Shanghai perceives and interprets air pollution. They highlighted narratives of public health, collective action, and legitimacy, showing that empirical corpus evidence should be accompanied by an understanding of how communities interpret and negotiate these discourses in practice.

These theoretical and empirical contributions suggest a twofold methodological strategy. Quantitative corpus analysis can identify collocational patterns, key lexical items, and diachronic shifts across genres, while ecolinguistic interpretation enables critical reflection on the ideological and ecological implications of such patterns. The methodological models developed by Liu and Li (2017) and Liu and Zhang (2018), combined with the civic perspective highlighted by Winter et al. (2021), reinforce the importance of integrating Stibbe’s ecolinguistic framework with Steffensen’s methodological guidance. This integration positions the present study to provide both empirical evidence and critical interpretation in advancing the analysis of environmental discourse.

Method

Corpus offers a unique perspective for collecting and analysing data, where the quantitative dimension enables researchers to identify linguistic patterns from large datasets (Biber & Conrad, 2001). By examining, for example, distribution, frequency, and N-grams, researchers can test hypotheses (Stefanowitsch, 2020). At the same time, qualitative insights are essential for interpreting the data, for instance, by clarifying the functions of the nominal collocations once extracted, as well as other important information about the use of nominal collocations. Corpus linguistics as a methodology emerged following the creation of the Brown Corpus of English in the 1960s, which laid the foundation for quantitative corpus research and subsequently informed qualitative corpus-based investigations (Hasko, 2020). Consequently, the present study employs corpus methods to identify nominal collocations that require further description and explanation.

There are five steps in obtaining the collocations of the word ‘pollution’ from the corpus. The first step involves selecting an appropriate corpus. Pollution has been a persistent issue in Indonesia for a considerable period, and therefore, a representative corpus is needed to capture how pollution has been represented. The Leipzig Corpora Collection (2013), particularly the Indonesian sub-section, was selected to conduct an ecolinguistic study supported by diachronic corpus evidence. A diachronic corpus is described as one that is carefully designed to represent a language or variety over time, thus allowing the tracing of linguistic change (Baker et al., 2006). The Leipzig Corpora Collection is considered highly representative due to its massive volume of data. The first step was therefore to download the Leipzig corpora news – Indonesian from the following link: https://corpora.uni-leipzig.de/en?corpusId=ind_news_2019. The consideration is that pollution cases in 2019 included the Southeast Asian haze crisis, which caused deaths and health problems in several countries, such as Indonesia, Malaysia, and Singapore, as well as a rise in global mortality rates due to pollution.

The second step involved importing the downloaded corpus into AntConc software (Anthony, 2023). In the third step, the word ‘pollution’ was entered into the search field, and the collocate function was applied. The software then automatically generated a list of collocations in descending order of frequency, accompanied by numerical data and contextual examples. The fourth step was to select the most frequent collocations for closer analysis, while filtering out adjectives that were not relevant to the word ‘pollution’. The fifth step was to categorise each collocation thematically according to its context to identify the main discursive representations of pollution in Indonesia.

Findings and Discussion

This study reveals that the word ‘pollution’ demonstrates strong nominal collocations in the Leipzig Corpora news – Indonesian section for the year 2019. Table 1 presents the nominal collocations of ‘pollution’ in 2019 as follows.

Table 1. Nominal Collocations of the word ‘pollution’
in the Leipzig Corpora – Indonesia 2019

No.	Nominal Collocations	Likelihood
1	<i>udara</i> ‘air’	2336.786
2	<i>asap</i> ‘haze’	314.919
3	Jakarta	131.782
4	<i>tingkat</i> ‘level’	123.806
5	<i>kendaraan</i> ‘vehicle’	101.783
6	<i>paparan</i> ‘exposure’	95.391
7	<i>lingkungan</i> ‘environment’	74.596
8	<i>debu</i> ‘dust’	71.893
9	<i>penyumbang</i> ‘contributor’	71.344
10	<i>kemacetan</i> ‘cogestion’	68.130

Based on the data retrieved from AntConc, Table 1 presents the top ten highest nominal collocations of the word ‘pollution’ from the Leipzig Corpora news – Indonesian in 2019 as follows.

1. “... *dan solusi jangka panjang bagi Indonesia, terutama untuk mengurangi **polusi udara** yang belakangan ini menjadi tantangan.*” (“... and a long-term solution for Indonesia, particularly to reduce **air pollution**, which has recently become a major challenge.”)
2. “... *sampah baik organik maupun non organik karena akan menambah kabut **asap** dan **polusi** yang membahayakan bagi kesehatan.*” (“... waste, both organic and non-organic, as it will add to the **haze** and **pollution** that pose health risks.”)
3. “*Pemerintah DKI **Jakarta** telah merespons permasalahan **polusi** udara dengan mengeluarkan Instruksi Gubernur (Ingub) Nomor 66 Tahun 2019 ...*” (“The **Jakarta** Provincial Government has responded to the issue of air **pollution** by issuing Governor’s Instruction No. 66 of 2019 ...”)
4. “... *mengklaim selama sepekan kebijakan perluasan ganjil genap diterapkan mampu mengurangi **tingkat polusi** udara dan kinerja lalu lintas ...*” (“... claiming that within a week of the odd-even traffic policy being extended, it was able to reduce **levels** of air **pollution** and improve traffic performance ...”)

5. “Jonan mengatakan **kendaraan listrik** akan mengurangi **polusi** dan tentunya lebih ramah lingkungan.” (“Jonan stated that electric **vehicles** would reduce **pollution** and, of course, be more environmentally friendly.”)
6. “... dan anak-anak mereka harus mendapatkan informasi yang akurat tentang **paparan polusi udara beracun** ...” (“... and their children must receive accurate information about **exposure** to toxic air **pollution** ...”)
7. “sebagai bentuk menjaga lingkungan karena jenis sumber energi ini ramah lingkungan bebas **polusi** dan tidak merusak alam.” “... as a way of protecting the environment, since this type of energy source is **environmentally** friendly, free from **pollution**, and does not damage nature.”
8. “memberikan perlindungan dari efek buruk sinar matahari serta **debu dan polusi**, menyembuhkan kerusakan rambut ...” (“... providing protection from the harmful effects of sunlight as well as **dust** and **pollution**, repairing hair damage ...”)
9. “Emisi gas buang kendaraan bermotor merupakan **penyumbang terbesar polusi udara**.” (“Exhaust emissions from motor vehicles are the largest **contributor** to air **pollution**.”)
10. “... pengoperasian LRT Jabodebek kelak dapat mengurangi **kemacetan dan polusi** di wilayah tersebut.” (“... the future operation of the Greater Jakarta LRT is expected to reduce **congestion** and **pollution** in the area.”)

The analysis of collocational patterns surrounding the word pollution in the Indonesian section of the Leipzig Corpora Collection reveals a consistent clustering around nominal associations. The top collocates include air, health, government policy, transportation, and toxic exposure, indicating that pollution is most frequently situated within discourses of environmental risk, institutional responsibility, and technical intervention. It aligns with earlier observations in corpus linguistics, which suggest that nominal collocates often anchor abstract phenomena in concrete social or environmental issues (Biber & Conrad, 2001; Stefanowitsch, 2020).

Beyond these lexical patterns, semantic and contextual associations highlight recurring themes of public health risks, urban management, and technological or policy solutions. Several examples foreground pollution as a health hazard, with collocates such as exposure and toxic air framing the problem through biomedical risk narratives. Others embed pollution within policy responses, as in references to Jakarta’s Governor’s Instruction or the odd-even traffic policy, which link environmental problems to urban governance and regulation. Technological optimism is also present, particularly in references to electric vehicles and light rail transport as solutions. At the same time, everyday lifestyle contexts emerge, such as the association of pollution with cosmetics advertisements, suggesting that environmental discourse is commodified alongside more critical framings.

From an ecolinguistic perspective, these collocational and semantic patterns reveal discursive frames that shape public interpretations of pollution. Narratives of risk are evident in references to health threats and toxic exposure, while responsibility is frequently attributed to government actors and technological interventions rather than civic participation. Victimhood is constructed through references to children and vulnerable populations, emphasising the human cost of environmental degradation. Technical solutionism is advanced through recurrent associations with electric mobility and infrastructural projects, which frame pollution as a problem solvable by innovation



and investment. It contrasts with Liu and Li's (2017) study of "smog" discourse in Chinese and Anglo-American newspapers, where competing framings emphasised scientific uncertainty and contested responsibility. In the Indonesian corpus, risk is amplified rather than downplayed, and government-led solutions dominate over civic narratives, a contrast to Winter et al. (2021), who emphasised grassroots perceptions and collective action in Shanghai.

Variation across genre and time further refines these findings. News discourse frequently foregrounds government measures and policy initiatives. At the same time, promotional or lifestyle texts embed pollution within narratives of consumption, such as personal care products that promise protection from environmental harm. The study suggests a gradual strengthening of policy-oriented and technological frames, particularly in reference to electric vehicles, renewable energy, and urban transport solutions, reflecting the broader societal shift towards sustainability discourses. It indicates that while the discourse of pollution in Indonesia continues to highlight health risks, it increasingly incorporates solution-oriented narratives that legitimise both governmental regulation and technological development as central responses.

These findings demonstrate how collocational analysis uncovers systematic patterns, how semantic and contextual interpretation highlights risk, responsibility, and solution-oriented associations, and how discursive frames shift and evolve. By integrating corpus evidence with ecolinguistic interpretation, the present study contributes both empirical insights and critical reflection to the analysis of environmental discourse.

Conclusion

While the present study has demonstrated the value of combining corpus methods with ecolinguistic interpretation in examining pollution discourse, several limitations must be acknowledged. The analysis relied on a single corpus source, which, although extensive, may not fully capture the diversity of Indonesian discourse. Furthermore, the focus on nominal collocates, while offering clear patterns, restricts the exploration of grammatical constructions and evaluative stances that may also carry ecological significance. Another limitation lies in the five-year span of the data, which does not allow for the examination of longer diachronic shifts or the influence of major global environmental events.

Future research could expand the scope by triangulating data from multiple corpora, including social media and policy documents, to more effectively capture contemporary discursive dynamics. Expanding the analysis to verbal and adjectival collocates may also provide richer insight into evaluative language and agency. Comparative studies across languages and cultural contexts would further test the applicability of ecolinguistic frameworks in diverse ecological discourses. By addressing these avenues, subsequent research can deepen understanding of how language shapes environmental awareness and contributes to ecological change.

References

- Anthony, L. (2023). *AntConc* (4.2.4). Waseda University.
Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.

- Baker, P. (2010). Corpus Methods in Linguistics. In L. Litosseliti (Ed.), *Research Methods in Linguistics* (pp. 93–113). Continuum.
- Baker, P., Gabrielatos, C., Khosravini, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh University Press.
- Biber, D. (2011). Corpus linguistics and the study of literature. *Scientific Study of Literature*, 1(1), 15–23. <https://doi.org/10.1075/ssol.1.1.02bib>
- Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure* (Vol. 28). John Benjamins Publishing Company.
- Biber, D., & Conrad, S. (2001). Quantitative Corpus-Based Research: Much More Than Bean Counting. *TESOL Quarterly*, 35(2), 331–336.
- Hasko, V. (2020). Qualitative Corpus Analysis. *The Encyclopedia of Applied Linguistics*, 1–7. <https://doi.org/10.1002/9781405198431.wbeal0974.pub2>
- Leipzig Corpora Collection-Indonesian. (2013). *Download Corpora Indonesian*. <https://wortschatz.uni-leipzig.de/en/download/Indonesian>
- Liu, M., & Li, C. (2017). Competing discursive constructions of China's smog in Chinese and Anglo-American English-language newspapers: A corpus-assisted discourse study. *Discourse & Communication*, 11(4), 386–403. <https://doi.org/10.1177/1750481317707379>
- Liu, M., & Zhang, Y. (2018). Discursive constructions of scientific (Un)certainty about the health risks of China's air pollution: A corpus-assisted discourse study. *Language & Communication*, 60, 1–10. <https://doi.org/10.1016/J.LANGCOM.2018.01.006>
- Stefanowitsch, A. (2020). Corpus linguistics: A Guide to the methodology. In *Textbook on Corpus Linguistics*. Language Science Press. <http://langsci-press.org/catalog/book/000>
- Steffensen, S. V. (2025). Surveying ecolinguistics. *Journal of World Languages*, 11(1), 1–49. <https://doi.org/10.1515/jwl-2024-0044>
- Stibbe, A. (2014). An Ecolinguistic Approach To Critical Discourse Studies. *Critical Discourse Studies*, 11(1), 117–128. <https://doi.org/10.1080/17405904.2013.845789>
- Stibbe, A. (2015). *Ecolinguistic: Language, ecology and the stories we live by* (Vol. 1). Routledge.
- Winter, A. K., Le, H., & Roberts, S. (2021). From Black to Blue Skies: Civil Society Perceptions of Air Pollution in Shanghai. *China Quarterly*, 248(1), 1059–1080. <https://doi.org/10.1017/S0305741021000588>