
| RESEARCH ARTICLE

Keywords and Semantic Domains in Annual Global Risks Reports through the Past Decade

Amira Hanafi Elzohiery

Lecturer of Linguistics, Department of English Language and Literature, Faculty of Arts, Cairo University, Cairo, Egypt

Corresponding Author: Amira Hanafi Elzohiery, **E-mail:** amiraelzohiery@cu.edu.eg

| ABSTRACT

Incorporating keyword analysis and semantic domain analysis, the present study linguistically investigated ten Global Risks Reports published by the World Economic Forum (WEF) from 2013 to 2024. The corpus comprised 174,800 word types and 405,366 word tokens. Using the features of Wmatrix and the concordance and collocation tools in Sketch Engine, the study identified the top themes, semantic domains, and patterns prevalent in the ten reports. The semantic domain analysis revealed deep insights into the details of these themes, patterns, and the types of language used compared to the keyword analysis. The results showed the obvious interplay between politics, the economy, technology, and the environment, with minimal attention given to humanitarian and social dynamics. Economic risks appear to be the most pressing risks according to these reports, with every theme ultimately boiling down to the economic situation. Additionally, the reports discussed governments' economics without addressing their impact on citizens and their socioeconomic conditions. This study could be a valuable contribution to the literature that could aid in the tracking of long-term linguistic choices found in risk reports using corpus linguistics. It could also serve as a valuable reference for governments, institutions, or corporations that seek to interpret and analyze global risk reports.

| KEYWORDS

Global Risks; Keyword Analysis, Semantic Analysis, Corpus Linguistics

| ARTICLE INFORMATION

ACCEPTED: 02 November 2024

PUBLISHED: 08 December 2024

DOI: 10.32996/jhsss.2024.6.12.3

1. Introduction

The concept of risk is as old as humanity itself since humans have always expressed their continuous concern over their security and well-being (Karner, 2023). The word "risk" was utilized in the sixteenth and seventeenth centuries to refer to sailing in uncharted waters, but later it was expanded to encompass various aspects of life, particularly in banking and investments (Giddens, 2003). The rise of capitalism, notably during the post-World War II era until the late 1970s, known as the *Golden Age of Capitalism*, further popularized the idea (Glyn et al., 1991). This period emphasized profit and loss calculation and financial risk management, making risk a crucial factor in decision-making processes, policy formulation, and environmental management. Later, a variety of social change factors, including the insurance sector, probability and statistics theory, the development of epidemiology, risk-based management, and decision theory, may have had an impact on the idea of risk over time (Zinn, 2018). Humanity now has an unprecedented ability to significantly jeopardize the conditions necessary for sustaining life on Earth, and this marks a pivotal moment in history (Global Challenges Foundation). Hansson's conceptualization of "risk" encompasses five senses (Hansson, 2004, 2011, and 2013, as cited in Boholm, Möller, & Hansson, 2016). The first sense defines risk as an unwanted event that may or may not occur. The second sense equates risk with the cause of an unwanted event that may or may not occur. The third sense associates risk with the probability of an unwanted event that may or may not occur. The fourth sense defines risk as the statistical expectation value of an unwanted event that may or may not occur. The fifth sense of risk refers to the fact that a decision is made under conditions of known probabilities. However, Boholm, Möller, & Hansson's study (2016) found that in risk analysis, "risk" is often defined quantitatively according to the third and fourth definitions, whereas everyday language sticks to the first and second definitions. In this paper, I report on a corpus analysis of samples of annual global risks

reports issued by the World Economic Forum (WEF) throughout the past decade. I aim to analyze the keywords and semantic domains in these reports to scientifically reach out to the common themes in these reports using tools of corpus linguistics. This could be a valuable contribution to the literature that aids in tracking the long-term linguistic choices found in risk reports using corpus linguistics.

2. Literature Review

Over the past few decades, the term “global risk” has gained significant attention (Kamer, 2023). In the 1980s, two sociologists, Perrow (1984) and Beck (1986), approached the concept of risk from different perspectives in the United States and Europe, respectively. Beck, in his book *Risk Society: Towards a New Modernity* (1986), focused on global risks that arise from industrial and technological advancements and which transcend national borders and affect all social classes. According to him, the risk society is characterized by the unequal distribution of both “goods” exemplified in wealth, and “bads” manifested in pollution, contamination, and other production by-products (Baxter, 2020). These risks are reflexive since they could have been prevented in contrast to the risks that result from the hazards of natural disasters. Perrow, in *Normal Accidents: Living with High-Risk Technology* (1984), highlighted the idea of tightly coupled systems and interdependencies in nuclear power plants, air traffic control systems, and financial markets. These systems are susceptible to the propagation of risks and failures due to their complexity, involving multiple interacting components and feedback loops. Years later, Nassim Nicholas Taleb developed the theory of the Black Swan in his book *The Black Swan* (2007) which refers to the non-existent consequences or events which, based on current understanding, are deemed impossible; however, they defy the laws of nature, disregard risk management and planning, and at the end, unexpectedly happen. For example, this can include positive and negative events such as the prefiltration of smartphones, the dissolution of the Soviet Union, the 2007–2008 financial crisis, and the outbreak of COVID 19.

Unfortunately, hazard mitigation has always failed to garner public support as it requires long-term investments that usually clash with short-term political interests, even though it is more cost-effective to prevent losses than to recover from them (Palmer, 2013). Nevertheless, what happened during the COVID-19 pandemic has undoubtedly necessitated a profound shift in the world’s perception of risk and the approaches to mitigate it. For several years, numerous publications, reports, and studies have consistently alarmed the world to many risks, such as “the Global Risks Report, Lloyd’s City Risk Index 2015–2025, Global Catastrophic Risks, the Cambridge Global Risk Index, and Emerging Risks in the 21st Century, and Natural Disaster Hotspots” (de Amorim & de Andrade Guerra, 2020). Lloyd’s City Risk Index 2015–2025 is a comprehensive study conducted by Lloyd’s, a British insurance market. The study focused on the economic impact of 301 global cities’ exposure to 18 types of risks such as market crashes, cyber-attacks, terrorism, and natural disasters. The Global Catastrophic Risks is an annual report published by the Global Challenges Foundation in Sweden since 2020, which draws attention to the hazards of weapons of mass destruction, pandemics, artificial intelligence, near-Earth asteroids, climate change, and supervolcanic eruptions. Also, the Cambridge Global Risk Index, first published in 2015, provides a comprehensive annual assessment of the economic and financial losses faced by cities as their analysis includes more than 12,000 catastrophe scenarios encompassing 22 threats that have the capacity to significantly disrupt the economic activity in 279 of the world’s most crucial cities. Among the risks publications is *Emerging Risks in the 21st Century: An Agenda for Action* (2003) published by the Organization for Economic Co-operation and Development (OECD), which offered insights into future challenges and potential strategies to manage emerging risks. As for the Natural Disaster Hotspots by the Socioeconomic Data and Applications Center (sedac), it only focuses on natural hazards: “earthquakes, volcanoes, landslides, floods, drought, and cyclones” (Natural Disaster Hotspots, 2006).

As for the Global Risks Report, following the World Economic Forum’s (WEF) Annual Meeting in Davos, Switzerland, in 2006, the WEF decided to begin the tradition of publishing the Global Risks Report and initiate its annual publication until the current year. The WEF is a non-profit organization founded in 1971 that helps to engage leaders from various sectors such as political, business, academic, and civil society to shape agendas and address complex challenges while maintaining independence, impartiality, and high standards of integrity (World Economic Forum, 2024). The WEF has several publications that include topics such as emerging technologies, net-zero industries, cybersecurity, and global cooperation. As for the Global Risks Report (World Economic Forum), it involves five categories of risks: economic, environmental, social, geopolitical, and technological. The reports are based on surveys from business leaders, policymakers, research, think tanks, workshops meet, data analysis, and collaborative partnerships. This study examined the WEF’s annual Global Risk reports from 2013 to 2024. Using corpus linguistic analysis, it analyzed keywords and semantic domains to identify its top themes and recurrent patterns.

2.1. Previous Research on the Language of Reports

Investigating the influence of discourse in the knowledge-driven era of capitalism is significant because language is central to presenting and shaping knowledge and human behavior within the capitalist system, which is often tied to power dynamics (Mayr, 2008). Numerous studies have examined the linguistic practices of reports. Some analyzed the generic features of academic reports such as the study by Veerappan, Bhar, and Chetty (2024) which examined laboratory reports from a genre-

based perspective. Others investigated the linguistic features of environmental reports such as Cui (2024) who studies the relationship between the United Nations Sustainable Development Reports and Apple's Environmental Progress Reports integrating corpus and CDA methods. The analysis specifically examines high-frequency verbs, nouns, and modal verbs. The findings indicated that the United Nations significantly influences corporate reports. In a similar vein, Gagnon, Steven, & Paulo, (2020) examined the lexical choices used in award-winning annual reports using tools of corpus linguistics. Similarly, Fløttum & Dahl's paper (2012) aimed to analyze the polyphonic, epistemic, deontic, and axiological markers in two reports: the Human Development Report 2007/2008 (HDR) and the World Development Report 2010 (WDR), that discussed climate change challenges. Their comparative linguistic analysis showed that the linguistic recourses in each report helped in shaping narratives about climate change from a development perspective. Yeh (2018) studied how women were discursively constructed in World Development Reports from 1998-2018 by analyzing three reports published by the World Bank. Their findings suggested that the image of women changed from that of the passive to the empowered one. Smeuninx, De Clerck, & Aerts (2016) studied the readability of sustainability reports by corporates using NLP analysis and corpus linguistics. Their findings indicated that Australian documents generally exhibit lower readability in comparison to documents from other regions, particularly within specific industries. On the other hand, U.S. documents tend to be positioned toward the more readable side of the spectrum (pp. 73).

Other studies have been devoted to examining corporate annual reports (CARs) from a linguistic perspective. Annual reports have garnered the attention of researchers since the early 1980s (Thomas, 1997). CARs typically present information about the company's operational performance and financial standing, which are significant to the financial community. The linguistic analysis of CARs has evolved significantly. Whereas CARs were historically oriented towards informing stakeholders and investors, CARs now increasingly promote the company's image through evaluative language, using evaluative words such as adjectives, adverbs, and action verbs. (Hussain, Ali, Kasim, & Jalaluddin, 2020). Using a different approach, Liu, Bilal, & Komal, (2022) studied the key topics and nomination strategies between the CEO statements in annual and Corporate Social Responsibility reports from a discourse historical approach. In addition, readability and obfuscation in corporate annual reports have been the focus of some recent studies such as Ponce, González, & Al-Mohareb, (2023); Noh, (2021); Fisher, van Staden, & Richards, (2020); Besuglov & Crasselt, (2020); and De Souza, Rissatti, Rover, & Borba (2019).

2.2. Investigating the Language of Risk

The discourse surrounding the semantics and lexis of risk, danger, and challenge has been the subject of extensive empirical investigation. Utilizing tools of corpus linguistics, several studies have explored how the definitional and semantic meanings of risk are represented in discourse, especially in media texts. For instance, Hamilton, Adolphs, and Nerlich (2007) conducted a corpus-based analysis to elucidate the discursive framing of risk, while the study by Hardy and Colombini (2011) was heavily influenced by Hamilton et al.'s (2007) corpus-based examination of all nominal and verbal uses of the lemma of risk. The latter study employed frequency analysis, concordance contextual analysis, collocation analysis, and an exploratory variation on "distinctive-colligeme analysis". Their findings indicated that the term 'RISK' was overwhelmingly used in a negative medical context, concentrated primarily in magazine and academic discourse.

Some researchers have explored the lemma of the word "risk" itself. For example, Boholm, Möller, and Hansson (2016) examined the behavioral profile of "risk", "safe", and "secure" and their derived adjectives in everyday discourse using corpus linguistics. They concluded that the word "risk" has a non-quantitative, everyday use that differs from its technical and academic applications. Zinn (2010, 2018, and 2020) also investigated the use of "risk" language in media, particularly within *The Times*, through a diachronic study on the discourse of its use in the news. His longitudinal research traced the social changes and explored how institutional and linguistic changes have shaped a new social reality that has been highly influenced by the word "risk". Zinn's findings showed that the 1970s marked the emergence of keywords that co-occurred with "at risk" and "at-risk," such as jobs, lives, children, and patients. He concluded that this reflected the social changes, empirical calculations of the future, technological innovation, ongoing medical research, and chronic health issues.

In a similar vein, Collins (2021) investigated how "risk" about pre-exposure prophylaxis (PrEP) was portrayed in news media in the UK, Ireland, and the US. Using corpus linguistics tools, the study investigated how media narratives reflect ideologies that can result in stigmatization and deter individuals who require treatment from seeking it. The analysis revealed that the concept of "risk" was manifested in a limited set of risk-related terminology within the data, primarily referring to PrEP's ability to "reduce the risk of HIV infection".

Semantics-wise, Boholm (2012) suggested that “risk” is related to agency and decision-making, offering examples of the use of “to risk” as a verb. However, in a later study (2017), Boholm’s reliance on dictionaries, thesauri, and other lexicographic resources rather than empirical studies and expert input resulted in a different lexical focus than specialized risk and safety research. Similarly, Galantino (2021) used a dictionary-based approach to examine how the words “risk”, “danger”, and “threat” were used in German and Italian press coverage of migration in 2015–2016. The findings demonstrated that much risk-related terminology was used, and it linked migration to adverse societal impacts and portrayed it as a security concern.

Merkelsen (2011) also utilized a semantic approach to examine the debate over risk definitions, arguing that the premises of a risk ontology are invalid, as risk and danger are conceptually different, in contrast to the claimed synonymy between the terms. Additionally, Mell and Muller (2021) employed a diachronic corpus-linguistic approach to analyze the use of the lemma of “risk” and “chance”, their typical collocations, and their connection to represented concepts in German parliamentary discourse from 1949 until 2017. The analysis revealed that risk thematization has been increasing, with a noticeable peak in 2009. It also showed that “risk” was mostly used in apocalyptic contexts such as health, economics, and, since the 1980s, technology.

The discourse surrounding the lexical semantics of risk, danger, and challenge has been the subject of extensive empirical investigation. However, the analysis of how the concept of risk is represented in global reports and policy documents remains a significant gap in the existing research. To the best of my knowledge, no study has analyzed the keywords and semantic domains in annual global risks reports especially issued by the WEF to explore their recurrent themes through the past decade. Corpus tools can help in understanding the evolution of patterns and top thematic concerns, and they offer insight into the recurring linguistic patterns that are challenging to yield through other traditional tools (Hunston, 2006).

3. Framework of Analysis

3.1. Keyword Analysis

Corpus linguistics uses empirical analysis to examine a huge body of texts, and it can provide beneficial tools for tackling pressing issues in social sciences such as climate change, criminology, healthcare, and policymaking. (McEnery, & Brookes, 2024). The present study uses keyword analysis which is considered an integral part in corpus linguistics. The word “keyword” was first used as KWIC (keyword in context) to refer to “search tokens” surrounded by lexical items from left and right in the corpus. Later, it was used to refer to keyword analysis after Scott (1996) introduced the Keywords Tool in WordSmith Tools, which has offered an easy method to extract a ranked list of words that are statistically significant in a corpus (Moreno-Ortiz, 2024). Through corpus linguistics tools, keyword analysis can pinpoint words or phrases that appear significantly more often in corpus compared to another corpus (Geluso & Hirsch, 2019). This differs from simple frequency lists (wordlists) that reveal the most commonly occurring lexical items. Frequency lists have been widely used to support language learning and teaching, as they can help focus on the lexical items that can benefit learners (Miller, 2020). Frequency lists (wordlists) can also help in the examination of diachronic variations in language, the most frequent words in a text, comparisons of vocabulary profiles across different texts, the pointing out of n-grams and collocates, and comparisons of language use across different genres or registers. Although a frequency list is a useful quantitative option in corpus linguistics, it is not 100% reliable in some cases. For example, if a frequency list is based on a large online news target corpus such as *The New York Times*, a researcher may claim that a certain word has a high frequency according to the available frequency list, yet this word may be used disproportionately in a certain specialized section of news such as sports reporting or finances and not utilized across all sections. Thus, the high frequency of this word might not reflect its accurate usage in the corpus under investigation (the target corpus); therefore, to gain a more nuanced understanding, there should be a reference corpus that can help in comparing the use of this word in the target corpus to its use in a reference corpus.

Thus, researchers may resort to keyword analysis as it compares the frequency list of the target corpus to the frequency list of a reference corpus (which might be theme-related) in order to highlight the difference in word rank or frequency based on the calculation of the p-value of any difference (Rayson & Potts, 2020). Therefore, any word identified as unusually frequent in the comparison is *key*. Consequently, a keyword list is a ranked set of words that stands out as representative of a corpus (Moreno-Ortiz, 2024). Stubbs (2010) refers to keywords as “tips of icebergs: pointers to complex lexical objects which represent the shared beliefs and values of a culture” (pp. 23). Extracting keywords can serve as a foundation for deeper analysis as they can be classified into the major themes, and in the present study, these themes can lead to pinpointing the recurring themes associated with risk throughout the past decade. This can enable researchers to uncover the concepts and ideologies concealed within thousands or millions of words by analyzing the keyword results. However, there are several limitations to keyword analysis as keywords primarily emphasize lexical differences, rather than semantic, grammatical, or functional differences. Also,

some words may not occur frequently enough to have a significant impact, which results in their exclusion from the keyword list and a tendency to be overlooked (Baker, 2006). For example, a keyword list may yield the word “happiness” as key in a certain text despite the existence of its synonyms such as joy, delight, pleasure, and contentment in the same text. However, these words may appear in the text less frequently; therefore, they might typically be excluded from the keyword list as keyword analysis does not reveal their semantic relationships. Pointing out the key semantic domains, consequently, can help overcome the limitations of keyword analysis and illustrate the recurring patterns in the examined texts.

3.2. Semantic Domains Analysis

While the raw data collected in a corpus are not always sufficient to address various research questions, corpus annotations become necessary. Annotation differs from simply adding metadata to a corpus as it specifically focuses on incorporating linguistic or contextual information to the text (Zufferey, 2020). Annotating a corpus is not an easy task, given the huge amount of data that might be available in the corpus. It depends on the research question, theoretical framework, and objectives of the corpus study. It also needs precision and careful revision from researchers. Although raw data can have their advantages, many types of annotation can enrich the corpus analysis. Tokenization serves as the basic method, classifying a text into individual words or occurrences. Another type is lemmatization which pinpoints and marks each word in the corpus with its base form. Also, part-of-speech tagging provides syntactic and morphological annotation by assigning labels to each tokenized word, indicating its part of speech and including its grammatical category. Syntactic parsing also annotates the text syntactically, which can be done either manually or automatically. In addition, there are phonetic and phonological annotations that depend on transcribing audio data, as well as sign and gesture annotations that can capture non-verbal communication. According to Gries and Berez (2017), semantic annotation, which is the focus of the present study, involves the examination of meanings of word forms in a corpus, often referred to as word sense disambiguation. This process depends on algorithms, the researchers’ knowledge, and various software techniques, as these algorithms assign a sense to each word from an inventory of possible meanings that match the context in which the word is used (Gries, & Berez, 2017). An example of a tool utilized in semantic annotation is Wmatrix which has a rule-based token and Multi-Word Expression semantic tagger. This tool relies on the UCREL Semantic Analysis System (USAS) which is a framework designed for the automatic semantic analysis of text. USAS is a multitier semantic tagging system that classifies words into 21 major semantic categories and 232 unique semantic fields. The complete tag set, along with examples for each semantic field, can be downloaded from the website: [UCREL USAS] (<http://ucrel.lancs.ac.uk/usas/>). For instance, words like “act,” “adventure,” and “approach” are semantically tagged under the category of activity or action, while words such as “actually” and “enormously” fall under the category of boosters (Rayson et al., 2004). The present study employs Wmatrix to pinpoint the semantic domains and generate a keyword list, supplemented by manual filtering to control the tool’s tendency to generate an excessive number of results.

4. Rationale of the Study

Rarely did I read studies that examined the language of global risk reports issued by the World Economic Forum (WEF) over the past decade. Furthermore, few studies investigated the keywords and semantic domains within these reports. Addressing this gap in the literature, therefore, was essential for tracing the recurrent themes and patterns that the ten risk reports might have exhibited during the past decade. Not only is this analysis significant for academic purposes, but also as a valuable reference for governments, institutions, or corporations that seek to interpret and analyze global risk reports.

5. Delimitations of the Study

This study examined ten reports issued by the WEF over the past decade using corpus linguistic tools to conduct keyword analysis and explore key semantic domains. However, neither did the research extend beyond 2013 nor did it include content analysis or the examination of graphs and images within these reports.

6. Research Question

The primary research question that guided the investigation was: What are the common themes and patterns that emerged in the risk reports from the last decade? This question was further broken down into three sub-questions: a) What are the main keywords and themes in the corpus of global risk reports? b) What are the major semantic domains in the corpus of global risk reports? c) What are the implications of the interplay between keyword analysis and semantic analysis?

7. Corpus and Methodology

To collect the specialized corpus for this study, I downloaded the ten global risk reports from the official WEF website: <https://www.weforum.org/publications/>. The reports, available in PDF format from 2013 to 2024, were converted into text files to facilitate the analysis using Wmatrix software. After removing images and refining the data, I uploaded the text files into Wmatrix. The resulting corpus comprised 174,800 word types and 405,366 word tokens. The corpus-based analysis utilized the features of Wmatrix (Rayson et al., 2004) software to generate keyword lists and point out semantic fields. To compare the two corpora, the British National Institutional Corpus (BNC) was utilized as a reference corpus. It is also available in Wmatrix (Rayson et al., 2004), as BNC serves as a prototypical corpus (Gries & Berez, 2017). After generating the list of keywords, I manually grouped them into 20 themes, which were subsequently analyzed in relation to the semantic domains highlighted using the semantic tagger in Wmatrix software. The total frequency of the semantic domains is computed and compared to that in the reference corpus using Log Likelihood (LL). The recurrent semantic categories were categorized and examined in relation to the themes derived from the keyword list. Afterward, the top words/concepts in the two lists were analyzed in context through the concordance and collocation functions (Word Sketch) available in Sketch Engine, a corpus analysis tool. Sketch Engine was developed by Kilgarriff in 2003 and can be accessed online: <https://www.sketchengine.eu> (Kilgarriff et al., 2014).

8. Results and Findings

8.1. Keyword Analysis

To answer the first research question regarding the top-used keywords, I uploaded the text files into Wmatrix and examined the most frequently occurred keywords. The top keywords, ranked by their log-likelihood statistical significance, are presented in Table 1. As a first step, it can be noticed that the top keywords reflect interconnected global challenges, such as the "economic" crisis, which can cause "social" unrest. Furthermore, environmental and energy dangers that may lead to infrastructure failures were exemplified in keywords such as "environmental" and "energy". Also, the top keyword list revealed the international scope of these challenges, as demonstrated by terms such as "global," "world," and "geopolitical." Keywords such as "risks," "crisis," and "failure" highlight the role of these reports in emphasizing the necessity for effective management strategies to mitigate adverse impacts on society, the economy, and the environment. Additionally, keywords such as "social," "human," and "civil" show the inclusion of individual well-being. Also, including "technologies" indicates that technological advancements have been part of the dangers that may benefit or complicate the aforementioned challenges.

Table 1. Top Keywords for the WEF Global Risks Reports with Frequency

No.	Keyword	Frequency	Log Likelihood
1.	risks	486	1460.61
2.	global	474	1458.78
3.	economic	297	790.05
4.	crises	159	489.43
5.	failure	143	428.65
6.	infrastructure	142	425.58
7.	world	171	319.13
8.	report	212	312.98
9.	environmental	119	310.38
10.	major	152	310.24
11.	social	142	304.81
12.	human	103	284.78
13.	technologies	92	283.14
14.	large-scale	82	252.36
15.	civil	87	249.58
16.	society	80	235.91
17.	governance	75	230.82
18.	geopolitical	74	227.74
19.	energy	77	226.76

As a second step, I sought to expand the analysis by examining all the yielded keywords and categorizing them into thematic groups to answer the second part of the first research sub-question which is pinpointing the thematic patterns of the WEF reports. Following a thorough analysis, I refined the list by excluding function words that were not pertinent to the study, ultimately grouping the remaining keywords into twenty themes, as illustrated in Table 2. These themes were organized according to the frequency of their associated keywords. The resultant twenty themes (as represented in Table 2) included the management of crises and the political factors that influence social and economic stability, as seen in the themes of *Crisis Management*, *Political Landscape*, *Conflict and Security*, and *Economic Challenges*. Some themes reflected the social structure and ethical dynamics, such as *Social Dynamics*, *Human Rights and Ethics*, *Socioeconomic Inequality*, and *Cultural and Societal Change*. Another group of themes addressed environmental sustainability and resources, as illustrated by *Environmental Concerns*, *Energy and Resources*, *Natural Disasters*, and *Infrastructure*. Additionally, the reports included themes related to technological and informational impacts, exemplified by *Emerging Technologies*, *Technological Impact*, *Data and Information*, and *Information and Media*. Finally, there were themes related to future trends and market adaptation, including *Market Trends*, *Future Trends*, *Urbanization*, and *Health and Wellbeing*. It is evident from Table 2 that throughout the ten years, the top recurrent themes were *Crises Management*, *Political Landscape*, *Conflict and Security*, *Economic Challenges*, and *Infrastructure*.

Table 2. All Themes with Keywords and their Frequencies

No.	Theme	Keywords with their Frequencies
1.	Crisis Management	risks (486), report (212), crises (159), major (152), geopolitical (74), instability (48), challenges (39)
2.	Political Landscape	global (474), world (171), international (78), governance (75), political (69), power (12), decision-makers (10), democracy (3)
3.	Conflict and Security	failure (143), civil (87), geopolitical (74), conflict (73), instability (48), security (45), military (35), attacks (33), crime (24), unrest (20), tensions (16), violence (10)
4.	Economic Challenges	economic (297), unemployment (75), fiscal (65), debt (29), deflation (30), inflation (23), liquidity (19), resources (17), asset (14), poverty (8), recession (2)
5.	Infrastructure	infrastructure (142), large-scale (82), systems (81), governance (75), institution (39), development (37), organizations (22), planning (16), management (14), services (11), regulatory (10)
6.	Environmental Concerns	environmental (119), water (93), climate (76), weather (43), resources (35), natural (31), biodiversity (21), ecosystems (19), floods (19)
7.	Technological Impact	technologies (92), networks (80), data (45), cyber (37), cyberattacks (29), applications (23), digital (18), computing (17), machine (11), innovations (11), robotics (10), automation (9), algorithms (3)
8.	Social Dynamics	social (142), society (80), migration (42), rights (32), populations (22), community (15), civic (13), freedom (13), diversity (8), inequality (6)
9.	Market Trends	financial (84), economy (73), prices (71), trading (26), commodities (13), investments (12), consumer (12), capital (7)
10.	Energy and Resources	water (93), energy (77), resource (35), oil (22), supply (21), sustainability (5), renewable (3), consumption (3)
11.	Emerging Technologies	networks (80), applications (23), advanced (22), innovations (20), digital (18), computing (17), automation (9), smart (10), robotics (10), connectivity (9)
12.	Data and Information	systems (81), information (62), data (45), digital (18), storage (10), monitoring (8), platforms (8)
13.	Future Trends	impacts (67), potentials (52), trends (30), implications (20), insights (10), changes (8), scenarios (6), shifts (6), evolution (5), transformations (4)
14.	Natural Disasters	catastrophes (45), weather (43), storms (38), disasters (23), floods (19), tsunamis (10), drought (5)
15.	Human Rights and Ethics	human (103), freedom (13), movements (10), human rights (10), protections (10), refugee (7), inequality (6)
16.	Urbanization	growth (24), transportation (23), planning (16), community (15), cities (15), services (12), investments (12), urbanization (9), housing (9), mobility (8), sustainability (5), demographics (2)

17.	Socioeconomic Inequality	income (38), underemployment (29), opportunities (22), resources (17), disparity (16), poverty (8), wealth (8), barriers (5), socioeconomic (4), access (3)
18.	Information and Media	information (62), communication (10), news (7), media (6), misinformation (3), censorship (2)
19.	Health and Wellbeing	health (33), infectious (22), chronic (13), treatments (10), nutrition (10), healthcare (8)
20.	Cultural and Societal Change	movements (10), diversity (8), awareness (8), norms (3), engagement (3), expressions (3), representations (2)

8.2. Semantic Domains Analysis

To address the second research question and expand the semantic analysis of the top keywords and their themes, the ten documents were uploaded to Wmatrix and compared with the British National Corpus (BNC) institutional reference corpus to obtain the most frequent semantic domains. This comparison yielded the top semantic categories, which were ranked according to their Log Likelihood in Table 3 (See Figure 1). This approach differs from keyword and thematic analysis, as those methods do not capture the different semantic equivalents of the top keywords. Subsequently, I manually matched the top 20 semantic domains yielded with themes I obtained from keyword analysis to gain a broader understanding of the dominant patterns associated with these domains (see Table 4). This resulted in fewer themes than the ones yielded through keyword analysis. In the following analysis, I examined the key semantic domains and their related keywords, supported by examples from the concordance lines relevant to each theme. I had grouped the eight themes into four major categories under which I discussed the relevant themes and examples.

Table 3. Top Semantic Fields for the WEF Global Risks Reports ordered according to the Log Likelihood

No.	Semantic Domain	Frequency	Log Likelihood
1.	Danger	913	2491.38
2.	Numbers	1581	1276.34
3.	Science and technology	305	788.06
4.	Cause and Effect	632	738.06
5.	Business	482	737.32
6.	Cheap	304	731.38
7.	The Universe	285	612.35
8.	Weather	210	564.16
9.	Geographical terms	372	526.11
10.	Change	468	463.16
11.	Belonging to a group	587	437.20
12.	Green Issues	197	413.54
13.	Important	366	341.56
14.	Difficult	317	320.63
15.	Money	315	311.13
16.	Damaging and destroying	200	303.51
17.	government	410	300.49
18.	Hindering	131	211.90
19.	evaluation: bad	76	209.84
20.	Violent/Angry	131	153.30

Figure 1 Word Cloud of the Prominent Semantic Domains in the Corpus



Table 4. Top Themes and their Semantic Domains using Wmatrix 5

No.	Theme	Semantic Domains and their Frequencies
1.	Crisis Management	danger (913), cause and effect (632), evaluation: bad (76), important (366)
2.	Economic Challenges	numbers (1581), money (315), business (482), cheap (304), change (468)
3.	Technological Impact	science and technology (305)
4.	Infrastructure	geographical terms (372)
5.	Environmental Concerns	the universe (285), weather (210), green issues (197)
6.	Political Landscape	government (410)
7.	Conflict and Security	damaging and destroying (200), violent/angry (131), hindering (131), difficult (317)
8.	Human Rights and Ethics	belonging to a group (587)

8.2.1 Economic and Infrastructure Challenges

By aligning the primary semantic domains with the themes identified through keyword analysis, several prevalent themes emerged. *Economic Challenges* were illustrated by semantic domains such as “numbers,” “money,” “cheap,” “business,” and “change.” Within the “money” category (315 instances), keywords like “financial,” “liquidity,” “currency,” “banks,” “loss,” “debt,” “spending,” “paying,” and “expenditure” stood out. These indicate that economic risks have been a major focus in the reports over the past decade, significantly impacting global stability. An examination of the concordance lines revealed that keywords such as “debt” and “loss” reflect broader economic crises (see Figure 2). Notably, the top collocate for “debt” is “public,” which illustrates how the governments’ borrowing levels might have a disastrous effect (See Example 1), and this finding is further supported by the concordance lines of the keyword “financial” (see Figure 3). These patterns suggest that economic challenges, especially those linked to public debt, have posed a significant risk and have been expected to result in a financial crisis similar to the catastrophe of 2008 (See Example 2).

Infrastructure is another theme that posed a global challenge, which was represented by semantic domains that included “geographical terms,” highlighting keywords that involved natural disasters or hydrological hazards such as “eruptions,” and “tide;” coastal areas such as “ocean,” “river;” “channels;” “coastal;” and “horizons;” and terrestrial areas such as “island;” and “mountains”. These keywords highlight the potential impact of the physical environment on infrastructure. Example 3 explicitly mentions climate change and its potential effects on ocean damage, which could impact ecosystems and human activities. They were consistent with the recurrent emphasis on environmental risks in most of the semantic domains (See Section 8.2.4).

- (1) They are also fuelled by the high levels of **public debt** in Japan and the US, where political gridlock has exacerbated perceptions.
- (2) Their social, economic and political fallouts could be far-reaching, as exemplified by the continuing impacts of the **financial** crisis of 2007-2008.

- (3) *The risk multiplier that climate change presents to water shortages, biodiversity loss, **ocean** damage and deforestation also creates a complex "heterarchy", rather than a simple hierarchy of environmental risks.*



Figure 2 The Concordance Lines of the Keyword "debt"



Figure 3 The Concordance Lines of the Keyword "financial" and its Top Collocate "crisis"

8.2.2 Political and Security Factors

The prevalent themes within this category include the *Political Landscape*, characterized by semantic domains such as "government," which involved keywords such as "political," "coup," "geopolitics," "democratic," "activists," "revolutions," "policy-maker," "nationalism," and "anti-terrorism". By examining the collocates of the word "political", it appeared that the top nouns modified by "political" were "deadlock", "polarization", and "instability" (See Figure 4), which suggests that the political atmosphere has been usually characterized by conflict and division. It also frequently collocates with "social", "religious", "economic", and "domestic". This indicates how the political risks have been multifaceted and intertwined with the social, economic, and cultural ones as in Examples 4, 5, and 6.

- (4) *...with strong links between this macroeconomic risk nexus and social issues, such as rising income inequality and **political** and social instability.*
- (5) *...scattered or isolated terrorist attacks carried out by individuals or non-state groups with ideological, **political** or religious goals, resulting in loss of life, severe injury and/or material damage Weapons of mass destruction.*
- (6) *Consequently, "resilience" has become a key policy and agency theme to counteract the growing sense of economic, **political** and social risk that changing climatic conditions pose.*

Another prevalent, related theme is *Conflict and Security* which were illustrated by domains including "damaging and destroying," "violent/angry," and "crime", involving keywords such as "collapse", "destruction", "broken", "harm", "accidents", "devastating", "wrecking", and "slashed". By examining the profile of the most frequent word "collapse" (493 times), it appeared that it mostly collocated with the word "state" followed by "ecosystem", which might indicate the interconnectedness between the environmental and political aspects as this could lead to environmental degradation and resource scarcity which in turn could destabilize governments and societies. Figure 5 illustrates this growing concern regarding the resilience of biodiversity as a result of ecosystem collapse.

The domain of "violent/angry" included "abuse", "assault", "fallout", "threaten", "fighting back", "vicious", "force", "hit", "attack", "riots", and "unrest". In Example (7), the use of the adjective "vicious" highlights the destructive and relentless nature of the relationship between climate change and nature loss. In Example (8), the noun "fallout" suggests the forceful and damaging

effects of the financial crisis. Similarly, in Example (9), the noun “abuse” suggests the possibility of harm or exploitation of technology. This corroborates the prevalence of these negative sentiments in risk reports, which can be there due nature of these reports and the complexity of global challenges.

- (7) Fifth, climate change exacerbates nature loss, which in turn reduces nature's resilience to climate change—a **vicious** circle.
- (8) When they were hit by the **fallout** of the financial crisis, they had the fiscal leeway to stimulate economic activity through the opposite of austerity..
- (9) Technological concerns include data mismanagement, loss of privacy, increase in surveillance, and possible **abuse** of new and more complex information technology, which is further explored in Part 2.4.

Additionally, *Crisis Management* was represented through semantic domains such as “danger”, “cause and effect”, “evaluation: bad”, and “important”. The word “danger” topped all the semantic domains (913 instances), and it included keywords such as “high-risk”, “jeopardize”, “pitfalls”, “risky”, “endanger”, “hazards”, “at risk”, “risk”, and “risks.” This was followed by the semantic domain of “cause and effect” which was significantly different with (a LL of 738.06.) This could be due to the nature of these risks reports; therefore, this semantic domain had tokens such as “attribute,” “dependent upon,” “depends/on,” “endanger,” “entail,” “give rise to,” “implications,” “influence,” “in response to,” “in connection with,” and “interconnect”, “link”. Unsurprisingly, the risk reports are saturated with this cautionary vocabulary. A collocational analysis of “risk” revealed its primary association with “global” (as in the reports’ title), followed by “environmental,” “systematic,” “economic,” and “geopolitical” (see Figure 6). Environmental risks, including water crises, natural disasters, biodiversity loss, and food shortages, were identified as the most significant threats (see Figure 7).



Figure 4 The Collocates “Political” and “Instability”



Figure 5 The collocates "ecosystem" and "collapse" and their relation to "biodiversity"

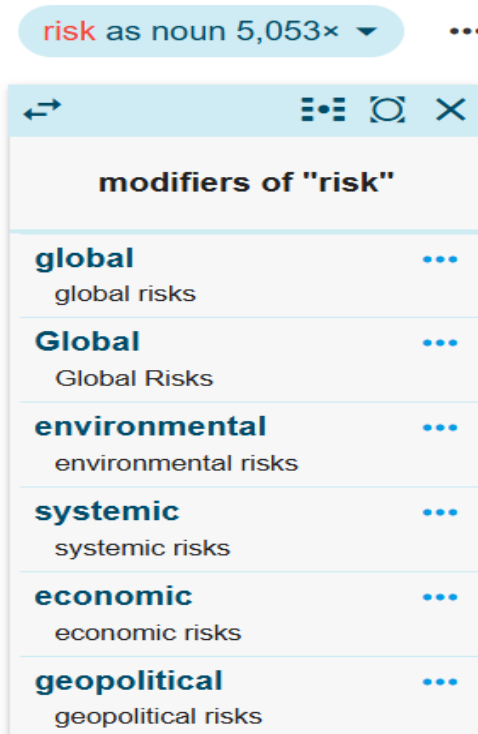


Figure 6 The Top Collocates of the Keyword "risk"



Figure 7 The Two Collocates "Environmental and Risk"

8.2.3 Technological Advancements

Continuing the pattern observed in previous themes, the dominant themes in this category encompassed *Technological Impact*, exemplified by semantic domains such as “science and technology” (305 instances) and represented in keywords such as “biotechnology”, “chemical weapons”, “experiments”, “chemistry”, “genetic and engineering”, “high tech”, “nuclear weapons”, and “techno-science”. The most frequently used word was “weapons” (238 times), primarily in the context of “chemical weapons” and “nuclear weapons.” The analysis revealed a consistent focus on the dangers of weapons of mass destruction (See Figure 8) and autonomous weapons, particularly highlighting concerns about Russia’s potential use of nuclear weapons, as illustrated in Example (10). In addition, the reports issued warnings regarding the potential negative consequences of the Fourth Industrial Revolution (4IR) technologies, as illustrated in Example (11).

(10) As an energy power, and possessing the world's second largest arsenal of **nuclear weapons**, Russia will continue to play an assertive role in the geopolitical order.

(11) There are many calls for deeper engagement on questions of ethics in the development and use of **4IR technologies**.

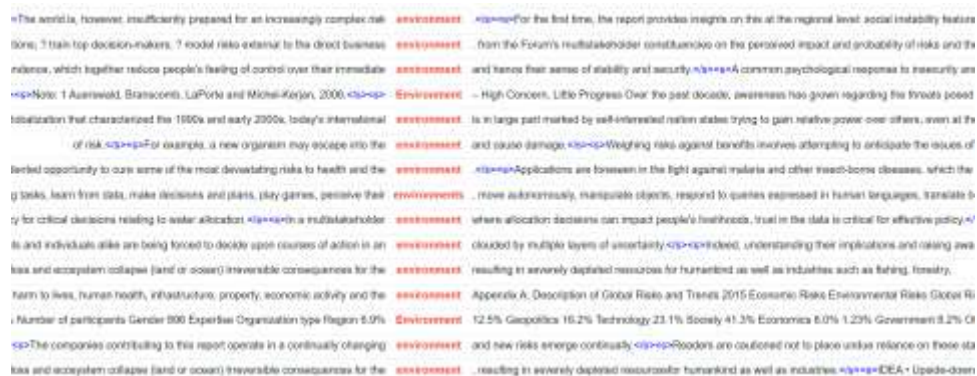


Figure 9 Concordance Lines for "environment"

9. Discussion

The interplay between keyword analysis and semantic analysis yielded a number of implications. Although the keyword analysis revealed a diverse range of themes, such as crisis management, political landscape, and social dynamics, the semantic domain analysis highlighted the predominance of economic, political, technological, and environmental risks, with a lesser focus on humanitarian, cultural, and social aspects. While economic, political, technological, and environmental risks were prioritized in both analyses, the effects of these risks on human rights, societal change, health, and socio-economic equality appeared to be marginalized according to the analysis. In the face of political and economic risks, the emphasis on human well-being should have been a prime concern; nevertheless, the analysis of the semantic domains of the reports implied a potential detachment between the top risks and their immediate and tangible impact on human lives. Both keyword analysis and semantic analysis were common in the top themes. On one hand, keyword analysis highlighted the “economic” file as one of the top-used keywords. On the other hand, the semantic domain analysis highlighted “numbers” and “business” as the most frequently occurring semantic domains. Both lists placed environmental concerns afterward. Nevertheless, the keyword analysis included the word “human” as one of the top-used keywords, but the effect of the risks on human beings found a lower place in the semantic domain analysis. The concordance analysis showed that the “human” and “belonging to a group” keywords were largely used descriptively to discuss issues related to institutions, corporations, or federations.

Prioritizing the hazards of economic and environmental issues expands upon the global dangers mentioned by Beck in his book *Risk Society: Towards a New Modernity* (1986). They agree that these dangers transcend national borders, but their effect on all social classes, an issue Beck emphasized in his book, was not clearly illustrated in the language of the reports under investigation. Also, the risks prioritized in the reports are in line with Perrow’s (1984) perception of dangers exemplified in nuclear power plants and financial markets. It is alarming that these risks have persisted for over forty years. However, the collocations of “risk” in the analyzed reports contrasted with Zinn’s findings (2010, 2018, 2020) which highlighted its linguistic association with “jobs,” “lives,” “children,” and “patients.”

When it comes to the language used, the semantic domain analysis highlighted the use of evaluative language which is similar to the use of evaluative language employed in CARs as Hussain, Ali, Kasim, & Jalaluddin (2020) illustrated in their research on the language of corporate annual reports. However, the evaluative language of the reports was cautionary and depressing. Keyword analysis merely focused on keywords related to crisis management, such as “risks,” “report,” “crises,” “major,” “geopolitical,” “instability,” and “challenges.” In contrast, the semantic analysis revealed a saturation of cautionary vocabulary in the reports, including words like “pitfalls,” “risky,” “endanger,” “hazards,” “at risk,” and “risks.” In addition, grim and ominous vocabulary was used in keywords such as “collapse,” “destruction,” “broken,” “harm,” “accidents,” “devastating,” “wrecking,” and “slashed”. This apocalyptic view echoes Mell and Muller’s (2021) findings that linked risk thematization to apocalyptic semantics in health, economics, and, since the 1980s, technology.

One of the main patterns that was evident throughout the analysis was the tight integration of the top four recurrent risks throughout the past decade: economic, political, environmental, and technological. Keyword analysis ranked “economic” as the third most frequently used word, with a Log Likelihood (LL) score of 790.05. A manual investigation of economy-related keywords revealed that “economic”, “unemployment”, “fiscal”, “debt”, “deflation”, “inflation”, “liquidity”, “resources”, “asset”, “poverty”, and “recession” were recurrent keywords in the reports over the past decade. Similarly, the semantic analysis ranked fields related to economic challenges highly, and that included fields of “numbers”, “money”, “business”, “cheap”, and “change”. This highlights the impact of business and money on global stability. Interestingly, both analyses highlighted “debt” as a major global economic risk. Further examination of the concordance lines indicated that “public debt” was one of the major

contributors to these economic challenges. An analysis of the collocational profiles of keywords associated with the rest of the themes like “political,” “fallout,” “collapse,” “environment,” and “community” revealed their semantic association with the word “economic.” While the keyword analysis highlighted limited references to risks related to unemployment, poverty, and socioeconomic disparity, these keywords were absent from the semantic analysis. This confirms again the minimal attention paid to the immediate and tangible impact of these risks on humans.

Another pattern that was common between the two analyses was the occurrence of the word “infrastructure.” Although in the keyword analysis, the word “infrastructure” was ranked as one of the top recurrent keywords, it did not appear as a keyword according to the semantic analysis. However, it appeared as a theme related to geographical areas. A collocational analysis of the words associated with “infrastructure” revealed that it was mostly used with “information” infrastructure. This might explain why the word itself did not appear in the semantic domain analysis, yet it emerged as a theme that yielded semantic fields such as “geographical terms”, which involved natural disasters or hydrological hazards like “eruptions” and “tide”, as well as coastal areas including “ocean,” “river,” “channels,” “coastal,” and “horizon.” This illustrates that the focus on infrastructure in global risks reports is less about man-made infrastructures and more about environmental factors.

Although both analyses ranked the political landscape as one of the top themes, the semantic analysis revealed more details about the characteristics of its atmosphere exemplified by keywords such as “coup,” “geopolitics,” “democratic,” “activists,” and “revolutions.” Further analysis revealed associations with keywords such as “polarization” and “deadlock.” Also, both analyses prioritized technological advancement as one of the top risks, but it ranked lower than economic ones. The semantic analysis revealed the types of these hazards, exemplified by “biotechnology,” “chemical weapons,” “experiments,” “chemistry,” “genetic engineering,” “high tech,” and “nuclear weapons.” However, the keyword analysis revealed another type with fewer instances which was “cyber attacks.” This suggests a prevailing concern with traditional technological risks in comparison to the emerging technological risks. Another pattern that yielded similar keywords was the environmental risks which encompassed natural disasters and human-made hazards.

10. Conclusion

Although there was extensive research on the lexical semantics of risk, danger, and challenge, no studies have investigated how risk is represented in global reports and policy documents from a linguistic perspective. Therefore, this study linguistically examined the WEF’s annual Global Risk Reports from 2013 to 2024. Using corpus linguistic analysis, it analyzed ten reports to pinpoint their common themes and recurrent patterns through keyword analysis and semantic domain analysis. Both methods helped in highlighting the main themes and semantic domains of the WEF reports over the past decade; however, the semantic domain analysis showed deeper insights regarding the details of these themes, the patterns, and the type of language used. The analysis showed how deep the interplay between politics, economy, technology, and the environment was, with minimal attention given to humanitarian and social dynamics. The reports did not emphasize the tangible risks affecting human beings. Economic risks appeared to be the top pressing risks according to these reports, with every theme ultimately boiling down to the economic situation. However, the reports discussed governments’ economics without addressing their impact on their citizens and their socioeconomic conditions. In addition, as expected, the evaluative language of the reports was cautionary and ominous. Examples from the concordances align, to some extent, with Hansson’s first and second conceptualizations of the term “risk” (Hansson, 2004, 2011, and 2013, as cited in Boholm, Möller, & Hansson, 2016), which defined risk as an unwanted event that may or may not occur. The second sense equated risk with the cause of such unwanted events. This study could make a valuable contribution to the literature by helping trace the long-term linguistic choices employed in risk reports through keyword analysis and semantic domain analysis. Although the WEF has announced that its risks reports include five categories of risks: economic, environmental, social, geopolitical, and technological, analyzing the language of the reports of the past decade can highlight the WEF’s bias and priorities in its focus on specific types of risks over others. Future analyses could explore how these linguistic choices and thematic patterns have influenced policy decisions and public opinion in recent years.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Ahling, Y. (2018). <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1482519>. Retrieved April 14, 2024, from DIVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1482519&dsid=-2369>
- [2] Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- [3] Baxter, J. (2020). Health and Environmental Risk. Kobayashi, A. (Ed.), *International Encyclopedia of Human Geography* (Vol. 2, pp. 303-307), Elsevier.
- [4] Beck, U. 1992 [1986]. *Risikogesellschaft. Auf dem Weg in eine andere Moderne [Risk society: Towards a new modernity]*. Frankfurt am Main: Suhrkamp.
- [5] Besuglov, B. & Crasselt, N. (2020). The effect of readability and language choice in management accounting reports on risk-taking: an experimental study. *Journal of Business Economics*, 91, 5–33. <https://doi.org/10.1007/s11573-020-00980-4>
- [6] Boholm, M., Möller, N. and Hansson, S.O. (2016), The Concepts of Risk, Safety, and Security: Applications in Everyday Language. *Risk Analysis*, 36: 320-338. <https://doi.org/10.1111/risa.12464>
- [7] Boholm, M. (2018). Risk and quantification: A linguistic study. *Risk Analysis*, 36 (6), 1243-1261. <https://doi.org/10.1111/risa.13258>
- [8] Boholm, M. (2017). The semantic field of risk. *Safety Science*, 92, 205-216. <https://doi.org/10.1016/j.ssci.2016.10.011>
- [9] Boholm, M. (2012). "The Semantic Distinction between "Risk" and "Danger": A Linguistic Analysis." *Risk Analysis: An Official Publication of the Society for Risk Analysis* 32 (2): 281–293. doi:10.1111/j.1539-6924.2011.01668.x.
- [10] Collins, L. C. (2021). Pre-exposure prophylaxis (PrEP) and 'risk' in the news. *Journal of Risk Research*, 25(3), 379–394. <https://doi.org/10.1080/13669877.2021.1894470>
- [11] Cui, C. (2024). A three-dimensional discourse analysis of SDGs reports and Apple's environmental progress reports: A corpus-based study. *Advances in Social Science, Education and Humanities Research*, 229-234. https://doi.org/10.2991/978-2-38476-222-4_27
- [12] De Amorim, W. S., & de Andrade Guerra, J. B. S. O. (2020). Pandemics, global risks and adaptation: Challenges for a changing world. *Research in Globalization*, 2, 100023. <https://doi.org/10.1016/j.resglo.2020.100023>
- [13] De Souza, J., Rissatti, J., Rover, S., & Borba, J. (2019). The linguistic complexities of narrative accounting disclosure on financial statements: An analysis based on readability characteristics. *Research in International Business and Finance*, 48, 59-74.
- [14] Fisher, R., van Staden, C.J., & Richards, G. (2020). Watch that tone: An investigation of the use and stylistic consequences of tone in corporate accountability disclosures", *Accounting, Auditing & Accountability Journal*, 33(1), 77-105. <https://doi.org/10.1108/AAAJ-10-2016-2745>
- [15] Fløttum, K., & Dahl, T. (2012). Different contexts, different "stories"? A linguistic comparison of two development reports on climate change. *Language & Communication*, 32(1), 14–23. <https://doi.org/10.1016/j.langcom.2011.11.002>
- [16] Gagnon, J., Young, S., & Alves, P. (2020). The linguistic properties of award-winning annual reports. *SSRN*. <http://dx.doi.org/10.2139/ssrn.3575679>
- [17] Galantino, M. G. (2021). The use of risk language in migration discourse. A comparative exploration of German and Italian newspapers. *Journal of Risk Research*, 25(3), 331–346. <https://doi.org/10.1080/13669877.2021.2001673>
- [18] Geluso, J., & Hirsch, R. (2019). The reference corpus matters: Comparing the effect of different reference corpora on keyword analysis. *Register Studies*, 1(2), 209–242.
- [19] Giddens, A. (2003). *Runaway World: How Globalization is Reshaping our Lives*. New York: Routledge.
- [20] Global Challenges Foundation, 2024. Global Risks: Definition.
- [21] Glyn, A. Armstrong, P., & Harrison, J. (1991). *Capitalism since 1945*. Basil Blackwell.
- [22] Gries, S., & Berez, A. (2017). Linguistic Annotation in/for Corpus Linguistics. In: N. Ide and J. Pustejovsky (eds.), *Handbook of Linguistic Annotation* (pp. 379-409). Springer Science+Business Media, Dordrecht. doi 10.1007/978-94-024-0881-2_15
- [23] Gutiérrez Ponce, H., Chamizo González, J., & Al-Mohareb, M. (2023). Annual reports readability from linguistic and communication perspectives: systematic literature review. *Business and Professional Communication Quarterly*, 86(4), 446-497. <https://doi.org/10.1177/23294906231151893>
- [24] Hamilton, C., Adolphs, S., & Nerlich, B. (2007). The meanings of 'risk': a view from corpus linguistics. *Discourse & Society*, 18(2), 163-181. <https://doi.org/10.1177/0957926507073374>
- [25] Hardy, D., & Colombini, C. (2011). A genre, collocational, and constructional analysis of RISK. *International Journal of Corpus Linguistics*, 16 (4), 462–485. <https://doi.org/10.1075/ijcl.16.4.02har>
- [26] Hunston, S. (2006). Corpus linguistics. *Encyclopedia of Language & Linguistics (Second Edition)*. <https://doi.org/10.1016/B0-08-044854-2/00944-5>
- [27] Hussain, S. S., Ali, A. M., Kasim, Z. M., Jalaluddin, I. (2020). A review on the rhetorical structure and linguistic features of corporate annual reports. *International Journal of Academic Research in Business and Social Sciences*, 10(9), 236-256. <http://dx.doi.org/10.6007/IJARBS/v10-i9/7591>
- [28] Karner, C. (Ed.). (2023). *Research Handbook on the Sociology of Globalization*. Edward Elgar Publishing Limited.
- [29] Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). *The Sketch Engine: Ten years on*. *Lexicography*, 1(1), 7-36.
- [30] Liu, Q., Bilal, & Komal, B. (2022). A Corpus-Based Comparison of the Chief Executive Officer Statements in Annual Reports and Corporate Social Responsibility Reports. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.851405>

- [32] Mayr, A. (2008). *Language and Power: An Introduction to Institutional Discourse (Advances in Sociolinguistics)*. London: Continuum
- [34] McEnery, T., & Brookes, G. (2024). Corpus linguistics and the social sciences. *Corpus Linguistics and Linguistic Theory*. <https://doi.org/10.1515/cllt-2024-0036>
- [35] Miller, D. (2020). Analysing Frequency Lists. In: Paquot, M., Gries, S.T. (eds) *A Practical Handbook of Corpus Linguistics* (pp. 77-97). Springer, Cham. https://doi.org/10.1007/978-3-030-46216-1_4
- [36] Merkelsen, H. (2011). The constitutive element of probabilistic agency in risk: a semantic analysis of risk, danger, chance, and hazard. *Journal of Risk Research*, 14(7), 881–897. <https://doi.org/10.1080/13669877.2011.571781>
- [37] Moreno-Ortiz, A. (2024). Keywords. In: Moreno-Ortiz, A (eds) *Making Sense of Large Social Media Corpora*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-52719-7_4
- [38] Müller, M., & Mell, R. M. (2021). 'Risk' in political discourse. A corpus approach to semantic change in German Bundestag debates. *Journal of Risk Research*, 25(3), 347–362. <https://doi.org/10.1080/13669877.2021.1913631>
- [39] Nicholas, T. (2007). *The Black Swan*. Penguin Books
- [40] Noh, M. (2021). Culture and annual report readability. *International Journal of Accounting & Information Management*, 29 (4), 583-602. <https://doi.org/10.1108/IJAIM-05-2021-0094>
- [41] Palmer, L. (2013). Providing aid before climate disasters strike. *Nature Climate Change*, 3, 857–858. <https://doi.org/10.1038/nclimate2016>
- [42] Perrow, C. 1984. *Normal accidents, living with high-risk technology*. Princeton, NJ: Princeton University Press.
- [43] Rayson, P., Archer, D., Piao, S., & McEnery, A. M. (2004). The UCREL semantic analysis system. In Proceedings of the beyond named entity recognition semantic labelling for NLP tasks workshop, Lisbon, Portugal, (pp. 7-12).
- [44] Rayson, P., & Potts, A. (2020). Analyzing Keywords Lists. In: Paquot, M., Gries, S.T. (eds), *A Practical Handbook of Corpus Linguistics* (pp. 119-139). Springer, Cham. https://doi.org/10.1007/978-3-030-46216-1_4
- [45] Smeuninx, N., De Clerck, B., & Aerts, W. (2016). Measuring the Readability of Sustainability Reports. *International Journal of Business Communication*, 57(1), 232948841667545. <https://doi.org/10.1177/2329488416675456>
- [48] Socioeconomic Data and Applications Center (sedac) (2024). Natural Disaster Hotspots.
- [49] Stubbs, M. (2010). Three Concepts of Keywords. In: Marina Bondi and Mike Scott (eds) *Keyness in Texts* (pp. 21-42). John Benjamins Publishing Company, Amsterdam/Philadelphia.
- [50] Thomas, J. (1997). Discourse in the marketplace: the making of meaning in annual reports. *The Journal of Business Communication*, 34, 47–66. <https://doi.org/10.1177/002194369703400103>
- [52] Veerappan, V., Bhar, S., and Chetty, T., (2024). Debunking method section of tertiary electrical engineering laboratory reports. Zainudin, M. and Rahmat, H. (Eds.), *Proceedings of the 4th International Conference on Communication, Language, Education and Social Sciences (CLESS 2023)*, Advances in Social Science, Education and Humanities Research (819), https://doi.org/10.2991/978-2-38476-196-8_14
- [53] World Economic Forum, 2024. About Us: Our Mission.
- [54] Zinn, J., & McDonald, D. (2015). Changing discourses of risk and health risk: A Corpus analysis of the usage of risk language in the New York Times. In *John Martyn Chamberlain* (Eds.), *Medicine, Risk, Discourse and Power* (pp.207-240). Routledge - Taylor & Francis.
- [55] Zinn, J. O. (2018). The Proliferation of 'at risk' in *The Times*: A Corpus Approach to Historical Social Change:1785-2009. *Historical Social Research / Historische Sozialforschung*, 43(2 (164)), 313–364. <http://www.jstor.org/stable/26454291>
- [56] Zinn, J. O. (2020). Health and illness as drivers of risk language in the news media – a case study of The Times. *Health, Risk & Society*, 22(7–8), 437–455. <https://doi.org/10.1080/13698575.2020.1862065>
- [57] Zinn, J. (2010). Risk as discourse: Interdisciplinary perspectives. *Critical Approaches to Discourse Analysis across Disciplines*, 4 (2), 106-124. <http://cadaad.net/journal>
- [58] Zufferey, S. (2020). How to Annotate a Corpus. In: S. Zufferey (ed.), *Introduction to Corpus Linguistics* (pp. 163-193). <https://doi.org/10.1002/9781119779728.ch7>