Linguistic Research 42(2): 399-422
DOI: 10.17250/khisli.42.2.202506.006

A diachronic analysis of IMO Assembly Resolutions (1959-2023) using Gries' new keyword approach*

Mengjiao Suo[a] · Yaochen Deng[b] · Hyunjong Hahm[c] · Se-Eun Jhang[a***]**
(National Korea Maritime and Ocean University[a] · Dalian University of Foreign Languages[b] · University of Guam[c])

Suo, Mengjiao, Yaochen Deng, Hyunjong Hahm, and Se-Eun Jhang. 2025. A diachronic analysis of IMO Assembly Resolutions (1959-2023) using Gries' new keyword approach. *Linguistic Research* 42(2): 399-422. This study employs the hybrid keyword extraction method proposed by Gries (2021) to analyze diachronic semantic changes of key maritime terms in International Maritime Organization Assembly Resolutions from 1959 to 2023. Traditional frequency-based methods primarily identify high-frequency function words, whereas text dispersion-based analysis highlights keywords consistently distributed across the corpus, ensuring thematic representativeness. Gries' (2021) hybrid method, integrating both frequency and dispersion measures, effectively identifies keywords that are not only contextually relevant but also semantically significant, providing a more nuanced understanding of specialized discourse. The corpus is divided into three distinct periods based on major regulatory and technological milestones, revealing significant semantic shifts in maritime discourse. Keywords such as "safety" and "pollution," initially having narrow meanings, have notably broadened their semantic scope in response to evolving international regulations, technological advancements, and heightened environmental awareness. This study contributes to corpus linguistics by demonstrating that advanced keyword extraction methods, such as Gries' hybrid approach, offer deeper insights into how specialized languages evolve in response to dynamic industry contexts. **(National Korea Maritime and Ocean University · Dalian University of Foreign Languages · University of Guam)**

Keywords IMO Assembly Resolutions, hybrid keyword extraction, diachronic analysis, Maritime English, semantic change

---

* This paper is an abridged version of Mengjiao Suo's PhD Dissertation that will be published in August, 2025.

** First author
*** Corresponding author

1. Introduction

Keyword analysis is a key method in corpus linguistics and corpus-assisted discourse studies, helping researchers identify terms characteristic of specific discourse domains (Egbert and Biber 2019). This method has advanced with developments in computational linguistics and machine learning, evolving from basic frequency counts to more sophisticated techniques like chi-square tests, log-likelihood ratios, and dispersion measures, which improve the precision and relevance of extracted keywords. Traditional methods, such as frequency-based analysis, often fail to capture the contextual relationships between words, leading to irrelevant term extraction. Text dispersion-based analysis addresses these limitations by focusing on evenly distributed keywords across the corpus, while Gries' (2021) hybrid method integrates both frequency and dispersion for more accurate and contextually meaningful results.

Maritime English is the standardized language mandated by the International Maritime Organization (IMO) for communication in the maritime industry. It plays a crucial role in ensuring clear and effective communication among seafarers from diverse linguistic backgrounds. This specialized branch of English encompasses a lexicon and set of expressions essential for maintaining safety, facilitating navigation, and conducting operations at sea. The IMO requires the use of Maritime English to mitigate misunderstandings and promote safety across the global maritime sector, where clarity and precision are of paramount importance. As noted by Trenker (2000), Maritime English acts as a fundamental communicative medium within the international maritime community, supporting safe navigation and enhancing global maritime trade. This specialized form of English includes distinct terminologies and communication protocols vital for ensuring safety and operational efficiency at sea. Given the significant impact of communication on navigation, vessel operations, and emergency management, research on Maritime English is essential. Previous research has emphasized the importance of formulaic expressions in Maritime English, particularly in institutional genres such as marine accident investigation reports, where native and non-native professionals demonstrate different usage patterns (Jhang et al. 2018).

The present study applies Gries' (2021) method of keyword extraction to explore diachronic changes in the IMO Assembly Resolutions over the past 60 years (1959–2023). The corpus was divided into three distinct periods (1959-1973, 1975-1999, and

2001-2023) to track the diachronic changes in maritime discourse and keyword usage. This division was based on key historical, regulatory, and technological milestones in the maritime industry, which had significant implications for the language used in the IMO Assembly Resolutions (IMOAR). The first period (1959-1973) focuses on the early years of international maritime safety regulations, particularly the implementation of the International Convention for the Safety of Life at Sea (SOLAS). The second period (1975-1999) corresponds to the growing emphasis on environmental concerns, marked by the introduction of more stringent pollution control measures and sustainable practices. Around 2000, environmental regulations in the maritime sector gradually became a focal point. For example, new provisions in the International Convention for the Prevention of Pollution from Ships (MARPOL) increased the requirements for controlling ship emissions and oil pollution. These new environmental measures and policies necessitated more precise terminology for discussing pollution prevention, waste management, and environmental impact assessments. As a result, the development of Maritime English saw an expansion of terms related to environmental protection and sustainable practices. The third period (2001-2023) reflects the industry's adaptation to contemporary challenges such as maritime security, globalization, and climate change, with new regulations addressing these evolving issues. This temporal division enables a clearer analysis of how language and keyword usage have evolved in response to these significant shifts in the maritime sector.

Building on previous studies of keyword analysis, this research seeks to address the following questions:

1) What are the advantages of Gries' (2021) methodology over traditional methods in the extraction of keywords from Maritime English?
2) What are the diachronic changes in the meanings of keywords in the IMOAR from 1959 to 2023?

This paper is structured as follows: Section 2 provides a literature review on the keyword analysis. Section 3 outlines the data collection, research method, and procedures. Section 4 presents the research results, examining diachronic changes in IMOAR over nearly 64 years. Finally, Section 5 concludes the study and discusses its pedagogical implications for the field of Maritime English.

2. Previous studies

Keyword extraction is a fundamental technique in corpus linguistics, playing a critical role in identifying salient terms within a corpus that encapsulate the main themes or topics of the text. Early methods of keyword extraction were predominantly based on simple statistical measures, particularly frequency analysis, which identifies the most frequently occurring words in a text or corpus. Scott (1997) developed WordSmith Tools, a foundational software enabling researchers to generate word frequency lists and highlight statistically prominent words relative to a reference corpus. While effective in identifying high-frequency terms, such methods often fail to capture semantic relevance, as they tend to prioritize frequent words that may lack thematic significance (Rayson and Garside 2000). Baker (2004) emphasizes combining frequency with contextual analysis in keyword studies.

In response to these limitations, more sophisticated statistical techniques, such as log-likelihood ratios (LLR) and chi-square tests, were introduced. These methods improved keyword extraction by focusing not just on word frequency but also on the relative significance of words within a specific text, compared to a general reference corpus (Scott and Tribble 2006). Additionally, dispersion measures were employed to address the contextual distribution of keywords across a corpus, ensuring that the extracted words were not merely localized to a few sections but were broadly representative of the entire text (Baker 2004). Xu (2021) categorized the evolution of keyword extraction into three approaches: (i) the traditional frequency-based approach pioneered by Scott (1997) and Scott and Tribble (2006); (ii) text dispersion-based analysis, as exemplified by Egbert and Biber (2019), which emphasizes the contextual distribution of keywords across a corpus; and (iii) a hybrid approach that combines frequency with dispersion measures. Liu (2025) summarized these three approaches, emphasizing their complementary strengths and their applicability across different research contexts. Text dispersion-based and hybrid approaches have been lauded for their effectiveness in domain-specific corpora, where identifying thematically central terms is critical. These methods are valuable for tracking keyword changes over time or across sub-corpora in discourse studies. For example, Xu and Jhang (2020) demonstrated the effectiveness of keyword analysis when examining specialised legal contracts, such as charter parties. They showed that salient terms reflect the communicative functions and legal specificity of maritime texts.

Gries (2021) introduced a keyword extraction method combining frequency and dispersion analysis using Kullback-Leibler (KL) divergence. This approach assesses both word frequency and distribution, addressing the limitations of earlier methods by identifying contextually relevant keywords. It is particularly effective in specialized corpora and valuable for diachronic or synchronic discourse studies, enabling researchers to track shifts in keyword significance over time or across sub-corpora.

Diachronic analysis in corpus linguistics refers to the study of language change overtime, typically through the examination of texts produced in different periods. This form of analysis is invaluable for tracking shifts in discourse, terminology, and thematic focus. In the context of Maritime English, diachronic keyword analysis offers insights into how the priorities and concerns of the maritime industry, as reflected in texts such as the IMOAR, have evolved in response to technological advancements, regulatory changes, and global events.

Research on diachronic changes has often employed keyword extraction methods to track the evolution of language use within specific domains. For example, studies on the British National Corpus (BNC) have demonstrated how the meanings and usage frequencies of certain words have shifted due to sociopolitical changes (Leech 2007). Similarly, Baker (2006) showed how keyword analysis could be used to trace the transformation of discourses in political and environmental texts over several decades. In such studies, keyword extraction is a method not only for identifying high-frequency words but also for observing how key concepts and terminologies gain or lose prominence over time. Maritime English is a highly specialized domain. Keyword extraction and diachronic analysis can reveal how the language of the maritime industry has changed in response to external factors such as new regulations, technological innovations and environmental concerns. For instance, terms like "navigation" and "safety", which were predominant in earlier decades, have gradually been accompanied by terms such as "sustainability" and "security" as the maritime industry adapts to emerging challenges like climate change and international security threats (Trenker 2000). Jhang and Lee (2016) examined themes and trends in maritime economics and logistics by analyzing author keywords of international journals. These changes in keyword usage reflect not only shifts in linguistic practice but also broader trends in maritime policy and global trade. Keyword analysis of maritime institutional texts has also been employed to identify specialized discourse patterns and thematic focus in legal genres (Lu et al. 2017). This provides a basis for investigating changes

in domain-specific language over time.

The present study aims to examine these diachronic changes in the language of maritime discourse. By comparing the results of Gries' (2021) method with those produced by traditional keyword extraction techniques, this research seeks to uncover the semantic shifts that have occurred in the maritime industry over the past six decades. The field of keyword analysis has evolved from basic frequency counts to sophisticated techniques that account for both the statistical significance and the contextual distribution of words. Gries' (2021) method represents a significant advancement in this area, offering a more precise approach to extracting keywords in specialized corpora. The application of this method to the analysis of Maritime English, particularly in the diachronic study of the IMOAR, promises to yield important findings on the semantic changes that have taken place within the maritime industry. As the field continues to advance, keyword extraction will remain a critical tool for understanding how language reflects and shapes domain-specific discourse over time.

3. Data and methodology

3.1 Data

In the present study, the texts in the IMOAR were selected as the study corpus. As the principal international body governing global maritime affairs, IMO plays a pivotal role in shaping maritime regulations, policies, and technical standards. IMOAR as official documents, possess a high degree of authority and credibility, effectively representing the dominant discourse and policy orientations within the international maritime domain across different periods. Spanning from 1959 to 2023, this research corpus encompasses over six decades of maritime governance and regulatory evolution, offering a comprehensive basis for examining the diachronic development of Maritime English.

IMOAR is a self-built, domain-specific corpus comprising 1,093 texts from 1959 to 2023 with the latest data updated to 2023. The corpus was compiled and finalized by the end of 2023. All texts were sourced from the official IMO website (https://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions) and converted from PDF

to TXT format to enable computational analysis. The division of the IMOAR corpus—P1 (1959–1973), P2 (1975–1999), and P3 (2001–2023)—is grounded in significant shifts in maritime governance priorities. P1 corresponds to the IMO's foundational phase, marked by the establishment of core conventions and culminating in the adoption of MARPOL in 1973, which signaled a pivotal move toward environmental regulation. P2 reflects a period of institutional consolidation and regulatory expansion, encompassing the implementation of key instruments such as revised SOLAS provisions and the renaming of IMCO to IMO. P3 begins in the post-9/11 era, characterized by a pronounced emphasis on maritime security (e.g., the ISPS Code), alongside emerging concerns related to climate change, digitalization, and sustainability. This periodization captures major transitions in international maritime policy and provides a coherent framework for diachronic discourse analysis. To provide an overview of the corpus composition and its diachronic segmentation, Table 1 summarizes key descriptive statistics across the three designated periods.

Table 1. Descriptive statistics of IMOAR and three periods

Statistical items	IMOAR	1959-1973	1975-1999	2001-2023
Texts	1,093	393	411	289
File size	15,424,671	1,773,641	5,106,841	7,982,218
Tokens (running words)	2,412,543	285,770	808,102	1,227,132
Tokens used for word list	2,221,547	271,340	755,125	1,107,030
Types (distinct words)	28,419	8,639	16,222	18,299
TTR	1.28	3.18	2.15	1.65
STTR	30.5	30.80	31.46	29.87
SD (Standard Deviation)	69.54	68.86	68.24	69.20

The choice of a reference corpus is critical for analyzing keywords in the IMOAR. Leech (2002) emphasizes that a reference corpus should be comprehensive and representative of general language trends, ensuring a robust basis for comparative analysis. Scott and Tribble (2006) highlight that it should be larger and more general than the study corpus to identify significant deviations indicating specialized usage or emerging trends. As the reference corpus for this study, BNC Baby—a subset of the British National Corpus—was selected. It comprises four one-million-word samples representing distinct genres: fiction, newspapers, academic writing, and spoken conversation. Table 2 summarizes its structure, providing a baseline for comparison with the study corpus.

Table 2. Descriptive statistics of BNC Baby

BNC Baby	Classification	Texts	Tokens	Types
Spoken	Conversation	30	947,786	17,766
Written	Academic	30	1,018,664	36,479
	Fiction	25	1,018,924	36,119
	Newspapers	97	973,733	46,182
	Subtotal	152	3,011,321	77,058
Total		182	3,959,108	80,309

Both corpora are sufficiently large and diverse, ensuring statistically reliable comparisons and flexible alignment across different genres. Accordingly, the corpora provide a sound foundation for conducting keyword analysis and diachronic comparisons with confidence. The scale of the data ensures that keyword analysis and diachronic comparisons can be conducted with confidence in the reliability and representativeness of the results.

3.2 Methodology

Keyword extraction is a foundational task in natural language processing (NLP) and is essential for applications such as information retrieval, content summarization, and data analysis. The traditional keyword method employed in this paper refers to frequency-based analysis using Scott's WordSmith Tools, which calculates keyness scores based on log-likelihood ratios (LLR). Additionally, text dispersion-based analysis, recently highlighted by Egbert and Biber (2019), emphasizes the distribution of keywords across texts. Gries' (2021) hybrid method, integrating both frequency and dispersion measures, is employed as the primary keyword extraction approach which provides a more comprehensive and contextually relevant means of identifying keywords within a corpus.

This study is based on Gries' (2021) research method and utilizes Python to extract the relevant keywords. Python, with its extensive NLP libraries, provides a powerful and efficient environment for keyword extraction, enabling the identification of key concepts within large text corpora. Initially, data is collected, cleaned,and organized from various sources, whether structured or unstructured, such as web pages, research articles, or customer reviews. During the text preprocessing stage, tasks like tokenization (splitting text into words or sentences), removal of punctuation and stop

words (common words like "and" or "the" that do not contribute significant meaning), and text simplification are performed. Feature extraction follows, where the text is transformed into numerical representations to quantify the importance of each term within the corpus. Subsequently, various algorithms are applied to extract the most relevant keywords, which are then reviewed to ensure they accurately reflect the corethemes and concepts of the corpus.

4. Results and discussions

4.1 Comparison of three keyword extraction methods

This section compares three keyword extraction methods (frequency-based analysis, text dispersion-based analysis, and Gries' hybrid method) used to analyze Maritime English, highlighting the advantages of Gries' (2021) method over the other two methods. Keyword extraction is a critical process in identifying relevant terminology within a specialized corpus, and the choice of method can significantly affect the accuracy and relevance of the extracted keywords. While a traditional frequency-based method has long been employed to identify high-frequency words, they often fail to capture the nuanced, context-specific terms that are essential in specialized domains such as Maritime English. The text dispersion-based method is used in keyword extraction to identify terms that are not only frequent but also widely distributed across a corpus. By considering both frequency and distribution, this method helps highlight terms that are thematically central and broadly representative of the entire text, thus improving the reliability of keyword analysis in discourse studies. In contrast, Gries' mixed method combines frequency and dispersion measures, allowing for the identification of more contextually relevant keywords.

Table 3 below presents the results of the three keyword extraction methods. Gries' hybrid method yields more semantically relevant and domain-specific keywords such as "pollution" and "requirements," whereas the frequency-and dispersion-based methods tend to include high-frequency function words like "should" and "of," which are less informative in domain-specific analyses.

Table 3. Top 30 keywords among three keyword extraction methods

Rank	Frequency-based method	Text dispersion-based method	Gries' hybrid method
1	should	annex	accordance
2	of	maritime	annex
3	ship	tonnage	resolution
4	the	inspections	ships
5	ships	shipboard	provisions
6	resolution	tankers	requirements
7	safety	certification	ship
8	code	bulkheads	convention
9	solas	navigation	maritime
10	maritime	recalling	safety
11	international	ballast	practicable
12	cargo	radiotelephone	operational
13	reg	safety	cargo
14	equipment	bulkhead	applicable
15	assembly	discharges	appropriate
16	convention	complying	assembly
17	organization	thereto	guidelines
18	or	piracy	regulations
19	annex	oil	res
20	adopted	cargo	regulation
21	res	ships	adopted
22	imo	plating	organization
23	be	pollution	compliance
24	guidelines	confirming	pollution
25	requirements	extinguishing	required
26	and	thermocouples	recalling
27	oil	inclining	appendix
28	regulation	shipowner	appliances
29	provided	secretariat	tankers
30	pollution	appliances	inspections

As shown in Table 3, keywords that occur with high frequency and high dispersion are likely to be selected across all three methods. To illustrate the differences in keyword outputs among the three methods, Table 3 lists the top 30 keywords identified by each approach. From Table 3, Gries' (2021) method extracts significantly more keywords than the traditional frequency-based method and the text dispersion-based method. For instance, Gries' method extracts terms like "accordance," "annex," "resolution," "ships," "requirements," and others, which are directly related to the

maritime field. In contrast, the traditional frequency-based method extracts high-frequency words like "should," "of," "ship," "the", and "ships," which are often function words and not as contextually relevant to the specialized maritime discourse. These high-frequency function words, though common, do not contribute significantly to domain-specific analysis. On the other hand, text dispersion extracts keywords including "annex," "maritime," "inspections," "shipboard," "tankers," "navigation," "ballast," "radiotelephone," "discharges," and "piracy." Thus text dispersion-based method focuses on how evenly a word is distributed across texts and highlights domain-specific terms, avoids common function words. One of the key advantages of Gries' method is that it improves the contextual relevance and specificity of the extracted keywords. Gries' method uses a combination of frequency and dispersion to extract keywords that are more reflective of the specialized vocabulary within the maritime domain. For example, terms such as "maritime," "cargo," "safety," and "regulations" are highly relevant to the maritime industry and reflect the growing focus on safety, regulatory compliance, and environmental protection. However, keywords like "should" and "of" extracted by the frequency-based method are less contextually meaningful and fail to capture the specific terminology that characterizes the maritime sector. Furthermore, the text dispersion-based method ignores the frequency strength. It may identify rare words if they are evenly distributed. The text dispersion-based method also produces keywords with less sensitive to semantic centrality.

In the present study, by combining frequency and dispersion measures, Gries' method improves keyword extraction by filtering out irrelevant high-frequency function words and identifying contextually significant terms. This enables the extraction of precise, domain-relevant keywords such as "pollution," "required," and "applicable," which reflect key maritime concerns including environmental protection, regulatory compliance, and operational standards. However, these two traditional methods which are frequency-based method and text dispersion-based method, while capable of identifying frequently occurring terms and extracted the related terms, does not offer the same level of accuracy in capturing domain-specific vocabulary. While LLR and dispersion methods have their individual merits, Gries' hybrid method provides a more refined, meaningful, and contextually relevant set of keywords. In the study of Maritime English, where clarity, regulation-specific language, and international standardization are important, this method proves significantly more

effective. It ensures that rare but important domain-specific terms are retained if they have a strong enough presence across the corpus. This makes it ideal for identifying core terminology in specialized fields like maritime regulations.

To better analyze the two-dimensional keyword significance of words in the corpus, we can also use a scatter plot to examine the relationship between word frequency and dispersion. By plotting normalized frequency against the difference in dispersion, we can identify which words exhibit high keyness, reflecting their significance within the domain-specific context. The words located towards the top-right corner of the plot represent those that are both frequent and widely distributed, signaling their relevance in the study corpus. This visualization illustrates how keywords extracted by Gries' method tend to cluster in the high-frequency, high-dispersion quadrant, underscoring their contextual relevance. In contrast, function words highlighted by frequency-based methods appear more scattered, indicating weaker semantic contribution.

To complement the tabular comparison, Figure 1 below visualizes the distribution of keywords across frequency and dispersion dimensions.

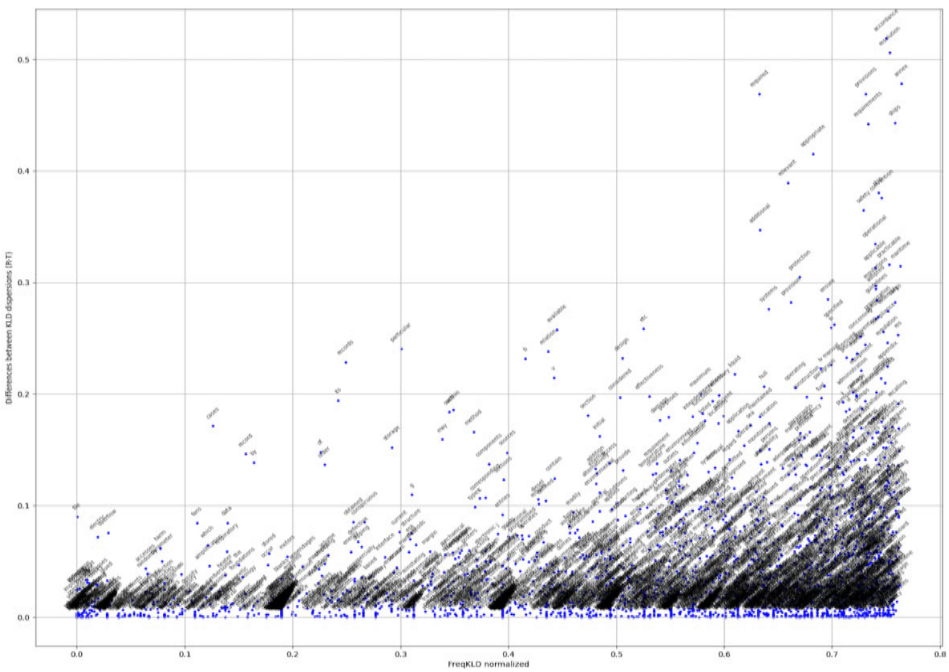

Figure 1. Distribution of keywords based on frequency and dispersion

In Figure 1 above, the values on the x-axis show how frequently certain words appear relative to others in the corpus, with words that have higher frequencies appearing further to the right. Y-axis represents the difference in dispersion, which indicates how widely a word is distributed across the corpus. A higher value suggests that the word is more evenly distributed across different sections of the corpus, while a lower value may indicate that the word appears more clustered within certain sections. Scatter points represent individual words from the study corpus, plotted based on their normalized frequency and dispersion. This figure illustrates a two-dimensional representation of keyword salience within the study corpus. All the keywords were extracted by Gries' hybrid method. Words located in the top-right quadrant are both highly distinctive in frequency and widely distributed, suggesting they are strong candidates for true keywords that are representative of the entire study corpus. These include terms such as "ballast," "tankers," "bulkheads," "corrosion," and "shipboard," which are characteristic of early maritime regulatory language.

Conversely, words appearing along the x-axis but with low y-axis (i.e., low dispersion differences) are frequently occurring but may be clustered in only a few documents, thus raising caution about their representativeness. Words positioned higher along the y-axis but with lower frequency-based distinctiveness may indicate items that are well distributed yet, not particularly more frequent in the study corpus, these may include generic or function-related terms.

In summary, the scatter plot could help us to observe the relationship between word frequency and dispersion in the target corpus, offering insights into the significance of keywords within the corpus. Words in the top-right quadrant represent significant domain-specific terms that are both frequent and dispersed, which are crucial for us to analyze keyword trends and changes.

4.2 Diachronic changes in the semantics of related keywords

In this section, we examine the diachronic changes of the semantics of related keywords within the maritime corpus from 1959 to 2023. The goal is to track how the meanings and usage of these terms have changed in response to shifts in the maritime industry's regulations, technology, and environmental concerns. By analyzing key terms across different periods, we can identify significant trends and transformations in the

discourse surrounding maritime issues. To explore the semantic changes, we first selected a set of related keywords based on their prominence and relevance within the maritime corpus. These keywords were extracted using Gries' hybrid method, which combines frequency and dispersion to identify contextually significant terms. Based on the purpose of IMO, the United Nations specialized agency with responsibility for the safety and security of shipping and the prevention of marine and atmospheric pollution by ships. IMO's work supports the UN sustainable development goals. Therefore, keywords such as "pollution" and "safety" reflect key themes in maritime discourse, including environmental protection, regulatory compliance, and maritime safety.

4.2.1 Semantic changes of "safety" in three different time-periods.

4.2.1.1 "Safety" in time-period (1959 to 1973)

In the early years (1959-1973) of the IMO, the term "safety" was closely associated with fundamental maritime safety practices, as reflected in the top ten collocate words, shown in Table 4.

Table 4. Collocate words about "safety" in 1959 to 1973

Collocates	Freq(Scaled)	FreqLR	Likelihood	Effect
maritime	1340	55	301.811	5.326
navigation	1420	34	150.345	4.549
sea	2560	33	107.077	3.655
life	1370	23	85.976	4.037
improve	330	15	85.214	5.473
measures	1320	22	81.929	4.026
standards	1170	18	64.294	3.911
relating	1000	17	63.927	4.055
ship	8650	41	61.825	2.212
ensure	3480	26	58.793	2.868

As seen in Table 4 above, the word "safety" frequently co-occurred with terms like "maritime," "navigation," "sea," and "life," signaling a strong focus on safeguarding lives at sea through basic regulatory measures, particularly under the framework of

the International Convention for the Safety of Life at Sea (SOLAS). These collocates reflect the initial regulatory priorities of the IMO, which centered around fundamental navigational safety and life preservation. The concentration of terms related to sea, navigation, and life also indicates that the semantic scope of "safety" during this period was narrower and more operationally focused compared to later decades. This period saw "safety" primarily defined in terms of physical security and codified responses to maritime emergencies, indicating a reactive approach to maritime safety during the early development of international regulations. The collocation patterns in this period emphasize the foundational role of ensuring maritime safety through structured and standardized measures, reflecting the IMO's early objectives to protect life at sea.

4.2.1.2 "Safety" in time-period (1975 to 1999)

By the 1990s, as the maritime industry globalized, the concept of "safety" expanded to include a more proactive approach. It began to frequently collocate with terms like "regulations," "guidelines," and "management," shifting from basic safety measures to a focus on risk management and integrated safety systems, as shown in Table 5 below. This evolution signals a more integrated understanding of maritime safety during this period.

To illustrate this shift in the semantic scope of "safety" during the second period, we present the top collocates of the term extracted from the corpus between 1975 and 1999. In the following table, the term "safety" increasingly collocated with regulatory and managerial terms such as "regulations," "ensure," and "management," suggesting a shift toward proactive, system-level safety strategies..

Table 5. Collocate words about "safety" in 1975 to 1999

Collocates	Freq(Scaled)	FreqLR	Likelihood	Effect
concerning	6490	233	631.842	3.217
prevention	5270	210	609.525	3.368
regulations	7050	222	549.36	3.028
certificate	4890	184	514.477	3.285
guidelines	12610	236	374.709	2.278
environmental	5810	159	353.707	2.826
pollution	2670	111	330.528	3.429

convention	13340	200	248.295	1.958
management	2990	72	144.09	2.641

As shown in Table 5, the concept of maritime safety expanded beyond basic physical protection to encompass broader operational and environmental concerns—such as pollution and regulatory compliance—reflecting a shift toward a more strategic and integrated safety management system from the 1980s onward. This shift signals a transition from reactive safety practices to a more systemic, risk-based approach, emphasizing organizational control and strategic regulation in maritime safety discourse. This trend highlights the growing complexity of the maritime industry and the need for robust safety practices that address both human and environmental risks in an increasingly globalized context.

4.2.1.3 “Safety” in time-period (2001 to 2023)

In recent years (2001 to 2023), “safety” has further evolved to reflect contemporary challenges in the maritime industry, particularly in the context of increased globalization and complex supply chains. Keywords such as “culture,” “compliance,” and “risk” have become prominent collocates. This evolution indicates a shift towards a holistic understanding of safety that integrates human factors, technology, and regulatory compliance. The current discourse around safety emphasizes not just adherence to regulations but the cultivation of a safety-oriented culture within maritime organizations, showcasing the term’s expanding semantic range.

Table 6 below presents the top collocates of “safety” from 2001 to 2023, highlighting the increasing associations with terms such as “management,“ ”compliance,“ and ”environmental,“ which reflect a holistic and institutionalized view of maritime safety.

Table 6. Collocate words about “safety” in 2001 to 2023

Collocates	Freq(Scaled)	FreqLR	Likelihood	Effect
security	6910	287	1187.489	4.319
management	6870	251	975.895	4.134
prevention	5470	215	865.916	4.239
protection	7960	219	733.036	3.724
environmental	3640	129	493.335	4.09
risk	5470	142	459.41	3.641

guidelines	5060	120	367.965	3.51
regulations	6540	130	356.663	3.256
system	21810	182	233.99	2.003
concerning	5250	92	231.445	3.074
fire	16390	148	208.022	2.117
compliance	3490	73	206.88	3.329
pollution	9010	106	193.916	2.499
cultural	462340	1409	191.554	0.55

The evolution of collocates like "management," "risk," "compliance," and "cultural" shows a shift toward a more integrated approach to maritime safety. Compared to earlier periods, the collocational profile in Table 6 highlights how maritime safety now encompasses psychological and institutional dimensions, illustrating the increasing importance of safety culture and organizational behavior. Safety now involves not just physical security but also human factors, organizational culture, risk management, and adherence to regulations. The emphasis on "risk" and" compliance" highlights the growing focus on addressing uncertainties and following international regulations. The presence of "environmental" and "pollution" reflects the sector's increasing environmental responsibilities, including protecting ecosystems from marine pollution. The focus on "prevention" and "cultural" points to a proactive approach, with a growing recognition of the importance of safety culture and prevention for long-term sustainability.

Data from 2001 to 2023 shows a clear trend toward a more holistic view of maritime safety, emphasizing risk management, compliance, safety culture, and environmental protection. This reflects the industry's adaptation to global challenges, technological advances, and stricter regulations, making safety a multifaceted priority.

4.2.2 Semantic changes of "pollution" in three different time-periods

4.2.2.1 "Pollution" in time-period (1959 to 1973)

The term "pollution" gained prominence in the IMOAR in response to significant environmental incidents, particularly in the 1970s. Early collocates included "prevention," "control," and "marine," reflecting a focus on developing regulations to mitigate pollution from shipping activities. The adoption of MARPOL illustrated

the growing recognition of environmental impacts and the need for formalized protocols to address these concerns. The semantic field of "pollution" was closely tied to discussions on regulatory frameworks and compliance measures aimed at reducing marine pollution.

Table 7 shows the collocation related to "pollution" from 1959 to 1973. These early collocates reflect an emerging awareness of marine pollution, focusing on basic contaminants such as oil and the necessity of initial regulatory responses. Specifically, the collocational profile of the period further illustrates these early concerns about maritime pollution by highlighting key terms that co-occurred with "pollution," particularly those emphasizing regulatory actions and marine-specific contexts.

Table 7. Collocate word of "pollution" in 1959 to 1973

Collocate	Freq(Scaled)	FreqLR	Likelihood	Effect
prevention	440	33	313.5	8.216
marine	900	32	255.401	7.139
oil	6100	28	110.431	4.185
sea	2560	19	92.297	4.879
control	4720	18	64.52	3.918
environment	560	7	41.059	5.631
reduce	370	5	30.087	5.743
ships	6070	12	28.611	2.97
anti	100	3	22.82	6.894
risk	980	5	20.61	4.338
seas	150	3	20.382	6.309
deliberate	20	2	20.15	8.631

Early discussions on maritime pollution primarily centered on controlling incidents like oil spills. However, the increasing emphasis on terms such as "prevention" and "reduce" indicates a shift toward proactive measures, including preventive regulations and cleaner technologies. The frequent use of "control" highlights a regulatory focus on effectively managing pollution risks. The rising importance of "environment," "marine," and "pollution" signifies a deeper integration of environmental concerns within maritime regulations, extending beyond direct marine impacts to encompass broader ecological and atmospheric effects. This emphasis reflects IMO's focus on ensuring adherence to global standards through enforceable regulations.

Taken together, the IMO's regulatory discourse has evolved from a reactive

approach to environmental control to a proactive, integrative strategy.

4.2.2.2 "Pollution" in time-period (1975 to 1999)

As environmental issues gained traction in the global discourse, the meaning of "pollution" began to encompass broader concerns about ecological sustainability. In the 1990s (1975 to 1999), keywords such as "sustainability," "emissions," and "impact" emerged as frequent collocates, indicating a shift towards considering the long-term environmental consequences of shipping practices.

To support this shift in discourse, Table 8 below presents collocates of "pollution" from 1975 to 1999, revealing a more complex semantic field involving sustainability and long-term environmental concerns.

Table 8. Collocate words of "pollution" in1975 to 1999

Collocates	Freq(Scaled)	FreqLR	Likelihood	Effect
ships	17570	127	330.772	3.14
control	10150	99	311.289	3.572
oil	8400	89	293.339	3.692
safety	13710	90	218.541	3.001
recalling	6000	59	186.022	3.584
environment	2980	31	100.828	3.665
sustainability	390	16	94.53	5.645
risk	1990	23	79.287	3.817
impact	500	15	79.28	5.193
noting	3910	29	76.437	3.177
incidents	1260	19	75.032	4.201
combating	250	12	74.674	5.871
emissions	7000	34	64.71	2.566

This period marked a transition from merely preventing pollution to promoting sustainable maritime practices, reflecting an evolving understanding of the relationship between maritime operations and environmental health. This table reveals a strong thematic focus on sustainability, marine environment, and emissions-related issues. This evolution underscores a growing integration of environmental sustainability into

maritime policy, where the term "pollution" increasingly interacts with terms denoting long-term ecological consequences. High-frequency collocates such as "ships," "control," and "oil" indicate that during this period, pollution was primarily discussed in relation to maritime transport and oil-related incidents. These collocates also have high likelihood ratios, suggesting statistically significant associations with the node word.

Terms like "safety" and "environment" also appear frequently, signaling a growing awareness of the broader implications of pollution beyond technical control. The presence of "sustainability," though less frequent, shows the early emergence of a more forward-looking discourse that ties pollution to long-term ecological concerns. Notably, "combating" and "impact" show high effect scores, indicating that even when these terms occur less often, their co-occurrence with "pollution" is particularly meaningful and non-random. This suggests a rhetorical shift toward framing pollution as a threat requiring active mitigation.

Overall, the data suggests that the discourse on pollution during this period was shaped by practical maritime challenges (e.g., *oil, ships, emissions*) while gradually incorporating environmental and sustainability-related perspectives, reflecting the evolving priorities of maritime governance and global ecological discourse.

4.2.2.3 "Pollution" in time-period (2001 to 2023)

In recent years, "pollution" in 2001 to 2023 has taken on an even broader semantic significance, particularly with the advent of climate change discussions. The term is now frequently associated with keywords like "air", "atmospheric", and "climate". This evolution highlights the growing awareness of the maritime industry's role in contributing to global warming and the urgency of addressing these challenges through innovative solutions. The contemporary discourse on pollution emphasizes not just the prevention of traditional pollutants but also the need for comprehensive strategies to reduce carbon footprints and promote environmental stewardship in the maritime sector.

To further demonstrate how the semantic scope of 'pollution' expanded in recent years, Table 9 below presents key collocates from 2001 to 2023 that reflect the maritime industry's increasing concern with air pollution, climate change, and sustainability. These terms such as "climate," "air," and "atmospheric" demonstrate the expanded

scope of environmental discourse in the maritime industry.

Table 9. Collocate words of "pollution" in 2001 to 2023

Collocate	Freq(Scaled)	FreqLR	Likelihood	Effect
prevention	5470	435	3105.153	6.483
oil	16910	150	423.366	3.319
control	17420	145	392.541	3.227
air	3760	69	288.125	4.368
incidents	2390	52	234.151	4.614
safety	21100	106	193.916	2.499
climate	4440	55	188.576	3.801
prevent	2990	43	159.452	4.016
damage	5240	51	152.092	3.453
shipboard	2270	36	140.153	4.157
atmospheric	300	20	134.609	6.229
emergency	8540	57	131.585	2.909
reduce	1470	30	131.299	4.521

Data from Table 9 (2001 to 2023) highlight a significant evolution in the maritime industry's approach to "pollution." This trend suggests a broadened semantic scope for "pollution," now linked not only to specific substances but also to broader frameworks such as carbon neutrality and global environmental governance .While the term "prevention" remains central, reflecting ongoing efforts to address traditional sources like "marine" and "oil," there is a marked increase in the prominence of "air" and "climate". This shift indicates a growing recognition of the importance of addressing air pollution and climate change, as the industry acknowledges its role in mitigating greenhouse gas emissions and global warming.

In addition to traditional concerns, the industry's focus has expanded to include proactive risk management and sustainability. Key terms like "control," "ships," and "shipboard" emphasize enhanced technical measures and emergency preparedness. Additionally, the rise of terms such as "reduce," "damage," and "combating" illustrates efforts to mitigate long-term ecological impacts. The inclusion of "atmospheric" and "emissions" reflects an increased attention to climate-related challenges, demonstrating the industry's adaptation to the complexities of global environmental issues and its commitment to sustainable practices.

In summary, both "safety" and "pollution" have undergone semantic broadening in response to evolving industry standards and global challenges.

5. Conclusion

This study demonstrated how Gries' (2021) keyword extraction methodology can effectively track the diachronic changes in the semantics of keywords in the IMOAR from 1959 to 2023. By comparing Gries' hybrid method with traditional frequency-based analysis and text dispersion-based analysis, the research demonstrates that Gries's method offers a more nuanced, context-sensitive, and efficient way to extract keywords, particularly in specialized domains like Maritime English.

The analysis revealed significant diachronic shifts in the meanings of key maritime terms, reflecting the evolving concerns of the maritime industry. Keywords such as "safety" and "pollution" have broadened inscope, aligning with global regulatory changes, technological advancements, and growing environmental consciousness. However, certain topics—such as marine management—have remained consistent focal points over the past six decades. Keywords like "prevention," "marine," and "oil" show relative stability, indicating the sustained relevance of these issues. The study underscores the importance of keyword analysis in capturing both change and continuity, offering a deeper understanding of how language evolves in response to external factors such aspolicy shifts and emerging challenges like climate change and international security.

Gries' methodology, combining frequency and dispersion, has proven invaluable in tracking these semantic shifts, offering insights into the specialized discourse of Maritime English. The findings highlight the increasing significance of terms related to sustainability, risk management, and environmental protection, marking a departure from earlier, more narrowly defined concepts of safety and pollution. This diachronic approach not only enhances our understanding of maritime discourse but also contributes to the development of more accurate and comprehensive keyword extraction methods for specialized fields.

This research has important implications for both academic studies of Maritime English and practical applications in maritime education, policy, and communication. As the maritime industry continues to face new challenges, the language of maritime safety, environmental protection, and international regulation will likely evolve further. In light of these ongoing developments, continued monitoring of keyword trends and their meanings will be essential for ensuring effective communication in this crucial global sector.

References

Baker, Paul. 2004. Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics* 32(4): 346–359.

Bondi, Marina and Mike Scott. 2010. *Keyness in texts*. Amsterdam: John Benjamins Publishing Company.

Egbert, Jesse and Douglas Biber. 2019. Incorporating text dispersion into keyword analyses. *Corpora* 14(1): 77–104.

Gries, Stefan Th. 2021. A new approach to (key) keywords analysis: Using frequency, and now also dispersion. *Research in Corpus Linguistics* 9(2): 1-33.

Jhang, Se-Eun and Suho Lee. 2016. A study of themes and trends in research of global maritime economics through keyword network analysis. *Journal of Korea Port Economic Association* 32(1): 79-95.

Jhang, Se-Eun, Sungkuk Kim, and Yilian Qi. 2018. Lexical bundles in ESP writing: Marine accident investigation reports. *Linguistic Research* 35(Special): 105-135.

Leech, Geoffrey. 2002. The importance of reference corpora. Presented at *the Conference Hizkuntza-corpusak. Oraina eta geroa*. Spain. October 24-25, 2002.

Liu, Cuilin. 2025. *Integrating bibliometric techniques with corpus-driven bigrams: A research trend analysis of cruise tourism*. Unpublished PhD Dissertation. National Korea Maritime and Ocean University.

Lu, Wenyu, Sungmin Lee, and Se-Eun Jhang. 2017. Keyness in maritime institutional law texts. *Linguistic Research* 34(1): 51-76.

Rayson, Paul and Roger Garside. 2000. Comparing corpora using frequency profiling. *Proceedings of the Workshop on Comparing Corpora*, 1–6. Association for Computational Linguistics.

Scott, Mike. 1997. PC analysis of key words—and key keywords. *System* 25(2): 233-245.

Scott, Mike and Chris Tribble. 2006. *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam: John Benjamins Publishing Company.

Tenkner, Peter. 2000. Maritime English: An attempt at an imperfect definition. *Proceedings of the Second IMLA Workshop on Maritime English in Asia*, 1-8. Dalian, China: Dalian Maritime University.

Xu, Lin. 2021. *Developing vocabulary lists in specialized Maritime English corpora*. Unpublished PhD Dissertation. National Korea Maritime and Ocean University.

Xu, Lin and Se-Eun Jhang. 2020. Keyword analyses of English charter parties. *Linguistic Research* 37(2): 267-288.

Mengjiao Suo

PhD Student
Department of English Language and Literature
National Korea Maritime and Ocean University
727 Taejong-ro, Yeongdo-Gu,
Busan 49112, Republic of Korea
E-mail: 605355902@qq.com

Yaochen Deng

Professor
The School of English Studies
Dalian University of Foreign Languages
6 West Section, South Lushun Road,
Dalian 116044, Liaoning, China
Email: deng_yaochen@163.com

Hyunjong Hahm

Associate Professor
Department of English and Applied Linguistics
University of Guam
UOG Station, Mangilao,
Guam 96923, USA
Email: hhahm@triton.uog.edu

Se-Eun Jhang

Professor
Department of English Language and Literature
National Korea Maritime and Ocean University
727 Taejong-ro, Yeongdo-Gu,
Busan 49112, Republic of Korea
Email: jhang@kmou.ac.kr

Received: 2025. 03. 10.
Revised: 2025. 06. 05.
Accepted: 2025. 06. 14.